Identification of a preferential substrate in glucose dehydrogenase from *Pycnoporus cinnabarinus*: a "sugar sweet" structural approach

Dottorando
Gabriele Cerutti

Docenti guida   Coordinatore

Prof.ssa Beatrice Vallone   **Prof. Stefano Gianni**
Dott.ssa Carmelinda Savino
Dott. Giuliano Sciara

Dicembre 2019

On the cover Alexander Calder, "Skybird", 1974, litograph

*"If you want to go fast, go alone.
If you want to go far, go together."*

African proverb

Table of contents

List of abbreviations

AA: auxiliary activity

AAD: aryl-alcohol dehydrogenase

AAO: aryl-alcohol oxidase

ABTS: 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)

*Af*GDH: glucose dehydrogenase from *Aspergillus flavus*

*An*GOX: glucose oxidase from *Aspergillus niger*

CAZy: carbohydrate active enzymes

CDH: cellobiose dehydrogenase

CHOX: choline oxidase

CLB: cellobiose

DET: direct electron transfer

DCIP: 2,6-dicholorophenolindophenol

FAD: flavin adenine dinucleotide

GDH: glucose dehydrogenase

GLC: D-glucose

GMC: glucose-methanol-choline

GOX: glucose oxidase

GTA: 1,3;1,4 β-glucotriose A

GTB: 1,3;1,4 β-glucotriose B

LDAs: lignin-degrading auxiliary enzymes

LGC: D-glucono-δ-lactone

LMB: laminaribiose

LMEs: lignin-modifying enzymes

LPMO: lytic polysaccharide monooxygenase

*Mt*CDH: cellobiose dehydrogenase from *Myriococcum thermophilum*

P2O: pyranose 2-oxidase

PDH: pyranose dehydrogenase

*Pc*GDH: glucose dehydrogenase from *Pycnoporus cinnabarinus*

*Pc*GDH-GLC: glucose-bound *Pc*GDH

*Pc*GDH-LMB: laminaribiose-bound *Pc*GDH

r.m.s.d.: root-mean-square deviation

SBS: external sugar-binding site

*Tm*P2O: pyranose 2-oxidase from *Trametes multicolor*

In all Figures amino acid residues are indicated with the one-letter code.

Chapter 1

Introduction

1.1 Glucose dehydrogenase and glucose oxidase: substrate specificity and biological relevance

Glucose dehydrogenase (GDH) and glucose oxidase (GOX) belong to the oxidoreductase superfamily (EC 1.1). These enzymes are named after their ability to catalyze electron transfer reactions from a donor (reductant) to an acceptor (oxidant) molecule, generally using nicotinamide adenine dinucleotide (phosphate) - NAD(P) - or flavin adenine dinucleotide - FAD - as cofactors. Their catalytic cycle can be divided into two half-reactions, depending on whether the enzyme acts as the reducing or the oxidizing agent. In the first half-reaction the enzyme in the oxidized resting state is reduced by the primary electron donor; in the second half-reaction the reduced enzyme acts as the secondary electron donor, transferring electrons to an external (secondary) electron acceptor.

While GOXs always use FAD as the cofactor, GDHs represent a much more heterogeneous family, in which the redox nonprotein component can be NAD(P), FAD, or pyrroloquinoline quinone (PQQ), depending on the function and the organism in which they are expressed[1]. For both GDH and GOX the primary electron donor is a sugar substrate that provides electrons from the reducing hydroxyl end, *i.e.* the anomeric end. In most cases these enzymes show preferential activity toward D-glucose, which is converted into D-glucono-δ-lactone. On the other hand, GDH and GOX transfer electrons to different secondary electron acceptors: in general, oxidases utilize molecular oxygen, producing hydrogen peroxide, while dehydrogenases transfer

electrons to various natural or artificial redox partners (Fig. 1.1). GOXs and only FAD-dependent GDHs belong to the flavoproteins superfamily known as glucose-methanol-choline (GMC) oxidoreductase[2, 3].

GOX is a β-D-glucose:oxygen 1-oxidoreductase (EC 1.1.3.4) expressed in fungi and insects. The biochemical properties of GOX have been extensively characterized. The catalytic efficiency has been examined on fifteen different hexoses and derivatives to understand which structural features of the substrate are essential for catalysis; D-glucose is by far the preferential substrate, showing a 20-fold higher affinity and a fivefold higher turnover rate compared to other monosaccharides. The pyranose ring oxygen and equatorially oriented hydroxyl groups at position 1 and 3 proved to be of importance for the enzyme-substrate complex formation[4].
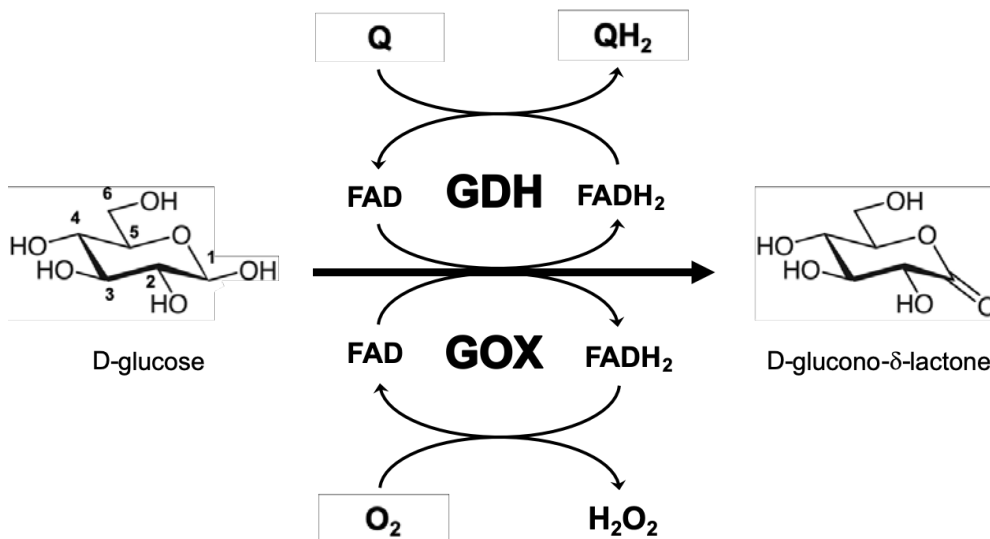
Figure 1.1. Catalytic cycle of GDH and GOX. Both enzymes accept electrons from D-glucose, which is oxidized at the C1 position and converted into D-glucono-δ-lactone, resulting in the reduction of the cofactor FAD. In GOX the secondary electron acceptor is molecular oxygen, which is converted into hydrogen peroxide; GDH instead transfers electrons to various different electron acceptors like quinones (Q). Hydrogen ions and electrons are omitted in the reaction scheme.

4

Nevertheless, the enzyme is able to act only on the β anomer of glucose, the mutarotation of the α form being rate limiting[5]. The pH dependency of glucose oxidation has been investigated for different GOXs and in all cases the optimum value is 5.5[4, 6]. The main biological function of GOX is to act as an anti-bacterial and anti-fungal agent through the production of hydrogen peroxide[7]. In the first reports after its discovery GOX was known as the anti-bacterial agent "notatin" since it was isolated from *Penicillium notatum*[8]. For example, in insects GOX is secreted from the hypopharyngeal gland of the bee into the nectar in order to provide acidic conditions and hydrogen peroxide, which are necessary to preserve honey[9]. The role of this enzyme in fungi includes assisting in plant infection and lignin degradation: once hydrogen peroxide is produced by GOX, it can be exploited by peroxidases to catalyze the oxidation of a variety of lignin model compounds[10, 11].

GDHs are classified according to the primary electron acceptor: NAD(P)-dependent GDHs are categorized in the EC 1.1.1 group, those utilizing PQQ are classified as EC 1.1.5, while the remaining GDHs belong to the EC 1.1.99 group, which includes those enzymes that utilize FAD as the cofactor. The systematic name that identifies the members of the latter group is D-glucose:acceptor 1-oxidoreductases since they use a variety of external electron acceptors, the FAD cofactor acting as the intermediary redox center. FAD-dependent GDHs have been found expressed in the periplasm of Gram-negative bacteria, in the cytoplasm of some insects and as extracellular enzymes in fungi. The first fungus-derived GDH from *Aspergillus oryzae* was reported in 1937[12], but the flavoprotein nature of this enzyme was identified thirty years later by Bak, who carried out the biochemical characterization of the purified enzyme[13]. Several monosaccharides were tested as electron donors and D-glucose turned out to be the best one. Unlike GOX, however, GDH did

not show stereospecificity toward the α or β anomer, both forms being oxidized at the same rate[14]. Because of the recent increasing interest in these enzymes for their potential use in biotechnological applications, several fungal GDHs have been recombinantly expressed and characterized[15-20]. Like *A. oryzae* GDH, they all show the highest catalytic efficiency for D-glucose, even though remarkable activity has been detected toward other sugars including mono and disaccharides, especially xylose. For all these enzymes the optimal pH is in the range between 5.5 and 7.5.

The functional characterization has historically been carried out using artificial redox dyes such as 2,6-dicholorophenolindophenol (DCIP). Its optical properties allow for an easy detection of the redox behavior in the experimental system: DCIP is blue in the oxidized form, with an absorption maximum at 600 nm, and colorless in the reduced form. But since DCIP is not a naturally occurring compound, the overall biological relevance of FAD-dependent GDHs still remains unclear: why did nature develop an alternative to GOXs and why do GDHs show preferential reducing activity toward aromatic acceptors rather than oxygen?

The available data on the biological role of these enzymes are very poor; since their discovery the scientific community has mainly focused on how to exploit the redox potential that these enzymes can accumulate rather than understanding their involvement in natural processes. The first remarkable hint about their physiological role was the observation that the expression yield of GDH from *A. oryzae* was enhanced by adding hydroquinone or benzoquinone in the growth media[13]. These naturally occurring compounds can form aromatic branched polymers in lignin, the major component of the woody plant cell wall. Lignin biosynthetic pathway includes a final polymerization step involving radical coupling of aromatic molecules produced by the action of

6

oxidative enzymes such as peroxidases and laccases. If the reactive radical intermediates are quenched by the redox activity of external enzymes, however, lignin cannot be produced. Many fungi are pathogenic plant parasites and their infection progresses through several stages. In the last necrotic stage, the plant reacts by increasing the free phenol content and upregulating the oxidative activity to produce radicals necessary for lignin biosynthesis. The neutralization of these radicals by fungal dehydrogenases can provide a powerful tool to inactivate the plant's defense mechanism.

This hypothesis regarding the function of GDH is strengthened by the observation that the quenching activity is achievable only if the enzyme has been first reduced by sugars, such as those released by the hydrolysis of the plant cell wall polysaccharide constituents[21]. Nowadays, this hypothesis on the biological function of GDH is the most embraced, since it contemplates the presence of sugars and aromatic electron acceptors in the naturally occurring environment where this enzyme is secreted. It is interesting to notice that the ligninolytic function is relevant also in the case of GOXs, confirming the close evolutionary relationship between GDH and GOX. From an evolutionary perspective, these two enzymes provide an example of the versatility of nature in choosing different chemical mechanisms to perform similar tasks.

The fascinating chemistry of GDH and GOX catalysis has also been responsible for an increasing interest in exploiting their redox properties in biotechnological applications. Nowadays, they are mainly employed as bioreceptors in biosensors in self-monitoring of blood glucose (see par. 1.4), but more recent applications include their use in biotransformation of aromatic compounds into high-value products (foods, cosmetics and pharmaceuticals) or industrial compounds (surfactants, adhesives and biomaterials) in the context of green chemistry and biorefinery[22].

7

1.2 Biochemistry of plant cell wall degradation

Cell walls are essential structural elements found in many organisms including prokaryotes, algae, fungi and plants; they are absent in animal cells. Their main role is to provide structural support and protection to the cell, and the shape needed to form tissues and organs. In plants, cell walls also play a crucial role in plant-microbe interaction, constituting the main defense barrier against potential pathogens. Plant cell walls can be divided into two categories: the primary walls that surround growing cells or cells capable of growth, and the secondary walls that are thickened structures surrounding fiber cells or vessel elements[23]. The secondary cell walls appear when the plant cell has stopped expanding, ensuring structural integrity and mechanical support, *e.g.* forming wood in some types of plants.

In terms of chemical composition, the woody material of plants is a complex mixture of three types of carbon-based polymer - cellulose, hemicellulose, and lignin - collectively called lignocellulosic biomass (Fig. 1.2). Cellulose is a linear polysaccharide consisting of glucose molecules condensed by a $\beta(1\rightarrow4)$ glycosidic bond; chains are held together in a crystalline structure by hydrogen bonds and van der Waals forces to form microfibrils[24]. Hemicellulose is a more heterogeneous polysaccharide: it can have either a branched or an unbranched structure and includes xyloglucans, xylans, mannans, glucomannans, and mixed $\beta(1\rightarrow3, 1\rightarrow4)$ glucans (Fig. 1.3)[25]. Lignin is a phenolic polymer largely composed of three major types of monomers (monolignols): *p*-coumaryl, coniferyl, and sinapyl alcohols[26], aromatic alcohols that differ in their extent of methoxylation. Monolignols are biosynthesized in the cytoplasm and then transported to the cell wall, where they are oxidized to phenoxy radicals that undergo polymerization by chemical coupling.
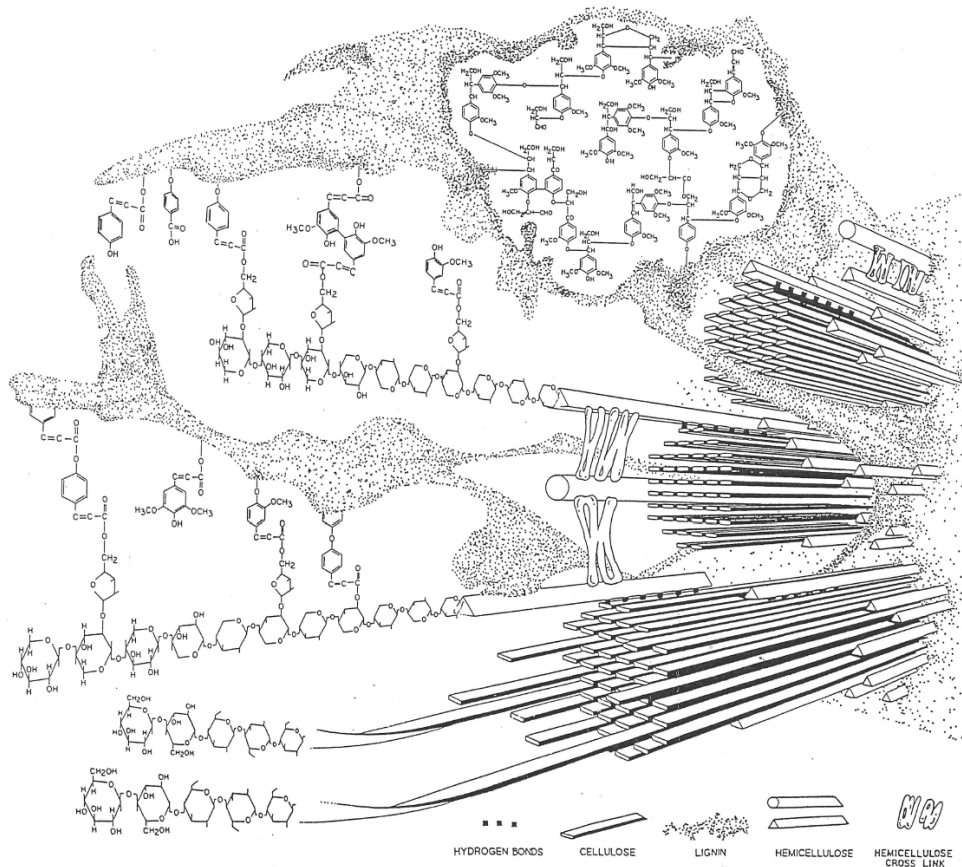
Figure 1.2. Structure of lignocellulose: the chemical description of the main components is included, together with the interactions responsible for the physical and chemical properties of this material. Cellulose microfibrils are held together through the adhesive action of hemicellulose and lignin[27].

Cellulose, hemicellulose and lignin form a composite matrix in which the hemicellulose and lignin polymers function as the recalcitrant adhesive for cellulose microfibrils (Fig. 1.2)[27]. This structural and chemical complexity requires an equally elaborated degradation system. Oxidative attack, cellulases, and mechanical disruption all reduce recalcitrance and improve access for depolymerizing enzymes.

Saprotrophic fungi are capable of degrading cellulose and hemicellulose through the collective action of multiple carbohydrate-active enzymes, typically acting together in a complementary and synergistic manner.

9

Glycoside hydrolases are the primary enzymes that cleave glycosidic bonds in cellulose and hemicellulose; they break the polysaccharide chain through endo- and exo-hydrolitic activity based on acid-base catalysis[28]. Their action is assisted by polysaccharide lyases and esterases, responsible for the removal of methyl, acetyl and phenolic esters. The crystalline packing of cellulose, however, makes most of the polysaccharide material inaccessible to hydrolases.

D-glucose

D-galactose

D-mannose

D-glucuronic acid

L-arabinose

D-xylose

L-Fucose

Xyloglucans

Mixed β(1→3, 1→4) glucans

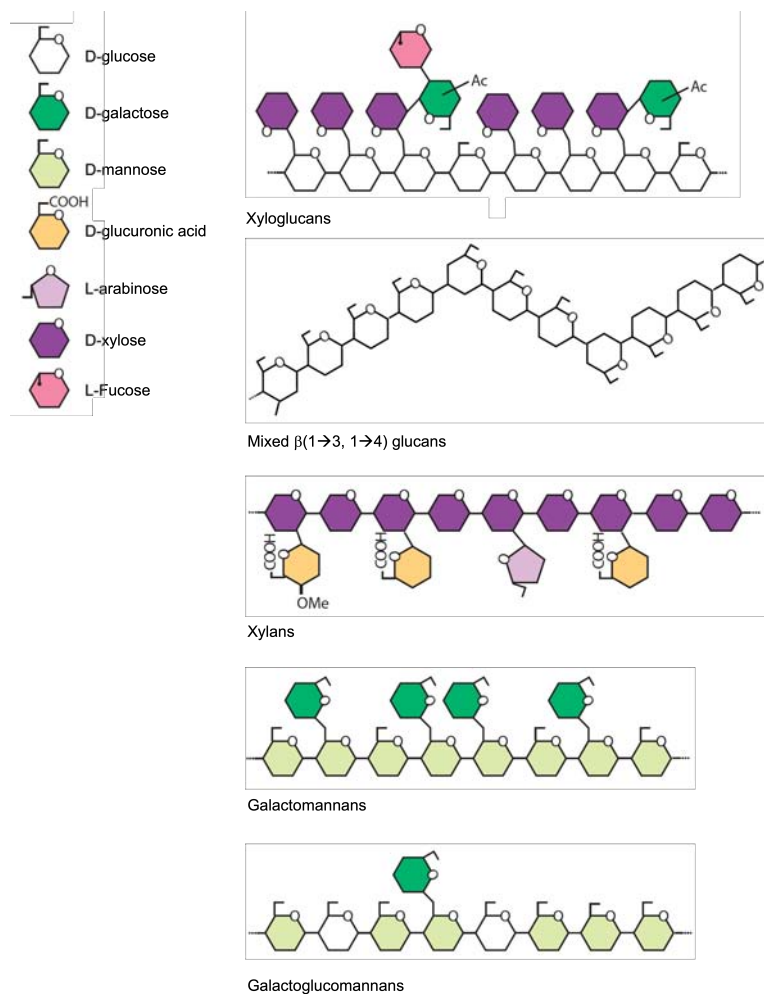Xylans

Galactomannans

Galactoglucomannans

Figure 1.3. Schematic representation of the types of hemicelluloses found in plant cell walls. Each sugar unit is colored according to the scheme on the top left[25].

10

Smaller "soluble" polysaccharide chains are produced by lytic polysaccharide monooxygenase (LPMO)[29]. These enzymes are able to bind the flat, solid, well-ordered surface of cellulose and catalyze the chain cleavage through an oxidative mechanism, providing new chain ends as starting points for hydrolases. LPMOs activate molecular oxygen and hydroxylate the polysaccharide backbone, eventually leading to chain breaks. Oxygen activation requires the reduction of the LPMO enzyme, which can be achieved through three electron transfer systems, shown in Figure 1.4: i) extracellular cellobiose dehydrogenase (CDH), ii) extracellular GMC oxidoreductases (*i.e.* GOX, GDH), and iii) plant-derived or fungal diphenols[30]. While CDH and endogenous diphenols directly reduce LPMO, GMC oxidoreductases contribute in an indirect way producing hydroquinones that act as redox mediators.

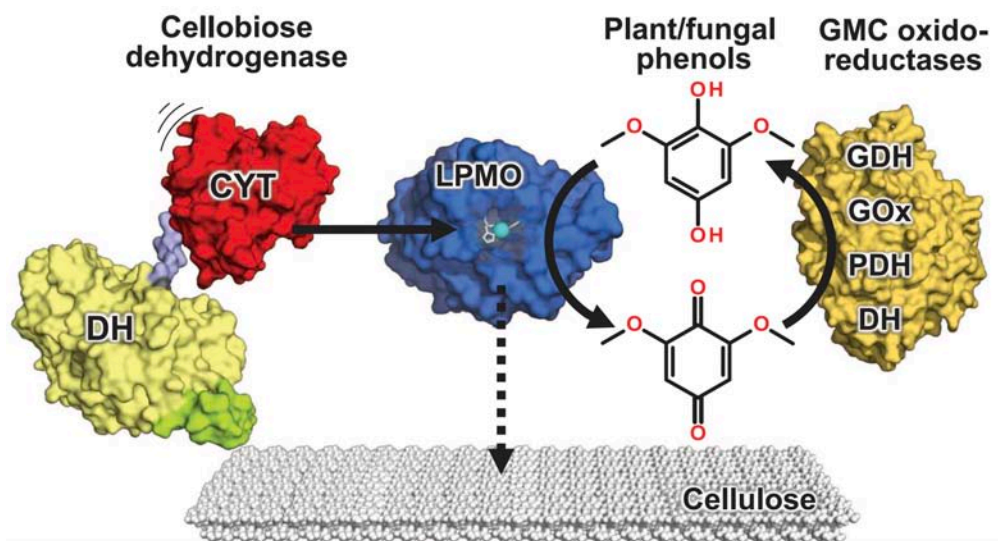

Figure 1.4. Three electron transfer systems reduce LPMO to initiate cellulose attack. CDH transfers electrons from its dehydrogenase domain (DH) via its mobile cytochrome domain (CYT) to the LPMO; various plant-derived and fungal phenols can act as direct sources of electrons; GMC oxidoreductases regenerate the reduced form of phenols, thus contributing indirectly to LPMO reduction[30].

11

The combined action of LPMO and GMC oxidoreductases thus provides a very efficient cellulose degradation system, generating a redox catalytic cycle where LPMOs are constantly fueled by the action of GMC oxidoreductases.

Lignin degradation is carried out by two different enzyme groups: lignin-modifying enzymes (LMEs) and lignin-degrading auxiliary enzymes (LDAs). LDAs are unable to degrade lignin on their own yet are necessary to complete the degradation process. LMEs are classified as phenol oxidases (laccases) and heme-containing peroxidases.

Laccases are extracellular, periplasmic and intracellular proteins that contain copper atoms in the catalytic site. Substrate oxidation by laccase is a single-electron reaction generating a free radical. The initial product of the reaction is typically an unstable phenoxy radical that can be converted to quinone in a second reaction catalyzed by the enzyme[31]. Studies on lignin model compounds have shown that the phenoxy radical formation leads to Cα oxidation, alkyl-aryl cleavage and Cα-Cβ cleavage[32]. The phenoxy species produced by laccases, however, are highly reactive and can re-polymerize easily through radical coupling.

Peroxidases attack lignin polymers in a relatively non-specific fashion because of their high redox potential; the catalytic cycle requires hydrogen peroxide to produce ferry oxo porphyrin species able to oxidize lignin-derived alcohols such as veratryl alcohol[27].

LDAs include many GMC oxidoreductases such as GDH, GOX, and CDH. All of these enzymes utilize molecules deriving from lignin, cellulose or hemicellulose degradation as electron donors and hence accumulate reduction potential. Once reduced, they can exhibit their auxiliary activities i) producing hydrogen peroxide from molecular oxygen, ii) quenching free radicals deriving from the activity of laccases, iii) reducing quinone compounds to fuel the

12

LPMO machinery. Hydrogen peroxide is required for the action of peroxidases, while the reduction of phenoxy radicals prevents them from re-associating into lignin. Moreover, some plants are found to actively secrete phenoxy radicals to build up lignin as a response to fungal infection in order to physically block fungal invasion; the reduction of these radicals by fungal enzyme systems counteracts these plant defense reactions[33].

The overall process of lignocellulose-degradation is the result of many different but complementary chemical reactions performed by a complex enzymatic system. Most of the enzymes involved are oxidoreductases and act trough an electron transfer mechanism. The complexity of the process is highlighted by the fact that few organisms can achieve lignocellulose breakdown; wood-rotting fungi are primary lignin degraders and are the only group of microorganisms capable of mineralization of this abundant plant product accounting for about 25% of removable organic matter in the biosphere[27]. It seems that the diversity in the fungal kingdom allowed these microorganisms to be successful in the occupation of different ecological niches from rotting plant material in water, to forests, and even to houses made of wood material. But our understanding of lignocellulose deconstruction at the molecular level is still fragmentary. Structure-function relationship for individual enzymes has been elucidated only for a small fraction of those involved in this process. A detailed structural characterization would pave the way for future investigations aimed at understanding the ability of fungi to adjust their enzyme repertoire to the substrate availability in different ecological niches.

13

1.3 Fungal FAD-dependent enzymes

Flavoproteins are ubiquitous flavin-containing enzymes involved in a variety of biologically relevant processes, mainly the catalysis of redox reactions. Flavoenzyme chemistry is remarkably versatile and complex, featuring unique, highly context-dependent mechanistic control that makes generalization between flavoenzymes difficult. Detailed mechanistic studies of flavoprotein oxidoreductases are typically hampered by the complexity of redox chemistry and the inherent intricacy of flavin chemistry.

Most flavoproteins contain a flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as a prosthetic group responsible for the redox catalytic properties. FMN is structurally composed by a riboflavin portion, an isoalloxazine ring, and a ribitol molecule covalently bound through a C-N bond, with the terminal ribitol carbon bound to a phosphate group. FAD results when an adenosine monophosphate is bridged together with FMN through the phosphate group.

The first flavoprotein was identified in 1933 by Warburg and Christian and named "the yellow enzyme"[34]; it is now well known that the flavin portion is responsible for this color in the oxidized resting form. Indeed, the redox and spectroscopic properties of flavoenzymes are due to the presence of the isoalloxazine ring, which can exist in four different oxidation states (Fig. 1.5). To accomplish their function most flavoenzymes exploit the ability of FAD to alternate between the yellow fully oxidized (quinone) and the colorless fully reduced (hydroquinone) state, each accessible through two subsequent electron transfer events. The oxidized form is expected to be planar, while the reduced form has a "butterfly" or bent shape[35]. It is well established that these two forms interchange through a radical intermediate state (semiquinone) deriving

from a single electron transfer. The existence of this intermediate redox state in flavoenzymes was reported in 1956, based on spectroscopic evidences[36]. Flavoproteins are encoded in the genome of both prokaryotic and eukaryotic organisms. Remarkably, a recent study has shown that the filamentous fungus *Neurospora crassa* is able to express about 100 different flavoproteins, more than any other species[37]. Since saprotrophic fungi need access to the plant biomass to survive, this arsenal of flavoproteins is thought to be necessary for the fungus to perform an efficient enzyme "combustion", in which many oxidoreductases cooperate together to convert lignin into CO_2 and H_2O.
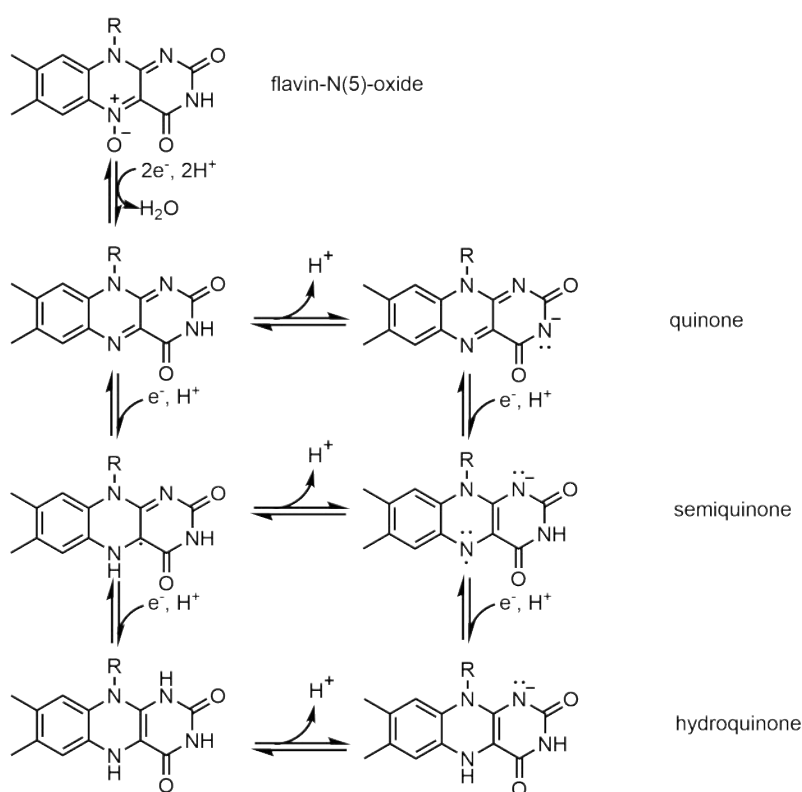

Figure 1.5*. The most common oxidation states of the isoalloxazine ring involved in flavoenzyme catalysis. Each form is accessible through electron transfer events. The forms resulting from acid-base equilibrium are also depicted.*

White-rot fungi represent an important ecological group, being the only organisms to effectively and selectively degrade lignin from wood. This peculiarity is considered of high potential in biotechnological processes, such as the bioconversion of agricultural by-products and the biodegradation of organopollutants, xenobiotics and industrial contaminants[38].

The white-rot fungus *Pycnoporus cinnabarinus* recently stood out as a promising microorganism of choice, mainly as a cell factory for the production of a wide range of enzymes that allow for an efficient and complete degradation of lignin[22]. The *Pycnoporus* genus belongs to the phylum Basidiomycota, class Agaricomycetes, order Polyporales, family Polyporacae, and accounts for four species; *P. cinnabarinus* is a common species widely distributed in the Northern Hemisphere[22]. It is characterized by a bright orange-red color of the basidiocarp (Fig. 1.6), due to its ability to synthesize pigments including cinnabarin and cinnabarinic acid. The genome sequencing of this organism has provided an overall view of the enzyme repertoire employed for lignin degradation, showing an outstanding abundance of carbohydrate-active enzymes[22].

Figure 1.6. The white-rot fungus Pycnoporus cinnabarinus growing on a wood surface.

16

This type of enzymes were originally grouped in the CAZy (Carbohydrate Active enZymes) database[39], which included glycoside hydrolases, glycosyltransferases and other families classified on the basis of sequence similarity. Recently, the CAZy database has been expanded and now includes lignin degrading proteins grouped in the "Auxiliary Activities" (AA) class[33, 40]. Like all enzyme classes in CAZy, the AA families were classified on the basis of a common structural or functional module depending on the available biochemical characterization.

The AA3 family accounts for the GMC oxidoreductases and can be divided into four subfamilies[33]; GOX and GDH belong to the AA3_2 subfamily, together with aryl-alcohol oxidase (AAO), aryl-alcohol dehydrogenase (AAD) and pyranose dehydrogenase (PDH). Originally GDHs were not considered in the AA3 classification[40], but a recent detailed phylogenetic analysis demonstrated the close evolutionary relationship with GOX. Four clades were identified, meaning that four different enzyme groups can be recognized as lineal descendants from a common ancestor: GOX and GDH class I, II and III[3, 33]. The GOX clade comprises all the GOXs characterized to date, from ascomycete organisms. GDH classes I and II contain ascomycete sequences as well; most of the currently characterized members belong to class I, while completely uncharacterized proteins belong to class II. GDH class III includes basiodiomycete sequences and just one member is characterized to date: GDH from *P. cinnabarinus* (*Pc*GDH)[17].

The close phylogenetic relationship between GOX and GDH suggests the appearance of a high degree of specialization from an early point in the evolution of these two enzymes. The classification of AA3_2 enzymes shows a marked multigenicity, more than any other AA3 subfamily. Multigenicity is a common feature in fungus-derived enzyme systems, especially in wood- and litter-degrading fungi, since they have to adapt to diverse ecological niches and

17

then deal with harmful compounds originating from organic matter degradation, secondary metabolism of antagonists or human activities[41]. This multigenicity is manifested by the presence of several paralogous genes in an organism as a result of gene duplication events, with the purpose of functional compensation upon eventual deficiencies of one of its members[42]. Moreover, gene duplication is followed by gene diversification, one of the most important mechanism through which nature designs enzyme isoforms with new functionalities[43].

The existence of multigenic AA family members has been explained in terms of functional redundancy, functional diversification including subtle differences in substrate specificity of individual enzymes, or fine tuning of the gene regulation and expression in order to adapt fungal organisms to the diversity of their natural substrates[40]. Plant cell walls are composite structures, and the multigenicity in the AA3 subfamilies might reflect necessities and adaptations to the degradation of these complex substrates. A recent analysis estimates the fungal species to be between 2.2 and 3.8 million[44], with only 3-6% being identified to date; this means that we are only looking at a fraction of the genetic diversity in fungi and there are still many more aspects of their biochemical properties to be discovered.

1.4 Glucose dehydrogenase as a glucose biosensor

Diabetes mellitus is a serious modern disease worldwide, detected in 9.6% of adults in the Unites States between 2003 and 2006[45]. Self-monitoring of blood glucose is an important component of modern therapy. It helps individuals to adjust their dietary intake, physical activity, and insulin doses, with the goal of improving glycemic control on a day-to-day basis. Most commercially available blood glucose monitor sensor strips are electrochemical biosensors consisting of an electrode and a glucose oxidizing enzyme. Three generations of biosensors have been developed, differing in the electron transfer processes they are based on (Fig. 1.7).

First-generation systems employed oxygen as the electron acceptor, determining glucose concentration by following either the consumption of oxygen or the liberation of hydrogen peroxide. In second-generation sensors,
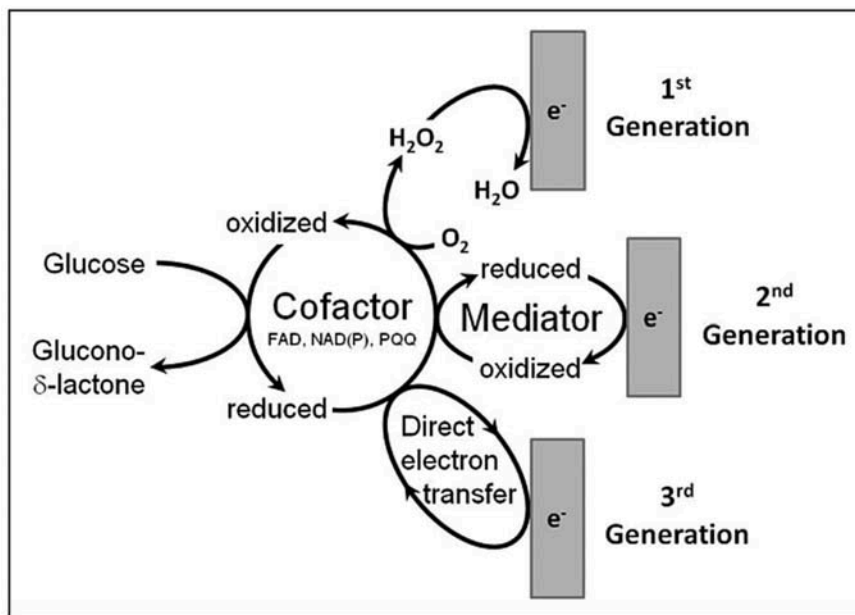

Figure 1.2. Schematic representation of the electrochemistry involved in blood glucose monitoring systems. The three generations differ in the redox mediator that transfer electrons to the electrode[1].

19

enzymes transfer electrons to artificial electron acceptors, also referred to as electron mediators or redox dyes, instead of oxygen to avoid interference from other redox species. The reduced acceptors are monitored colorimetrically or electrochemically. The third-generation sensor employs enzymes with the ability to transfer electrons directly to the electrode, without the need for mediators or oxygen; these enzymes are called direct electron transfer (DET) type enzymes. The latter type of system is considered the most promising since it does not require the use of artificial electron mediators and errors due to variations in the concentration of oxygen in blood samples are eliminated. Considering that redox mediators are usually toxic, a mediator-free system is ideal for implantable devices that continuously monitor glucose concentration by means of a sensor implanted subcutaneously.

Glucose oxidoreductases capable of DET, however, are limited at the moment. Currently, the most relevant biotechnological application of GDH is its use in second-generation systems, which are the most popular commercially available sensors. In particular, fungal FAD-dependent GDHs show many advantages compared to GOX and other GDHs. They are insensitive to molecular oxygen, therefore variations in the oxygen concentration in blood, which might lead to an underestimation of the glucose level, do not affect the final measurement. They also show a narrow substrate specificity, especially the lack of activity toward maltose, which is contained in the infusion solution for diabetes[46]. There are some difficulties, however, related to the coupling of the enzyme component with the electrochemically active electron acceptors that still need to be overcome.

In second-generation biosensors a redox mediator replaces oxygen in carrying electrons from the enzyme to the surface of the working electrode; the reduced mediator is then re-oxidized at the electrode, providing an amperometric signal and regenerating the oxidized form of the mediator. The most common

electron mediators employed so far are ferrocene, quinines and ferricyanide: they do not react with oxygen, are stable in both the oxidized and reduced forms regardless of pH, show reversible electron transfer kinetics, and react rapidly with the enzyme[47]. A common problem found in blood glucose monitoring systems is the interference of other substances present in blood that are electroactive at the operating potential of the amperometric sensor, such as ascorbic acid, cholesterol and uric acid. This type of interference can be avoided by applying an electrode potential as low as possible.

An artificial acceptor that has been recently gained attention for this purpose is hexaammineruthenium (III) chloride; in this compound the redox couple Ru (II/III) has a lower redox potential compared to the ferro-ferricyanide redox couple. Fungal FAD-dependent GDHs, however, are not able to utilize hexaamineruthenium as electron acceptor, while GOXs can be coupled to this acceptor for commercial blood glucose monitoring sensors. The comparison of the available 3D structures of fungal GDHs and GOXs allowed for the identification of the amino acid residues responsible for this different behavior; GDH from *Aspergillus flavus* (*Af*GDH) has been mutated to successfully construct a GDH able to utilize hexaammineruthenium as an electron acceptor[48].

Although most of the GDH-based blood glucose sensors are second generation systems, some examples belonging to the third generation were also developed. Enzymes can obtain quasi-DET ability when a mediator is bound to their surface; this way electrons are transferred from the reduced cofactor to the mediator and then from the mediator to the electrode. A quasi-DET biosensor has been produced by coupling GDH with an artificial mediator, N-hydroxysuccinimidyl ester 1-propoxy-5-ethylphenazinium ethyl sulfate (arPES), that spontaneously binds to amine groups of the enzyme[49].

21

Unfortunately, few naturally occurring enzymes are capable of DET-reaction with electrodes. Among these, fungal flavocytochrome CDHs turned out to efficiently perform DET. They are composed of two domains (Fig. 1.4): a heme *b*-binding cytochrome domain (CYT) and a FAD-binding dehydrogenase catalytic domain (DH), which are connected by a flexible linker that allows for flavin-to-heme interdomain electron transfer.

A fungal FAD-dependent GDH able to perform DET was successfully produced by linking the catalytic domain *Af*GDH to the *P. chrisosporium* CDH-derived heme *b*-binding cytochrome domain. This construct showed efficient flavin-to-heme interdomain electron transfer and maintained the original substrate specificity of fungal GDH, thus combining the two main advantages of the single enzymes to produce a powerful tool for continuous blood glucose monitoring[50].

1.5 Structural features of GMC oxidoreductases

The members of the GMC (glucose-methanol-choline) superfamily share a common fold and carry a covalently or non-covalently bound FAD cofactor. These enzymes are typically composed of two major structural domains: a FAD-binding domain and a substrate-binding domain. The flavin-binding domain is highly conserved in all members of the GMC family and shows the canonical βαβ dinucleotide-binding motif. The sequence and structure of the substrate-binding domain of GMC oxidoreductases are less preserved, which reflects the diversity in substrate specificity within this family. The overall structure can be described following the nomenclature first introduced for GOX from *Aspergillus niger* (*An*GOX)[51].

The FAD-binding domain consists of a central five-stranded parallel β-sheet (sheet A) surrounded by three α-helices and an additional crossover connection composed of a three-stranded antiparallel β-sheet (sheet B). One of those three helices is part of the βαβ-motif interacting with the ADP moiety of FAD. In AAO the sheet B is connected to another β-sheet motif of two antiparallel β-strands (sheet B')[52]; this feature is typical of this enzyme only. GOXs[53] instead display a long N-terminal tail wrapping around the bottom part of the FAD-binding domain that is absent[52, 54], shorter[55, 56] or too flexible to be structurally defined[57] in other GMC oxidoreductases (N-T in Fig. 1.8).

In most cases the cofactor is non-covalently bound to the apoenzyme and occupies a narrow channel. In choline oxidase (CHOX)[55], pyranose 2-oxidase (P2O)[57] and PDH[56], however, the FAD is covalently attached to the protein by an 8α-(N3)-histidyl bond: the Nε2 atom of a histidine residue is linked to the FAD C8M atom of the isoalloxazine ring. Regardless of the binding mode, the FAD molecule adopts an extended conformation, with the adenine and the

23

isoalloxazine moieties distal to each other. The most prominent interactions between the protein and the cofactor consist in a network of hydrogen bonds, predominantly involving the NH and the CO groups of the main-chain of residues located at the N-terminus and several bridging water molecules. The principal binding sites of FAD are on the ADP part of the molecule; in particular, the two phosphate groups proved to be essential for efficient binding of the cofactor to the apoenzyme in reconstitution experiments[58]. Moreover, the helix forming the βαβ-motif is thought to stabilize the negative charge on the phosphate groups through its helix dipole effect[59].
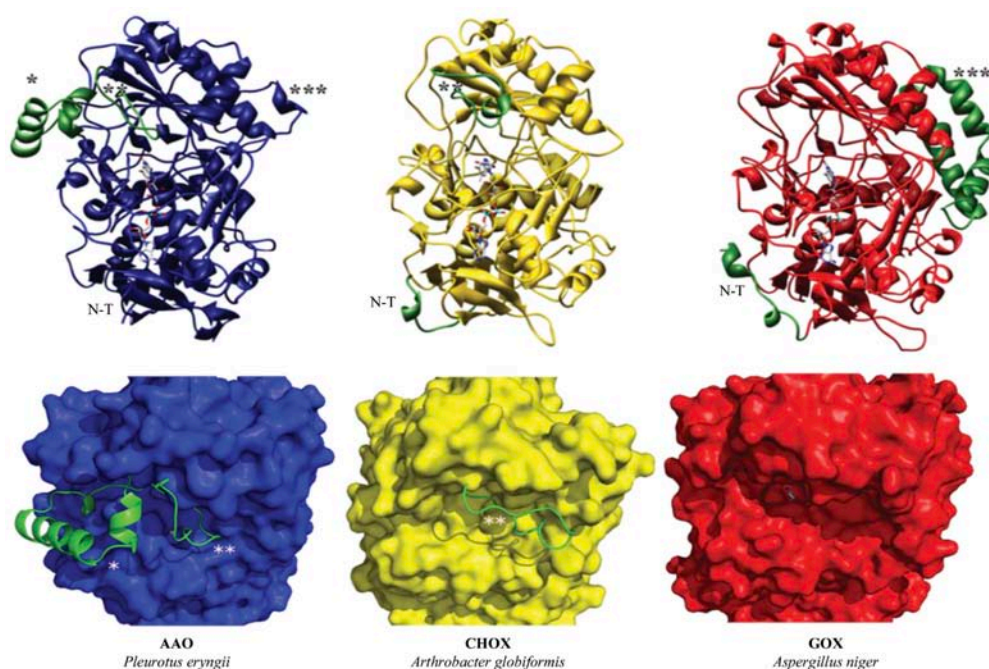

Figure 1.3*. Structural comparison of ligand-free AAO (blue), CHOX (yellow) and GOX (red) in the same orientation. Top row shows ribbons representation: structural elements typical of each subfamily are highlighted in green using stars; the N-terminal segment has different length in the three enzymes (N-T); the helical insertions in the substrate-binding domain are marked as *** (away from the active site); structural elements at the entrance of the active site are marked as * or **. Bottom row shows colored molecular surfaces, highlighting the differences in active site accessibility (green ribbons)[52].*

A consensus N-terminal signature Gly-X-Gly-X-X-Gly-(X)$_{18}$-Glu is part of the central core of the FAD-binding domain, with most of these residues located in the loop connecting the first β-strand and the α-helix. The carboxyl group of the conserved glutamate residue is involved in strong hydrogen bonds with the ribose moiety of FAD and with the backbone nitrogen of a glycine residue forming the βαβ-motif. The binding of ribose to the side chain carboxyl group of an acidic residue is observed in most nucleotide-binding proteins, generally with Asp for NAD- and Glu for FAD-binding proteins, except for some P2Os that bind the ribose moiety of FAD through an Asp residue[57]. The cofactor is deeply buried in the apoenzyme, with direct access to bulk solvent only limited to the part of the isoalloxazine ring containing the N5 atom directly involved in the catalysis.

As mentioned before, structural differences among members of the GMC oxidoreductase family are more pronounced in the substrate-binding domain, although all enzymes show a six-stranded antiparallel β-sheet (sheet C) flanked by two long α-helices in this region. The number and the position of helices surrounding the sheet C varies according to the member of the family: GOX and GDH possess a 50 amino acid insertion, folding into three more helices, that lacks in CHOX and that is reduced to 13 residues and a unique helix in AAO (*** in Fig. 1.8)[52]. On the other hand, a 14-17 amino acid insertion in the loop connecting two antiparallel β-strands forming the sheet C is present in AAO and CHOX while is absent in GOX and GDH (** in Fig. 1.8). Another 30-residue insertion present only in AAO folds as a group of three helices protruding from the central core of sheet C (* in Fig. 1.8).

Together, these two insertions generate a platform above the active site of AAO that notably restricts access to the isoalloxazine ring of FAD[52]. P2O, PDH and the dehydrogenase domain of CDH show a similar architecture in

25

which the active site is shielded by rigid structural components[56, 57, 60]. GDH and GOX instead show a large, deep pocket shaped like a funnel that does not show any structural element hampering the substrate accessibility to the flavin system[53, 54].

The catalytic site of GMC oxidoreductases is located at the bottom of a cavity in which the isoalloxazine moiety of FAD points toward the aqueous environment. Each member of the superfamily is characterized by a unique geometry of the active site that is responsible for their catalytic properties. The protein environment surrounding the flavin moiety exerts a strong modulation on the cofactor redox properties and on the enzyme affinity for electron donors and acceptors. As a consequence, small changes in the proximity of the cofactor can affect a range of parameters that are difficult to predict, monitor and assess. Given the complexity associated with the chemistry of flavoenzymes, it is also generally difficult to obtain stable and ordered crystal structures of oxidoreductase-ligand complexes, offering relatively few well-characterized binding modes.

The residues responsible for catalysis have previously been reported in a variety of GMC oxidoreductases. A highly conserved His/His or His/Asn pair in the proximity of the isoalloxazine ring is expected to function as a catalytic pair on the basis of crystal structures, site-directed mutagenesis, pH-dependence studies, or theoretical calculations[61]. The His/His pair is found in GDH, GOX, PDH and AAO, while CHOX, CDH and P2O use a His/Asn constellation. Both residues are localized on the *re*-face of the flavin portion: the universally conserved His lies on top of the isoalloxazine ring, while the His/Asn residues are in a lateral position, close to the pyrimidine portion of the isoalloxazine ring (Fig. 1.9). Even if these two residues proved to be essential for catalysis, their specific functional role is still a matter of debate. Since the principal His is conserved in all members of the superfamily, it was originally

proposed to act as the catalytic base responsible for proton abstraction in the substrate oxidation half-reaction (Fig. 1.9). This hypothesis is in agreement with studies carried out on most GMC oxidoredutases[61-64], with the remarkable exception of cholesterol oxidase, in which the sub-atomic resolution crystal structures show protonation of the catalytic His Nε2 atom in the pH range 4.5-7.3; the authors propose that the principal His acts instead as a hydrogen bond donor group assisting in optimizing substrate geometry and orbital overlapping[65, 66]. This residue has been pointed out as critical for oxygen activation in GOX[67], but this is not likely to be the case for other GMC oxidoreductases, such as GDH, CDH and PDH, that show no or poor ability to activate oxygen.
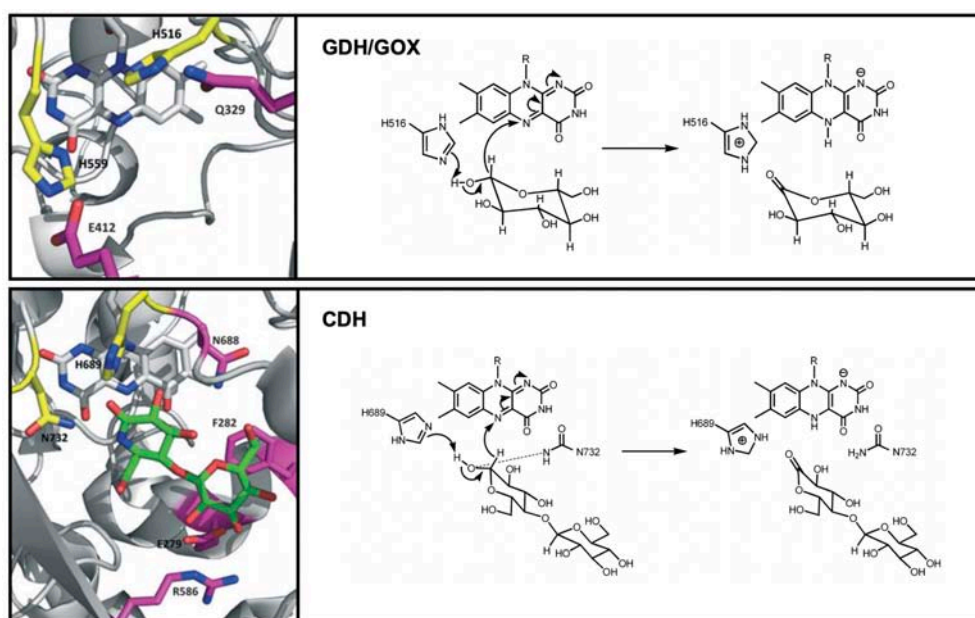

Figure 1.4*. Top panel: active site geometry in GOX (PDB entry 1GAL[51]) and proposed catalytic mechanism for the substrate oxidation half-reaction, valid also for GDH; Bottom panel: active site geometry in cellobiolactam-bound CDH (PDB entry 1NAA[64]) and proposed catalytic mechanism. In both structures the C atoms of the His/His (His516/His559 in GOX) and His/Asn (His689/Asn732 in CDH) conserved residue are in yellow. In both cases the conserved His is thought to act as the catalytic base[61].*

27

The structural determinants responsible for the oxidase or dehydrogenase activity in these enzymes have not been univocally identified yet. The structural characterization of PDH by Tan et al. suggested to take into account the distance between the side-chain nitrogen atoms of the catalytic pair as a possible discriminating factor[56]. The structural analysis carried out on *Af*GDH allowed the authors to identify three residues in the proximity of the active site able to act as molecular switches that shift the enzyme activity from oxidase to dehydrogenase[54].

The geometry of the isoalloxazine ring is strongly enzyme-dependent. The structure of this portion can be difficult to rationalize because of the high reactivity of FAD. Most of the available GMC oxidoreductase structures are modelled with an oxidized FAD, although distortions from the typical planar geometry have been observed. In the first GOX structure solved, Hecht at al. suggested that the Asn residue located at the *si*-face in the middle of the flavin system provides a backstop preventing a planar conformation of the ring; this hypothesis is corroborated by the low B-factors observed for this residue[51]. Deviations from planarity have been associated with redox potential and substrate binding[68, 69] but can also be a consequence of FAD reduction by radiation-generated electrons[70].

Interesting cases of particular flavin oxidation state are represented by CHOX and PDH, since a flavin C(4a)-oxygen adduct is observed in the crystallographic structures. The structural analysis of CHOX did not allow a distinction between a hydroxyflavin C(4a)-OH or a peroxyflavin C(4a)-OOH adduct; nevertheless, these data have been interpreted as the first direct visualization of a biochemically relevant intermediate involved in flavoenzyme catalysis[71]. This interpretation cannot be applied to the case of PDH, since this enzyme shows no physiological oxygen reactivity; most likely it represents an artifact arising from attack by reactive oxygen species on the

C(4a) position during irradiation of the protein crystal[56]. Conversely, a C(4a)-hydroperoxyflavin intermediate has been detected during the P2O catalytic cycle by means of absorption spectroscopy in rapid kinetics, but no structural evidence has been reported[72].

Crystal structures of ligand-bound GMC oxidoreductases have been solved for GDH, CDH, AAO and P2O. In the cases of GDH and AAO, the final products of the oxidative half-reaction have been employed for structure determination, using soaking and co-crystallization respectively[54, 73]. Ligand-bound P2O structures have been obtained using fluorinated analogs of glucose[62, 74] while CDH structure has been solved in complex with the inhibitor cellobionolactam[60, 64]. In GDH and in CDH, comparison between the ligand-free and the ligand-bound structures show minor differences in the active site, with a single displacement involving a Glu or an Arg side chain respectively upon ligand binding. AAO displays a highly pre-organized active site with almost no conformational changes between the ligand-free and ligand-bound structures. P2O is the only member of this family showing a significant conformational rearrangement that involves a substrate recognition loop acting with a "switch-and-lock" mechanism[74]. The structure of *Af*GDH in complex with D-glucono-δ-lactone explains the lack of substrate specificity observed in other GDHs, *i.e.* their residual activity toward other sugars (see par. 1.1). The absence of residues that recognize the sixth hydroxyl group of glucose in the active site may account for this enzyme activity toward xylose[54].

The quaternary structure of fungal GMC oxidoreductases ranges from monomeric to homo-oligomeric. Most of these enzymes are extensively glycosylated, with glycans contributing up to 65-70% of the molecular weight[16]. Native fully glycosylated proteins can be expressed in eukaryotic fungal host, not in bacterial cell cultures. Despite the lack of glycosylation,

29

functional enzymes have been recombinantly expressed in bacterial host[16, 18-20, 75]. Glycosylations are thought to play an essential role in thermal stabilization and in the folding process[76], since some of these proteins require *in vitro* refolding from inclusion bodies when produced in *E. coli*[18, 75]. In terms of electrochemical behavior, deglycosylation increases the efficiency of direct electron transfer ability, presumably by decreasing the distance between the enzyme's redox center and the electrode, which are believed to be too distant in the native glycosylated enzyme[77].

Chapter 2

Aims of the thesis

This PhD thesis focuses on the biochemical characterization of a redox enzyme likely to play an essential role in the vital cycle of the white-rot fungus *Pycnoporus cinnabarinus*: glucose dehydrogenase (GDH). In response to the structural complexity and heterogeneity of the different plant cell wall polymers, saprotrophic fungi produce a complex arsenal of enzymes to gain access to the carbon source. White-rot fungi, like *P. cinnabarinus*, developed the ability to rapidly and completely break down hemicellulose and lignin, causing the appearance of white cellulose leftovers on the invaded wood. This outcome is caused by the enzyme combustion performed by a cocktail of secreted redox proteins whose mechanism of action remains unclear.

Auxiliary enzymes, like GDH, prevent the re-polimerization of phenoxy radicals deriving from the action of lignin modifying enzymes; their reducing ability is fueled by the electron donor properties of carbohydrates deriving from the polysaccharide constituents of the plant cell wall. A preliminary biochemical characterization of *Pc*GDH was carried out in terms of substrate specificity and catalytic properties, but no structural information is currently available for this enzyme and many aspects of its function still lack a molecular description and clarification. To date, only one fungal GDH (from *Aspergillus flavus*) has been structurally characterized, which is 36.4% identical to *Pc*GDH in terms of primary structure. Notably, *A. flavus* and *P. cinnabarinus* differ significantly in terms of morphology, hosts, and growth conditions. The structural characterization coupled with the functional analysis of *Pc*GDH would pave the way for a deeper understanding of the molecular mechanisms at the base of the unique properties of this organism. Such information would

31

allow us to design of an engineered dehydrogenase highly specific and efficient toward glucose with negligible residual activity toward alternative sugars, thus ideal for usage in blood glucose monitoring sensor strips. Within the biorefinery concept, all the biomass components, including cellulose, hemicellulose and lignin, could be exploited in biochemical processes designed to maximize economic and environmental benefits while minimizing waste. Lignin, which is the second most abundant biopolymer on earth after cellulose, is by far the less valorized component. Filamentous fungi, like *P. cinnabarinus*, represent a source of novel enzymes displaying useful properties for modifying lignin and aromatic compounds. This potential is indeed currently underutilized due to insufficient knowledge about the enzymatic mechanisms involved. A detailed biochemical characterization of lignin-modifying and lignin-degrading auxiliary enzymes would lead to new and improved use of those systems in industrial biotechnologies, while expanding the knowledge on natural biocatalysts for lignin and aromatic utilization.

In the following chapters of the thesis I describe the experimental procedures and the results obtained in the characterization of *Pc*GDH by means of X-ray crystallography and spectrophotometric enzymologic methods. A special focus is put on the three reported crystallographic structures that I determined, since they turned out to be crucial in the formulation and the subsequent confirmation of a hypothesis on the physiological function of this enzyme.

Chapter 3

Materials and methods

3.1 Cloning and expression of *Pc*GDH

*Pc*GDH was recombinantly expressed and purified as described in Piumi et al.[17] at the Aix-Marseille University thanks to a collaboration with the group of Dr. Giuliano Sciara, Department of "Biodiversité et Biotechnologie Fongique" (INRA UMR1163).

The gene sequence encoding *Pc*GDH is available at the European Nucleotide Archive (ENA), EMBL-EBI (accession number [EMBL: PRJEB5237, scf184803.g17]. The open reading frame (ORF) was predicted from the genomic sequence by translating it in the six frames and detecting splicing sites following the general pattern GTRMGT...YAG. This prediction was corroborated by BLAST using the *P. cinnabarinus* cDNA library[22] with the Augustus gene prediction website (http://bioinf.uni-greifswald.de/augustus/). The ORF sequence was aligned against other organisms to check the splicing prediction (tblastx on NCBI). The synthesized cDNA was cloned in the pAN52.4 (EMBL accession: Z32699) expression vector using a restriction cloning approach with BssHII and HindIII enzymes. The 22 amino acids of the GDH signal peptide (MPHSRLLATIGAVALLARCVTA) were replaced by the 24 amino acid glucoamylase (GLA) prepro sequence from *A. niger* (MGFRSLLALSGLVCNGLANVISKR). Additional segments were incorporated: one encoding a His-Tag (CACCAT)×3 that was integrated downstream and two restriction sites (MluI and HindIII) at the 5′ and 3′ ends, respectively, for subcloning in the expression vector. After codon optimization for *A. niger*, the predicted cDNA sequence (accession number KJ934222) with

all its additional fragments was synthesized (GeneArt, Carlsbad, USA), propagated in *E. coli* JM109 (Promega, Charbonnières, France), and sequence-checked (GATC Biotech, Konstanz, Germany).

In the final expression cassette, the *Aspergillus nidulans* glyceraldehyde-3-phosphate dehydrogenase encoding gene (*gpdA*) promoter, the 5′ untranslated region of the *gpdA* mRNA and the *A. nidulans trpC* terminator were used to drive expression of the inserted encoding sequence. *E. coli* JM109 strain was used for vector construction and propagation. *A. niger* strain D15#26, orotidine-5'-phosphate decarboxylase deficient mutant (*pyrG-*), was used for heterologous expression of *Pc*GDH: the *Pc*GDH expression vector and the pAB4-1 vector containing the *pyrG* selection marker[78] were co-transformed in a 10:1 ratio, as described by Punt and van den Hondel[79].

Transformants of *A. niger* were selected for uridine prototrophy for growth on selective solid minimum medium (without uridine) containing 70 mM $NaNO_3$, 7 mM KCl, 11 mM KH_2HPO_4, 2 mM $MgSO_4$, and 1% (w/v) glucose and trace elements (1000x stock; 76 mM $ZnSO_4$, 178 mM H_3BO_3, 25 mM $MnCl_2$, 18 mM $FeSO_4$, 7.1 mM $CoCl_2$, 6.4 mM $CuSO_4$, 6.2 mM Na_2MoO_4, and 174 mM EDTA). Recombinant protein expression in suspension culture was carried out inoculating 2×10^6 spores mL^{-1} in 3 L of liquid minimum culture medium containing 70 mM $NaNO_3$, 7 mM KCl, 200 mM Na_2HPO_4, 2 mM $MgSO_4$, 5% (w/v) glucose and trace elements; pH was daily adjusted to 5.5 with 1 M citric acid. Extracellular *Pc*GDH production was carried out at 30 °C for 10 days in a shaker incubator at 120 rpm.

3.2 Purification of recombinant *Pc*GDH

After 10 days of growth, 3 L of culture were harvested clarified on GF/D and GF/F glass fiber filters (Whatman, Maidstone, UK), sterile-filtered (0.45 μm),

and concentrated ten-fold by ultrafiltration through a polyether-sulfone membrane with a 10 kDa molecular mass cutoff (Vivaflow crossflow cassette, Sartorius, Les Ulis, France). The retentate was adjusted to pH 7.8 and the His-tagged recombinant protein was purified on a Chelating Sepharose Fast Flow column (5 mL Ni Histrap, GE Healthcare, Velizy-Villacoublay, France) equilibrated with five column volumes of binding buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM imidazole, pH=7.8). After washing in five column volumes with binding buffer, bound proteins were eluted with six column volumes of 150 mM imidazole in binding buffer at a flow rate of 5 mL min^{-1} and collected in 1 mL fractions.

Protein purity was assessed using 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The protein bands were stained with Coomassie Brilliant Blue R-250 and molecular mass was determined against reference proteins (pre-stained protein ladder, Euromedex, Souffelweyersheim, France). *Pc*GDH protein band was identified by Western Blot analysis. Electrophoresed proteins were electroblotted onto polyvinylidene difluoride membranes (iBlot, Life Technologies, Saint-Aubin France). Membranes were then incubated in tris-borate saline (TBS) blocking solution (10 mM Tris-HCl, 150 mM NaCl, 5% w/v milk powder, pH=7.4) overnight at 4 °C, washed with TBS (10 mM Tris-HCl, 150 mM NaCl, pH=7.4) and with TBS Tween (10 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20, pH=7.4), and finally treated with blocking solution containing monoclonal anti-polyhistidine antibody serum at a 1:2000 dilution (Sigma-Aldrich, Saint-Quentin-Fallavier, France). Signal detection used the 5-bromo-4-chloro-3-indolyl phosphate-nitro blue tetrazolium assay (Roche Applied Science, Meylan, France) performed according to the manufacturer's procedure.

35

The molecular weight of recombinant glycosylated *Pc*GDH used in calculations is 70 kDa as estimated by the SDS-PAGE performed on the native glycosylated protein.

3.3 Crystallization and crystal handling

Purified *Pc*GDH was exchanged in buffer with no imidazole and concentrated to 17.4 mg mL^{-1} for crystallization trials. Crystallization plates were initially set up automatically using a Crystal Phenix robot (Art Robbins Instruments, Sunnyvale, CA, USA) in sitting-drop vapor-diffusion mode at 294 K using Intelli-plate 96 well (Art Robbins Instruments, Sunnyvale, CA, USA). Single crystals appeared using the AmSO$_4$ suite (Qiagen, Hilden, Germany) crystallization screen. The reproducibility of crystallization conditions was tested setting up hand-made plates using the hanging-drop vapor diffusion method at 294 K, which also allowed production of larger crystals, shown in Figure 3.1.
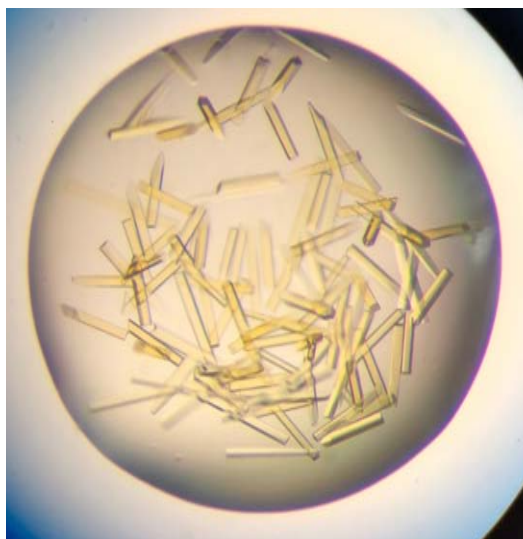

Figure 3.5. Crystals of ligand-free PcGDH. The yellow color is due to the oxidized FAD cofactor.

36

Well-diffracting single crystals grew by mixing 1 µL of protein solution and 1 µL of reservoir solution containing 2 M $(NH_4)_2SO_4$ and equilibrating the droplet against 0.5 mL of reservoir solution. For cryoprotection, ligand-free *Pc*GDH crystals were transferred in a solution containing 2 M $(NH_4)_2SO_4$ and 2 M $LiSO_4$, then flash-frozen in liquid nitrogen.

*Pc*GDH-GLC crystals were prepared by soaking the ligand-free *Pc*GDH crystals in a solution containing 2 M $(NH_4)_2SO_4$ and 30% D-glucose (Carlo Erba, Milan, Italy); after about 12 min incubation, crystals were transferred in 2 M $(NH_4)_2SO_4$, 30% D-glucose and 25% glycerol for cryoprotection, then flash-frozen in liquid nitrogen.

Analogously, *Pc*GDH-LMB crystals were obtained through soaking for 20 min in a solution containing 2 M $(NH_4)_2SO_4$ and 80% laminaribiose (Megazyme, Bray, Ireland), which also acts as cryoprotectant at such a high concentration. After flash-freezing in liquid nitrogen, crystals were used for data collection at synchrotron radiation sources.

3.4 Structure determination and refinement

X-ray diffraction data of ligand-free *Pc*GDH and *Pc*GDH-LMB were collected at the Diamond synchrotron (Harwell, UK), beamline I24; data for *Pc*GDH-GLC were collected at ELETTRA (Trieste, Italy), beamline XRD2. All datasets were collected at 100 K using a PILATUS detector. Data were indexed, scaled and integrated using the XDS package[80].

Molecular replacement was carried out using MOLREP[81] from the CCP4 suite[82]. Given the 36.4% sequence identity, the structure of GDH from *Aspergillus flavus* at 1.78 Å resolution[54] (PDB entry 4YNT) was used as the search model to calculate the initial crystallographic phases of ligand-free

*Pc*GDH, whose structure was then employed to obtain the initial phases of both *Pc*GDH-GLC and *Pc*GDH-LMB. Iterative automated and manual structure refinement was carried out using REFMAC5[83] and COOT[84], both implemented in the CCP4 suite. TLS cycles were included during the automated refinement[85]. 5% of the reflections were excluded from refinement for cross validation by means of free R-factor[86]. Manual adjustment was performed with the F_o-F_c map contoured at 3σ and the $2F_o$-F_c map at 1σ.

Validation of the models, including Ramachandran statistics and B-factor analysis, was carried out using the Multimetric model geometry validation tool implemented in the CCP4 suite. Protein sequence alignment was performed using ClustalW[87]; structural superposition was carried out using Superpose[88] implemented in the CCP4 suite. Figures were produced using Chimera[89].

3.5 Spectrophotometric assays for enzymatic activity

Functional assays have been carried out with the standard DCIP method: in every experiment DCIP has been used as the secondary electron acceptor in the half reaction responsible for FAD oxidation. The redox cycle is shown in Figure 3.2. Enzymatic activity was evaluated spectrophotometrically monitoring the variation in absorbance at 520 nm (ΔA_{520}), at which the pH dependence of the spectral properties of DCIP is least pronounced; the extinction coefficient of DCIP at 520 nm employed in calculations was ε_{520}=6800 M^{-1} cm^{-1}.

Pilot experiments monitoring the pH-dependence of the *Pc*GDH enzymatic activity revealed the optimal pH to fall in the range between 5.5 and 6 (data not shown), which is consistent with what has been previously observed[17]. For this reason, all the experiments herein reported were performed at pH 5.5.
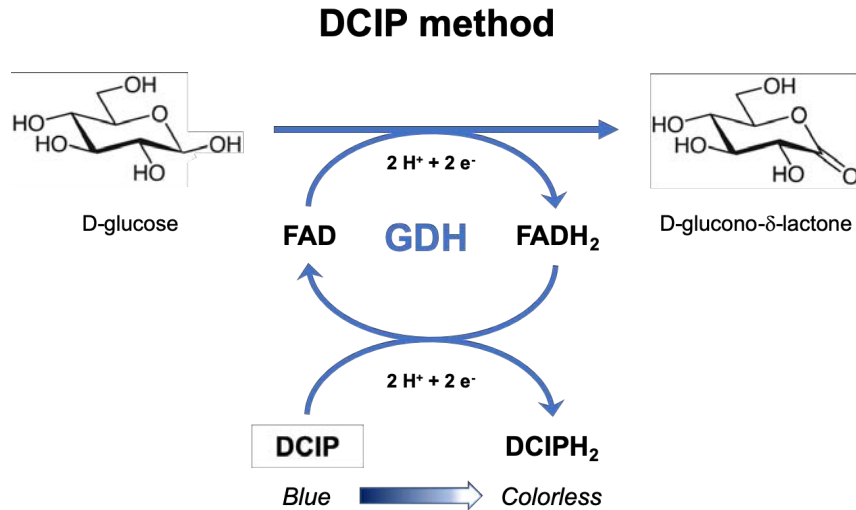
DCIP method

Figure 3.2. Schematic representation of the chemical principle of the DCIP standard method. The FAD cofactor in GDH is reduced by the sugar substrate (D-glucose in the example reported in the picture), which is converted to the C1-oxidized lactone (D-glucono-δ-lactone). Once reduced, the cofactor transfers electrons to DCIP, that is blue in the oxidized state and colorless in the reduced state. GDH activity can then be assessed by monitoring the decrease in absorbance of the oxidized DCIP acceptor.

Substrate specificity was assessed screening five different sugar substrates: D-glucose, cellobiose, laminaribiose, 1,3;1,4 β-glucotriose A (GTA) and 1,3;1,4 β-glucotriose B (GTB) (Megazyme, Bray, Ireland). All experiments were carried out using 39.7 nM *Pc*GDH, 0.4 mM DCIP, 50 mM citrate-phosphate buffer, pH=5.5; all substrates were employed at a concentration of 44 mM. The reaction was allowed to proceed for 60 s at 303 K.

Kinetic constants were determined only for D-glucose and laminaribiose. In the case of glucose, A_{520} was measured at time intervals of 15 s for 105 s total; this time range was chosen because DCIP concentration (A_{520}) linearly decreases over time in the first 105 s. Each measurement was taken in triplicate. Substrate specificity and kinetics constants were determined in a sample final volume of 0.1 mL in a 0.2 mL 96-well plate (Corning Costar, Corning, NY, USA) using a Multiskan GO microplate spectrophotomer (Thermo Scientific, Waltham, MA, USA). Solutions without enzyme were first

39

equilibrated at 303 K in the 96-well plate; upon addition of the enzyme, solutions were automatically mixed and ΔA_{520} was recorded. Initial velocities were measured at various substrate concentrations, and data were fit using Kaleidagraph software package by following the hyperbolic equation of Michaelis-Menten:

$$V = \frac{V_{max}[S]}{K_M + [S]}$$

where V is the initial velocity of the reaction, V_{max} is the maximum velocity of the reaction corresponding to the substrate saturation of the enzyme, K_M is the Michaelis-Menten constant and $[S]$ is the substrate analytical concentration.

Chapter 4

Results and discussion

4.1 Structural features of ligand-free *Pc*GDH

4.1.1 *Overall structure*

Crystals of *Pc*GDH are bright yellow, suggesting that the cofactor is in the oxidized resting state (Fig. 3.1). They belong to the P $2_12_12_1$ space group and contain one protein molecule per asymmetric unit. The calculated Matthews coefficient is 2.35, corresponding to a solvent content of 47.7%. Non-crystallographic symmetries are not observed, which is consistent with *Pc*GDH being a monomer. The highest resolution observed for crystal diffraction was 1.43 Å, but the final model of ligand-free *Pc*GDH is refined to 1.5 Å resolution. Ramachandran analysis indicates that most residues (94.89%) fall in the favored region, 30 residues (5.11%) are in the allowed region and there are no outliers. A total of 5 proline residues (258, 348, 422, 482, 493) are in *cis* configuration. Data collection and refinement statistics are summarized in Table 4.1. Secondary structure assignments and *Pc*GDH sequence alignment with *An*GOX and *Af*GDH are shown in Figure 4.1.

The overall structure of *Pc*GDH consists of 20 α-helices and 18 β-strands organized in two distinct domains, the FAD-binding domain and the substrate-binding domain (Fig. 4.2). The β-strands are arranged in 5 β-sheets named following the nomenclature introduced for *An*GOX.

Table 4.1. *Data collection and refinement statistics.*
Values in parentheses are for the highest-resolution shell.

	Ligand-free *Pc*GDH	*Pc*GDH-GLC	*Pc*GDH-LMB
Data collection			
Space group	P2$_1$2$_1$2$_1$		
Unit-cell dimensions (Å)	a=48.87 b=61.59 c=195.09		
Resolution range (Å)	97.55-1.43 (1.57-1.43)	97.93-1.77 (1.87-1.77)	97.15-1.57 (1.75-1.57)
Number of observations	1391283 (295026)	721028 (106450)	1071157 (300980)
Unique reflections	109762 (25836)	112115 (18083)	82347 (22997)
Completeness (%)	99.9 (99.7)	99.9 (99.3)	99.8 (99.3)
Redundancy	12.7 (11.4)	6.4 (5.9)	13.0 (13.0)
I/σ(I)	14.5 (1.72)	15.8 (2.2)	7.0 (0.48)
R$_{merge}$[a] (%)	7.1 (139.0)	8.1 (67.0)	17.5 (559.5)
CC$_{1/2}$	100 (79.5)	99.9 (85.1)	99.9 (25.4)
Wilson B-value (Å^2)	22.0	31.8	26.0
Refinement			
Resolution range (Å)	97.55-1.50	97.93-1.77	97.15-1.70
Molecules per asymmetric unit	1	1	1
R$_{work}$/R$_{free}$[b]	0.167/0.198	0.183/0.230	0.177/0.224
Deviations from ideal geometry			
Bond (Å)	0.0119	0.0117	0.0130
Angles (Å)	1.597	1.619	2.023
Ramachandran plot (%) Favored/allowed/outliers	94.89/5.11/0	95.4/4.6/0	95.4/4.6/0
Mean B-factors (Å^2)			
Protein	14.2	16.1	28.6
FAD/GLC/LMB	13.7/-/-	17.9/39.4/-	28.7/-/47.4
Water/Sulphate	41.34/61.7	44.3/98.3	51.7/107.7
Number of atoms			
Protein	5221	4589	5157
FAD/GLC/LMB	53/-/-	53/84/-	53/-/92
Water/Sulfate	649/55	457/35	273/15

a. $R_{merge} = \sum_i \sum_j |\, I_{i,j} - \langle I_j \rangle \,| \,/\, \sum_i \sum_j I_{i,j}$, where *i* runs over multiple observations of the same intensity, and *j* runs over all crystallographically unique intensities.

b. $R_{work} = \sum \big| |F_{obs}| - |F_{calc}| \big| \,/\, \sum |F_{obs}| \,/\, \sum_i \sum_j I_{i,j}$, where |F$_{obs}$|>0. R$_{free}$ is based on 5% of the data randomly selected and is not used in the refinement.
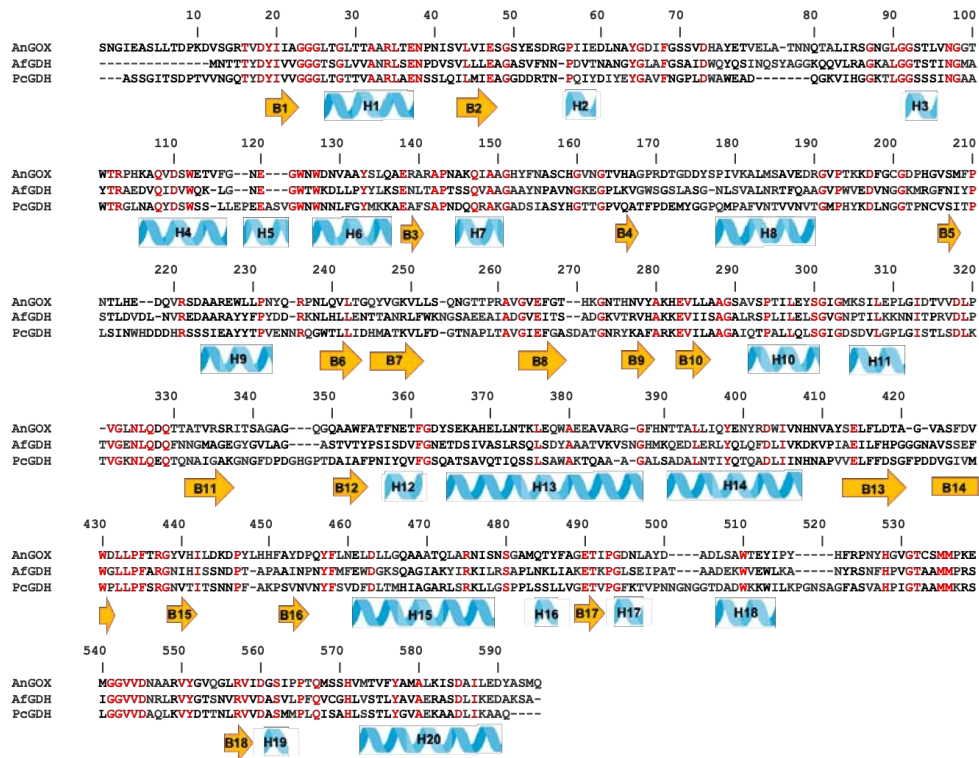
Figure 4.6. Sequence alignment of PcGDH, AfGDH and AnGOX based on structural superposition. Numbers refer to PcGDH sequence; conserved residues are colored in red. Secondary structure elements are assigned for PcGDH: β-strands are in orange and α-helices in light blue.

The FAD-binding domain is formed by the five-stranded (B1, B2, B6, B10, B18) parallel β-sheet A sandwiched between the three-stranded (B7, B8, B9) antiparallel β-sheet B and three α-helices (H1, H9, H20); B1, H1 and B2 form the βαβ-motif involved in the stabilization of the FAD cofactor.

The substrate-binding domain consists of a central six-stranded antiparallel β-sheet C (B5, B11, B12, B13, B14, B17) supported by seven α-helices (H8, H12, H13, H14, H15, H17, H18). The long helix H15 is in the focal point of the bent sheet and covers most of its back side together with the heavily bent H13 and the short H12. The sides of sheet C are covered by the helices H8, H14, H17, H18. Sheet C forms one side of a deep pocket (substrate-binding

43

cavity) containing the active site flavin near to the bottom and opposite to the β-sheet. The FAD-binding domain and the substrate-binding domain are linked through two types of connections: one consisting of three long segments connecting the two domain, another involving two extended two-stranded β-sheets (sheet D, parallel, and sheet E, anti-parallel) located at the interface between the two domains.

The protein surface has the overall shape of a bean, curved at the level of the substrate-binding cavity; it is possible to identify two faces, a convex face and a concave face, the latter providing access to the active site (Fig. 4.3).

The amino acid sequence of *Pc*GDH contains three potential N-glycosylation sites, Asn38, Asn188 and Asn439, identified by the consensus sequence Asn-X-Ser/Thr; two of these glycosylations are visible in the structure.
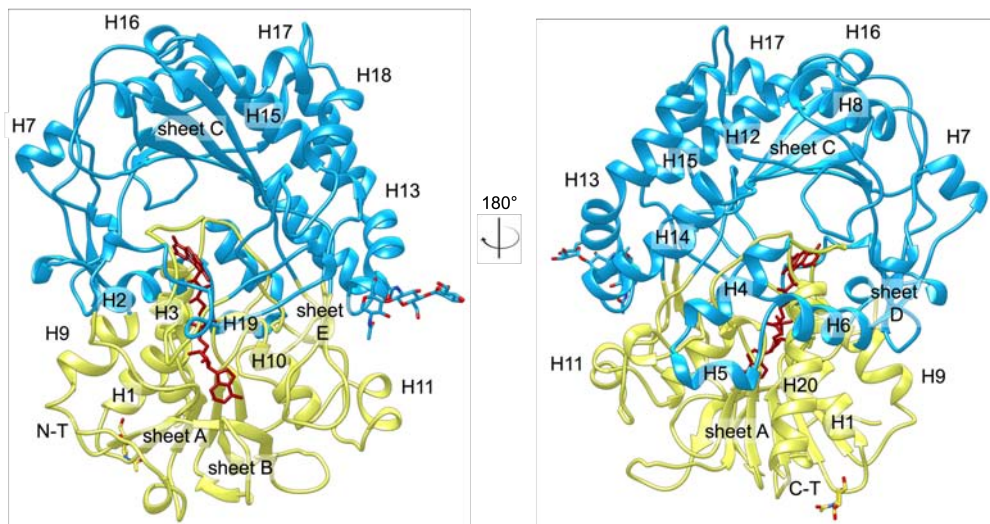

***Figure 4.2**. Two different views of the overall three-dimensional structure of ligand-free PcGDH in ribbon representation. The FAD-binding domain is shown in yellow, the substrate-binding domain in light blue, the FAD cofactor in dark red sticks. Secondary structure elements are assigned following the nomenclature in Figure 4.1. Glycosylations are represented in sticks and colored by heteroatom.*
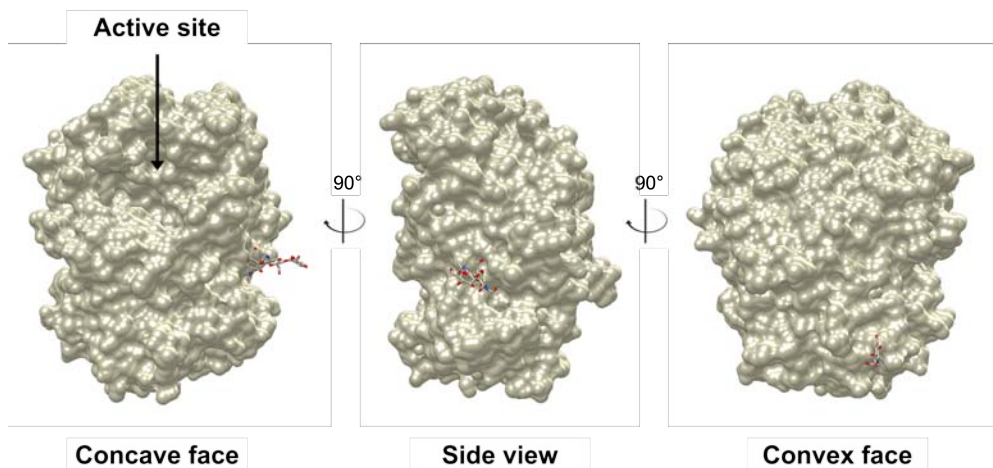
Figure 4.3. Molecular surface of ligand-free PcGDH observed from three different perspectives. The overall shape reminds of a bean with a convex face and a concave face. The substrate-binding cavity located on the concave face of the protein provides access to the active site. Glycosylations are represented in sticks and colored by heteroatom.

The Asn439 site shows a well-defined electron density for the first three sugar molecules: two $\beta(1{\rightarrow}4)$ linked N-acetylglucosamines and a $\beta(1{\rightarrow}4)$ linked mannose. Despite the less defined density at Asn38, the first covalently bound N-acetylglucosamine can be identified. Asn188 clearly shows a double conformation of the side chain, with an unmodelled blurred density in its proximity that might belong to the N-linked glycan.

The overall electron density is of very high quality except for three loop regions: the one connecting β-strands B8 and B9 (residues 269-275), part of the loop connecting B7 and B8 (residues 252-259) and the loop connecting B13 and B14 (residues 418-424). Modeling of these portions required the introduction of double conformations. Equal occupancies satisfied both the F_o-F_c and the $2F_o$-F_c map in the case of the B7-B8 and the B8-B9 loops, while the B13-B14 loop was modelled with a mixed occupancy. The main conformation, accounting for 70% of the protein molecules in the crystal, allows B13 to extend until Gly420 and B14 to start from Asp423. The remaining 30%

population has been modelled following the F_o-F_c map, since no clear density was identified in the $2F_o$-F_c map. After several rounds of refinement the F_o-F_c map still showed a diffuse missing density, indicating no definite conformation for the 30% conformer. As indicated also by the high B-factors in this region, the B13-B14 loop is endowed with high flexibility in the ligand-free form of *Pc*GDH, despite the hydrogen bonds typical of a β-sheet structure stabilize the main conformer.

4.1.2 *Active site geometry: the FAD-binding cleft and the reactive center*

The FAD molecule is non-covalently bound to the apoenzyme and occupies a narrow channel lined by ordered regions devoid of secondary structure, mainly the loop connecting the β-strand B1 and the helix H1 belonging to the βαβ-motif. The cofactor average B-factor (13.7 Å^2) is comparable to the average value observed for the protein atoms (14.2 Å^2).

Hydrogen bonds dominate the cofactor-protein interactions, which mainly involve the backbone NH and CO groups and several water molecules (Fig. 4.4). More in details, the adenine moiety interacts with the NH and CO groups of Ala247 through N1A and N6A atoms respectively; N6A is also stabilized by the presence of a water molecule. N3A interacts with the NH group of Ala48. The ribose moiety forms strong hydrogen bonds with the carboxyl group of Glu47, conserved in all GMC oxidoreductases. The pyrophosphate group forms an extended network of hydrogen bonds, five of them established with water molecules. The first phosphate group interacts with the NH group of Leu26 and with both the amino group and the side chain of Ser93; similarly, the second phosphate group interacts with Ala561 NH group and both the NH and the hydroxyl groups of Thr27.
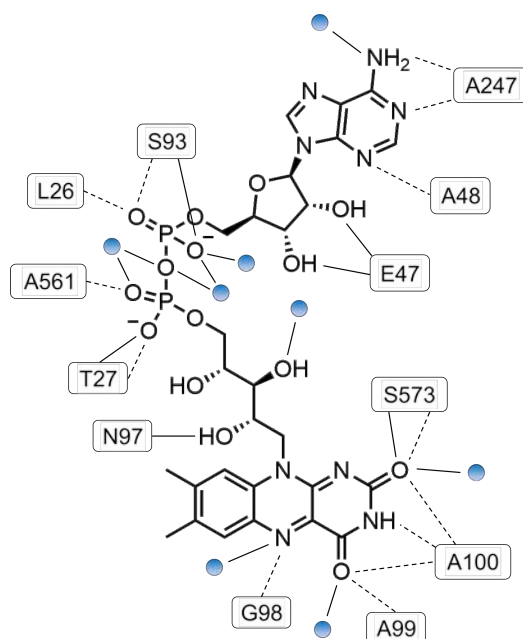
46

Figure 4.4. *Schematic representation of the FAD-protein interactions. Hydrogen bonds (distance <3.2 Å) are indicated by thick lines if they involve the amino acid side chain, dotted lines if they involve the NH or CO groups. Water molecules are depicted as blue circles.*

The negative charge of the pyrophosphate group is also stabilized by the orientation of helix H1, whose polarity at the N-terminus allows for charge compensation. The ribityl chain only forms two weak hydrogen bonds, with Asn97 side chain and with a water molecule. The isoalloxazine ring is rigidly fixed (Fig. 4.5A), especially at the pyrimidine moiety by hydrogen bonds of O2F, N3F and O4F. O2F interacts with the CO group of Ala100, a water molecule and both the NH group and the side chain of Ser573; N3F interacts with the CO group of Ala100; O4F interacts with both the NH groups of Ala99 and Ala100. Moreover, N5F is involved in a hydrogen bond with the NH group of Gly98. Interestingly, the hydroxyl group of Ser573 is found in an equally occupied double conformation: in one conformer it interacts with O2F, in the other is involved in a hydrogen bond with Asn97 side chain.
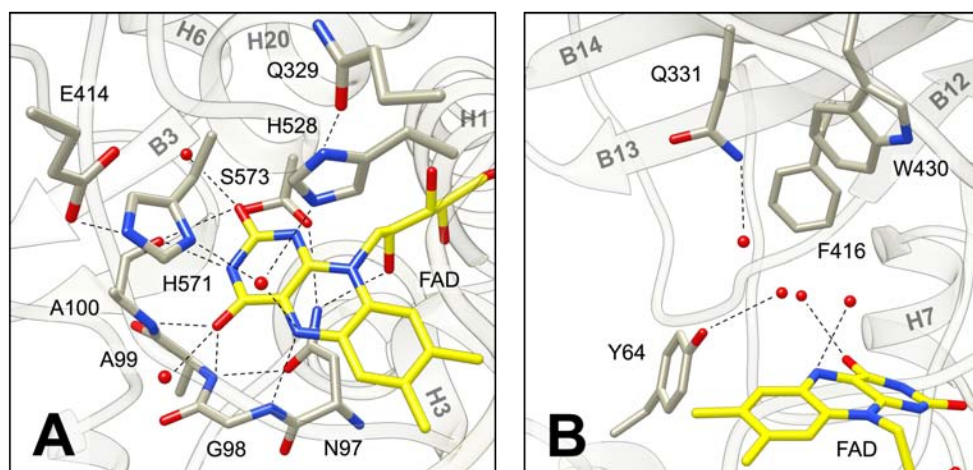
47

Figure 4.5*. Close-up view of the active site flavin - panel A: interactions between the isoalloxazine ring and the protein; panel B: residues facing the active site. The protein is shown in gray, the FAD cofactor in yellow, colored by heteroatom. The isoalloxazine ring is rigidly fixed through a dense network of hydrogen bonds. Most contacts involve the protein backbone except for Ser573, which stabilizes the cofactor also with its side chain. Ser573 is modelled in double conformation: one conformer interacts with the flavin system, the other interacts with Asn97 side chain. Asn97 is located on the si-face of the isoalloxazine ring, while the two catalytic histidines His528 and His571 lie on the re-face: the orientation of their imidazole ring is due to hydrogen bonding with Gln329 and Glu414 respectively. Except for Gln331, most of the residues directly facing the active site possess aromatic side chains: Tyr64, Phe416 and Trp430. Water molecules are depicted as red spheres. Hydrogen bonds (distance <3.2 Å) are represented with a dotted line.*

The electron density map of the isoalloxazine ring clearly shows a distortion from planarity (Fig. 4.6): the pyrimidine moiety is ~9° bent from the pteridine plane at the level of the reactive N5 atom.

This conformation of the ring in oxidized FAD has already been observed in other GMC oxidoreductases, in which an asparagine residue has been pointed out as critical in preventing the isoalloxazine ring to adopt a planar conformation. This interpretation is valid also for *Pc*GDH, where the conserved Asn97 is located right below the central part of the isoalloxazine ring and is further stabilized by Ser573 (Fig. 4.5A). This type of stabilization is observed also in CDH from *Myriococcum thermophilum* (*Mt*CDH) where the hydroxyl group belongs to a threonine residue instead of a serine; also in *Mt*CDH the cofactor is in a bent oxidized form.
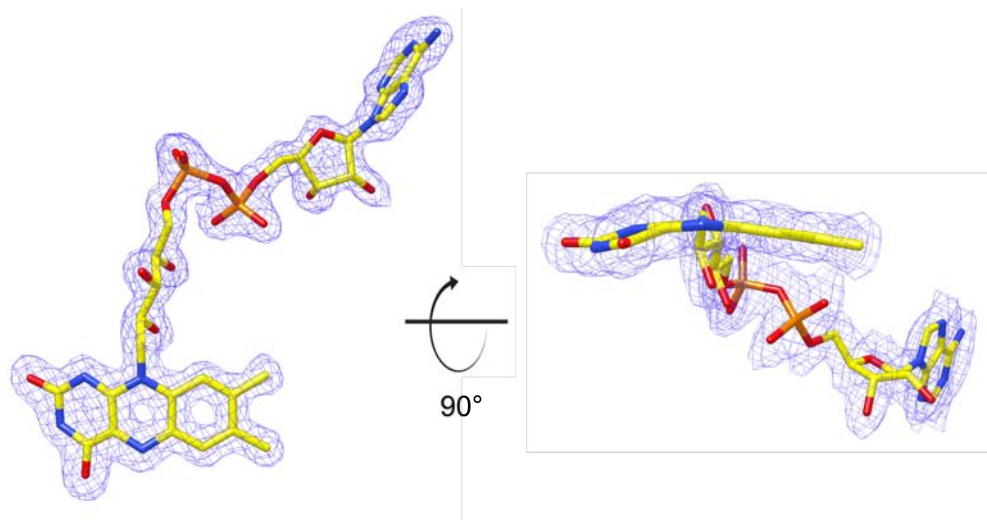
Figure 4.6. Two different views of the $2F_o$-F_c electron density map of the FAD cofactor contoured at 1σ. The FAD cofactor is shown in yellow sticks, colored by heteroatom; it adopts an extended conformation, with the adenine and the isoalloxazine moieties distal to each other. A distortion of ~9° from planarity was measured in comparison with the canonical isoalloxazine conformation of the oxidized cofactor.

In terms of biochemical reactivity, a distorted conformation of the isoalloxazine ring in the oxidized state implies that the flavin system can switch from the oxidized to the reduced form with minor conformational changes. The extensive network of hydrogen bonds involving the cofactor in *Pc*GDH corroborates this interpretation, since only a major reorganization of the protein environment could sustain a change in FAD conformation. The architecture of the protein backbone thus restrains the geometry of the cofactor, causing a modulation of the FAD redox potential[68]. Considering that fungal GDHs are thought to be involved in lignin degradation mainly as reducing agents, able to neutralize free radicals deriving from the host plant, an easily reducible flavin system is likely to be nature's choice in optimizing the efficiency of these enzymes.

The active site is located at the bottom of a large funnel-shaped cavity and is directly accessible from the solvent environment. As in other GMC

49

oxidoreductases, the catalytic pair His528/His571 is positioned in the proximity of the FAD, on the *re*-face of the isoalloxazine ring. The imidazole rings are oriented by Gln329 and Glu414 side chains, involved in a strong hydrogen bond with His528 and His571 respectively (Fig. 4.5A). Except for Gln331, most of the residues directly facing the active site possess aromatic side chains: Tyr64, Phe416 and Trp430 (Fig. 4.5B).

4.2 Functional clues from structure: *Pc*GDH in complex with glucose

4.2.1 *Substrate binding causes a structural rearrangement*

Crystals of *Pc*GDH soaked in glucose diffracted at 1.77 Å resolution. Ramachandran analysis indicates that 560 residues (95.4%) fall in the favored region while 27 of them (4.6%) are in the allowed region. Data collection and refinement statistics are summarized in Table 4.1. A total of seven glucose molecules are bound in the structure of glucose-soaked *Pc*GDH (*Pc*GDH-GLC): three on the external surface of the protein, four in the substrate-binding cavity, numbered according to their proximity to the active site.

*Pc*GDH in complex with glucose shows very few differences in the overall architecture compared to the ligand-free form: superposition of the two structures Cα atoms yields a root-mean-square deviation (r.m.s.d.) average value of 0.34 Å. The analysis of the r.m.s.d. values plotted as a function of the residue number (Fig. 4.7), however, shows Cα atoms displacement values exceeding 10 Å for residues forming the B13-B14 loop. Indeed, in *Pc*GDH-GLC this loop is locked in a single, well-defined conformation that restricts the access to the active site: residues Ser419 and Asp424 act as hinges bending the loop with a 90° angle toward the inner portion of the cavity (Fig. 4.8). The loop wraps on itself with a very tight geometry, allowed by the presence of a *cis*-proline (Pro422) at the turning point of the two antiparallel β-strands. This arrangement is completely different from the one observed in the structure of ligand-free *Pc*GDH where some extent of mobility for this loop was observed. It had to be modeled in two conformations, one of which points away from the active site and thus contributes to the wide opening of the substrate-binding cavity.
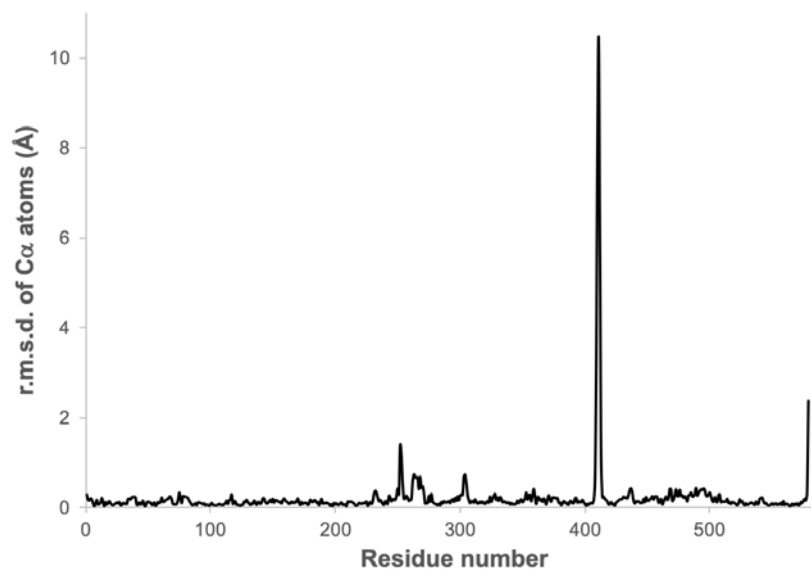
51

Figure 4.7*. Root-mean-square deviation of Cα atoms as a function of the residue number superposing the structures of ligand-free PcGDH and PcGDH-GLC. The highest r.m.s.d. values are observed for residues 420-423, corresponding to the B13-B14 loop.*

The Phe421 residue plays a key role in the B13-B14 loop conformational transition since it faces the solvent in the ligand-free protein and binding of glucose causes its phenyl ring to move ~17 Å closer to the active site. This major displacement allows Phe421 to interact with one of the glucose molecules found in the proximity of the FAD cofactor (Fig. 4.8).

No significant conformational changes are observed in residues directly facing the active site. The isoalloxazine ring of the cofactor, however, shifts by ~0.42 Å toward the substrate at the level of the N5 reactive center. The ring distortion from planarity is also more pronounced than in the ligand-free form, with a measured bending angle of ~22° (instead of ~9°) between the pteridine moiety and the pyrimidine moiety; this geometry is even more similar to the one typical of a reduced FAD in comparison with what observed in the ligand-free structure.
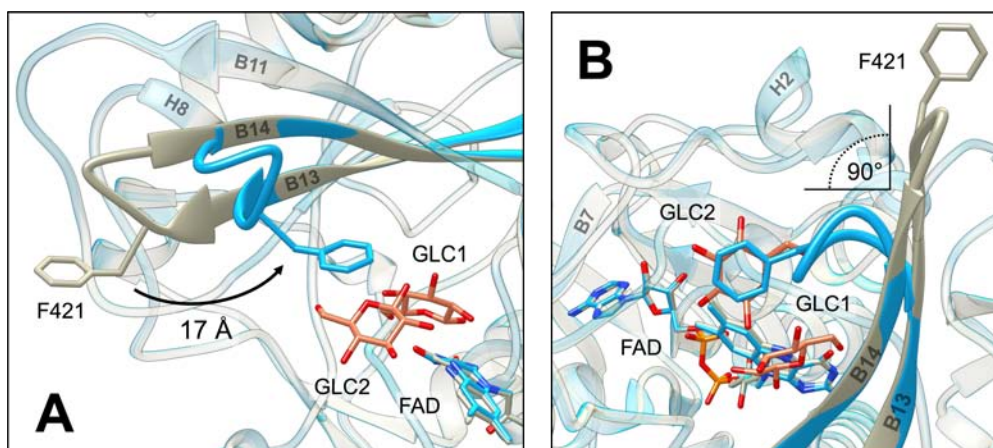
Figure 4.8. Superposition of ligand-free PcGDH (gray) and PcGDH-GLC (light blue) structures; glucose molecules are colored in orange, by heteroatom. Side view (panel A) and top view (panel B) of the active site and the B13-B14 loop: Phe421 (shown in sticks), which faces the bulk in the ligand-free structure, shifts ~17 Å toward the active site upon binding of glucose, and interacts with the second glucose molecule through CH-π interactions. This movement causes the B13-B14 loop to wrap on itself and to rotate by ~90° toward the active site.

Nevertheless, structure refinement carried out using the reduced FAD were not successful; the electron density map unequivocally identifies the cofactor as oxidized, despite the major distortion from planarity. The change in FAD conformation might represent a substrate-induced adaptation of the flavin system, which enhances the structural distortions already observed in the ligand-free structure[69].

4.2.2 *PcGDH-GLC active site*

Four glucose molecules are found in the proximity of the *Pc*GDH-GLC active site: two of them (GLC3 and GLC4) adhere to the edge of the funnel-shaped cavity leading to the active site, while the other two (GLC1 and GLC2) are very close to the FAD cofactor, arranged in line with one another. The quality of the electron density map allowed to orient the latter two molecules in a unique fashion, but it indicates a certain degree of mobility. Indeed, the average

53

B-factors of these two ligands (44.6 Å^2 for GLC1, 68.7 Å^2 for GLC2) are higher than the average value calculated for the protein atoms (16.1 Å^2).

The first glucose molecule located in the active site is found at the bottom of the substrate-binding cavity, directly facing the cofactor (GLC1 in Fig. 4.9). The reactive hydroxyl group points toward the Nε2 atom of His528, in a position that would allow the proton abstraction hypothesized in the catalytic mechanism. The O2 hydroxyl group interacts with Val526 CO group and a water molecule; the side chain of Gln331 stabilizes the sugar O3 hydroxyl group, which is also involved in a hydrogen bond with the O1 hydroxyl group of the second glucose molecule found in the active site. The first ligand also interacts with the O4F atom of the pyrimidine portion of FAD through its CH$_2$OH group.
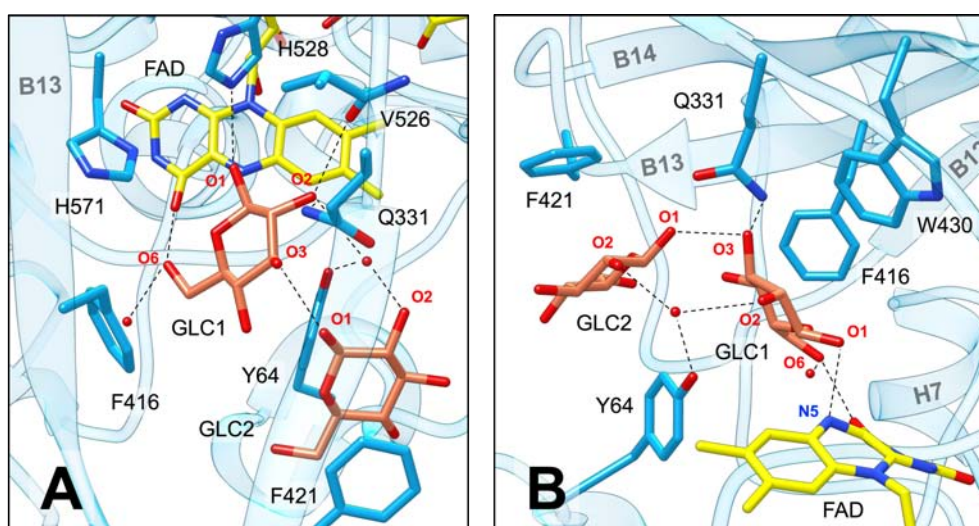

Figure 4.9. *Active site of PcGDH-GLC - panel A: top view; panel B: side view. The protein is shown in light blue, the glucose molecules in orange, the FAD cofactor in yellow, colored by heteroatom. The two glucose molecules are numbered according to their proximity to the flavin system. GLC1 is oriented toward the cofactor and His528 with its reactive hydroxyl group (O1). The O3 atom is very close to the O1 atom of the following glucose molecule (~2.7 Å). Both ligands are sandwiched between the aromatic residues Phe421, Tyr64, Phe416 and Trp430, which define the pathway connecting the solvent to the active site. Water molecules are depicted as red spheres. Hydrogen bonds (distance <3.2 Å) are represented with a dotted line. Panel A: B14 and H15 are omitted in the structure to allow visualization from the top of the active site.*

The glucose molecule is further stabilized by carbohydrate-aromatic interactions, *i.e.* interactions of an aromatic π-system with the C-H bonds of the sugar. The two aromatic rings of Trp430 and Phe416 lie almost on the same plane, forming an aromatic platform above the β-face of the glucose molecule. The same type of interaction is found on the other side of the ligand, which faces the aromatic ring of Tyr64.

The second glucose molecule found in the proximity of the active site is oriented with its reducing O1 hydroxyl group pointing toward the O3 and O4 hydroxyl group of the first glucose molecule (GLC2 in Fig. 4.9). The ligand does not take part in any hydrogen bond with the protein environment, but its CH groups on the β-face are engaged in CH-π interactions with Phe421 belonging to the B13-B14 loop, whose phenyl ring is perfectly parallel to the glucose pyranose ring (Fig. 4.9B). The α-face is stabilized by the aromatic contribution of Tyr64 phenol ring; this residue stabilizes both the α-faces of the two glucose molecules and together with Phe416 and Phe421 orient them with a precise geometry in the active site.

It is thus possible to point out the two main forces contributing to the overall process of substrate binding. The two main energetic contributions are i) the aromatic stabilization due to electron-rich residues and ii) the steric effects of the bulky aromatic residues interacting with the substrate, which impose a strict directionality to the path connecting the aqueous environment to the flavin system. Remarkably, the movement of Phe421, followed by the entire B13-B14 loop, could escort the second glucose molecule to the proximity of the active site, causing a narrowing of the substrate-binding cavity.

Moreover, two additional glucose molecules are bound at the periphery of this cavity (Fig. 4.10). Their position suggests that they define a diffusion pathway from the bulk to the flavin system, indicating that the wide substrate-binding
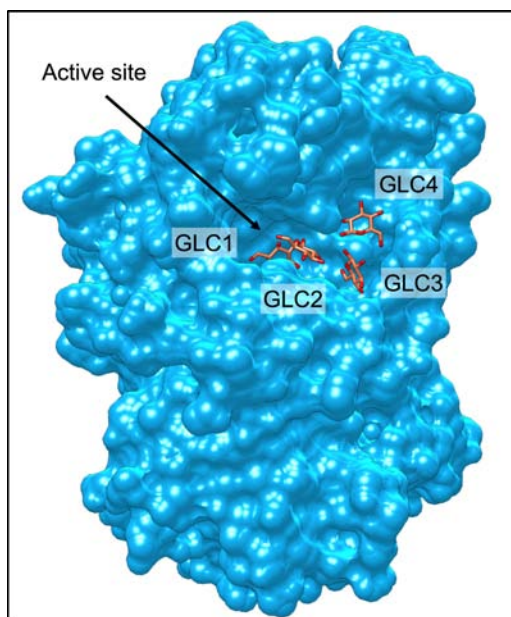
Figure 4.10. Surface representation of the funnel-shaped substrate-binding cavity, on the concave face of PcGDH. Four glucose molecules bind to this region, numbered according to their proximity to the active site. They seem to be caught in the diffusion process from the bulk to the flavin system. PcGDH surface is shown in light blue, glucose molecules in orange sticks, colored by heteroatom.

cavity might act as a funnel conveying the substrate molecules to the reactive center.

4.2.3 *External sugar-binding sites*

A peculiar feature of the *Pc*GDH-GLC structure is the presence of three glucose molecules on the external surface of the protein. Both the initial F_o-F_c and $2F_o$-F_c maps showed well-defined electron densities for these glucose molecules; they are rigidly fixed, showing B-factors comparable to the protein counterpart. They bind to three external small cavities that will be referred to as external sugar-binding sites (SBSs), numbered according to their proximity to the active site cavity; all of them are accessible from the convex face of *Pc*GDH and are shown in Figure 4.11.

56

SBS1 is a small pocket between sheet D and H7, resulting from the protruding loop connecting B5 and H9. Most of the glucose-protein contacts form a dense network of hydrogen bonds involving the sugar hydroxyl groups. The O2 atom is stabilized by the hydroxyl group and the CO group of Ser141, together with the Ala142 CO group. The O3 atom tightly interacts with the hydroxyl group of Thr168, the CO group of Phe169 and a water molecule mediating the interaction with Asp219 side chain, the NH_2 group of Gln147 might also be involved in this hydrogen bond network; the O4 atom interacts with a water molecule and the carboxyl group of Asp171. The glucose molecule in SBS1 is further stabilized by the aromatic side chain of Trp215, engaged in CH-π interactions with the α-face of the sugar ring.

SBS2 is located between H1 and H9, right under the sheet D. The type of interactions between the glucose molecule and the protein environment in SBS2 are similar to those found in SBS1. The O2 atom interacts with the carboxyl group of Glu37, the terminal nitrogen atom of Arg236 and a fixed water molecule, which also stabilizes the O3 atom together with the CO group of Tyr228; the O4 hydroxyl group is involved in hydrogen bonds with two CO group belonging to Tyr228 and Ala227. The pyranose ring is sandwiched between the side chains of Pro164, Tyr228 and Val232. The CH-π stabilization provided by an aromatic residue observed in SBS1 is missing in SBS2.

While SBS1 and SBS2 are on the same side of the protein, SBS3 is located on the convex face of the substrate-binding domain, between helices H8 and H12. The electron density of the glucose molecule bound to this site is weaker compared to the one observed in the other two SBSs; the ligand also shows higher temperature-factors and has been modelled with an 80% occupancy. It interacts with the protein through three hydrogen bonds: the O3 and O4 atoms
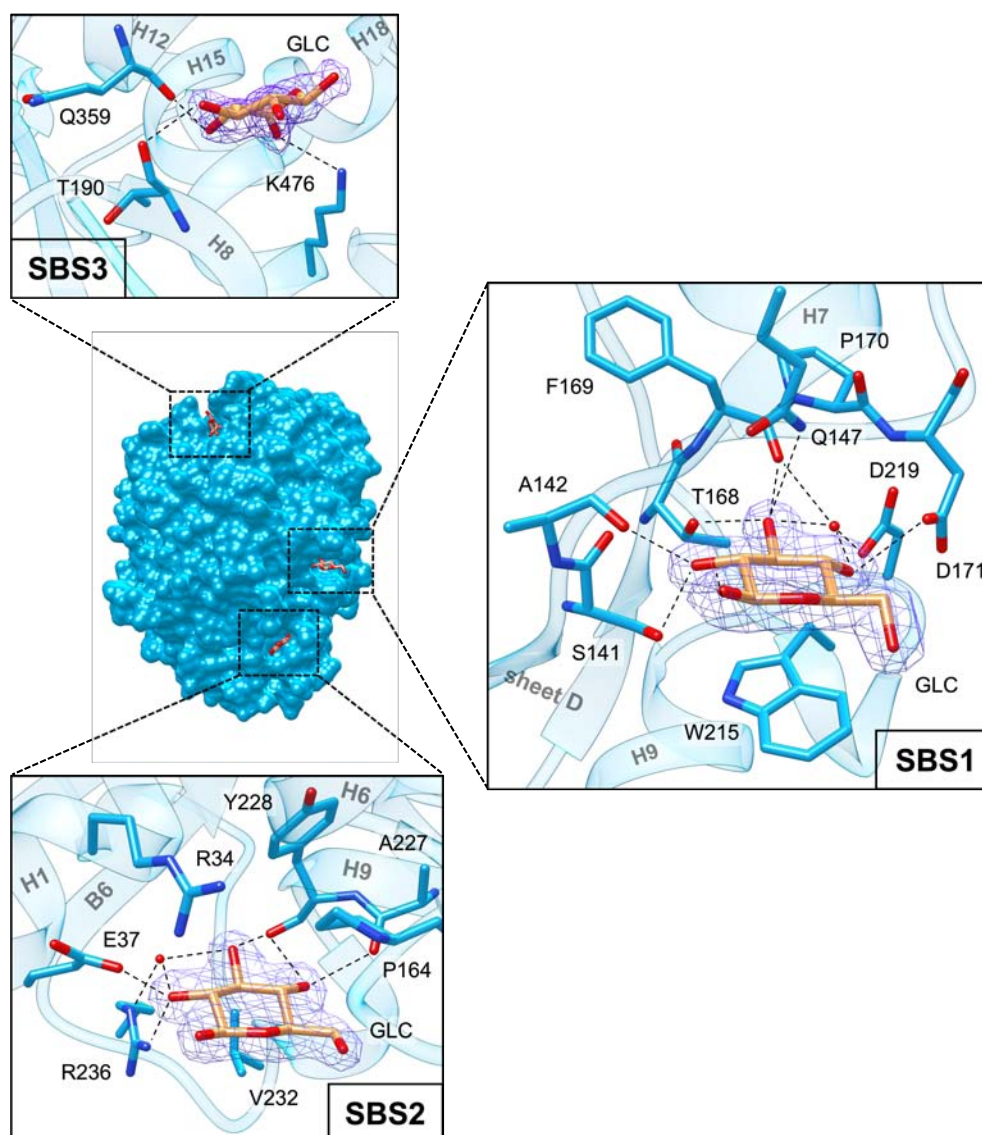
Figure 4.11*. External sugar-binding sites (SBSs) found in PcGDH-GLC. All three sites are located on the convex face of PcGDH, whose surface is represented in the center and colored in light blue. Glucose molecules are represented in sticks and colored in orange, by heteroatom. The three sites are numbered according to their proximity to the active site, located on the concave face of the protein (not shown). The three boxes show the detail of the interactions between the protein and the ligands found in the three sites. Water molecules are depicted as red spheres. Hydrogen bonds (distance <3.2 Å) are represented with a dotted line. The $2F_o$-F_c electron density map is also depicted for the glucose molecules, contoured at 1σ.*

interact with the CO group of Thr190 and Gln359 respectively, while the O6 atom interacts with the amino group of Lys476.

All three SBS are directly accessible from the aqueous environment. In SBS1 and SBS2 the glucose molecule is anchored to the protein through its three O2, O3 and O4 hydroxyl groups, with the CH_2OH group and the O1 reducing end exposed to the bulk. Conversely, the glucose molecule found in SBS3 appears more loosely attached with the O1 and O2 hydroxyl groups pointing in the direction of the solvent.

The superposition of ligand-free *Pc*GDH and *Pc*GDH-GLC reveals that residues in SBS2 are pre-organized, with no difference in the side chain arrangement of the two structures. On the other hand, binding of glucose to SBS1 causes a reorganization of residues Ser141, Asp171 and Trp215: their mobile side chains modelled as double conformers in ligand-free *Pc*GDH are locked in a single conformation upon binding of glucose. A similar behavior is found in SBS3, where Lys476 can be modelled as a single conformer only when glucose is bound.

This is the first time that secondary sugar-binding sites are found in a GMC oxidoreductase. The molecular bases of sugar recognition in these sites restrict the hypothetical natural binders to a limited group of carbohydrates. Only monosaccharides or oligosaccharides containing three free non-reducing hydroxyl groups at C2, C3 and C4 can externally bind *Pc*GDH. Potential naturally occurring ligands are the non-reducing ends of polysaccharides like cellulose or hemicellulose, together with all their possible cleavage products. Interestingly, also branched polysaccharides such as xyloglucans, glucuronoxylans, galactomannan and galactoglucomannans found in hemicellulose (Fig. 1.5) are plausible binders. The units forming the branches of these polysaccharides consist of xylose, glucuronic acid and galactose possessing free O2, O3 and O4 hydroxyl group. The stereochemistry of these

59

groups in xylose and glucuronic acid is identical to glucose, while is inverted at the O4 in galactose. If the external SBSs identified in *Pc*GDH are able to recognize these branching units, they might have a physiological relevance. As a possibility, they might be of importance in guiding the secreted enzyme toward the polysaccharide matrix of lignocellulose where its action is required.

4.3 Structure of *Pc*GDH bound to laminaribiose

4.3.1 *A puzzling substrate recognition mechanism*

The structural analysis of *Pc*GDH-GLC gave some hints about novel binding modes of the physiological substrate glucose to the enzyme. Interestingly, two glucose molecules were found in the proximity of the cofactor, not one. These two ligands show a defined orientation and are also very close to each other: the O3 atom of the first glucose molecule and the O1 of the following are only ~2.7 Å apart (Fig. 4.9). Moreover, the largest conformational change observed upon binding of glucose is primarily caused by the B13-B14 loop Phe421 recruitment in binding the second glucose molecule, not the one that directly faces the cofactor. These observations suggested that *Pc*GDH might be also able to bind a disaccharide. The orientation of the two glucose molecules is compatible with a hypothetical disaccharide containing a β(1→3) linkage, so laminaribiose was tested for the structural analysis. Soaking in laminaribiose was successful and the structure of laminaribiose-bound *Pc*GDH (*Pc*GDH-LMB) was solved and refined to 1.7 Å resolution. Data collection and refinement statistics are summarized in Table 4.1. Ramachandran analysis indicates that most residues (95.4%) fall in the favored region, 27 residues (4.6%) are in the allowed region and there are no outliers.

Superposition of ligand-free and laminaribiose-bound *Pc*GDH structures gave an average r.m.s.d. of 0.41 Å for the Cα atoms. As observed in *Pc*GDH-GLC, however, the Cα displacement vs residue number plot shows r.m.s.d. values exceeding 10 Å only for residues forming the B13-B14 loop (data not shown). Superposition of *Pc*GDH-LMB and *Pc*GDH-GLC models gave an average r.m.s.d. of 0.24 Å for the Cα atoms. Indeed, the two sugar-bound structures of

*Pc*GDH are almost identical, showing that binding of laminaribiose causes the same rearrangements observed in the case of glucose.

As in the case of *Pc*GDH-GLC, one molecule of laminaribiose occupies each of the three external sugar-binding sites; however, only one disaccharide molecule is found in the active site.

The electron density map indicates the latter to adopt a double conformation: the primary conformer is modelled with a 60% occupancy and the secondary one with a 40% occupancy. Figure 4.12 shows how the two glucosyl units of the disaccharide are oriented just like the two glucose molecules found in the active site of *Pc*GDH-GLC (Fig. 4.9). The two conformations of laminaribiose differ in the orientation of the reducing unit: they are flipped by 180° with the O1 and O6 atoms in opposite positions (Fig. 4.12A).
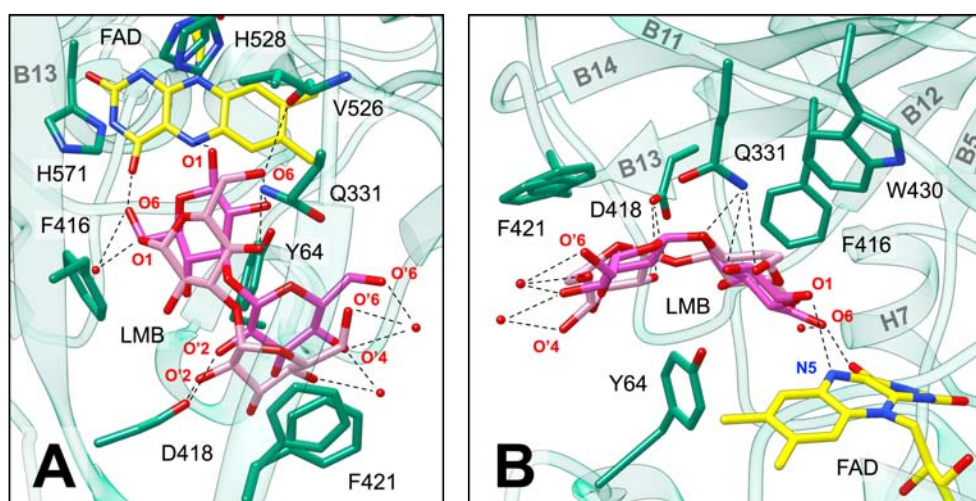

Figure 4.12. Active site of PcGDH-LMB - panel A: top view; panel B: side view. The protein is shown in green, the FAD cofactor in yellow, laminaribiose in magenta (primary conformer) or pink (secondary conformer), colored by heteroatom. The primary conformer is closer to the FAD cofactor than the secondary one, and oriented with the reactive O1 atom toward His528 and the FAD N5 atom. His528 and Phe421 are found in double conformation. Water molecules are depicted as red spheres. Hydrogen bonds (distance <3.2 Å) are represented with a dotted line. Panel A: B14 and H15 are omitted in the structure to allow visualization from the top of the active site.

In the primary conformer the reducing hydroxyl group is oriented toward the catalytic His528 and the FAD reactive center, like in the glucose-bound structure. The distance between the laminaribiose O1 atom and the FAD N5 atom is ~3.2 Å, while is ~2.6 Å in the case of glucose. This increased distance implies the loss of the hydrogen bond with the catalytic His528, which is also found in double conformation. The secondary laminaribiose conformer is shifted by ~1.0 Å backward with respect of the primary one, so even the interaction with the cofactor is missing in this conformer.

The amino acid residues involved in hydrogen bonds with the ligand are the same observed in PcGDH-GLC, but they contact different hydroxyl group of laminaribiose. The CO group of Val526 interacts with the O6 atom of the secondary conformer. The NH_2 group of Gln331 interacts with the O2 atom of the primary laminaribiose conformer and with both the O6 and O4 of the secondary conformer. Asp418, which does not interact with any glucose molecule in PcGDH-GLC, is involved in a hydrogen bond with the O'2 atom of laminaribiose in both conformers. The CH-π aromatic interactions found in PcGDH-GLC are conserved in PcGDH-LMB, since the orientation of the two pyranose rings of the disaccharide is the same observed for the monosaccharide counterpart. In the B13-B14 loop, Phe421 side chain is found in double conformation, each conformer interacting with the non-reducing unit of the two laminaribiose conformations. No further laminaribiose molecules were found in the proximity of the substrate-binding cavity.

The structure of PcGDH in complex with laminaribiose confirms the hypothesis about the ability of PcGDH to bind a disaccharide in the active site. Moreover, the functional analysis (described in par. 4.5) demonstrated that laminaribiose acts indeed as a substrate of PcGDH. Interpretation of the electron density map, however, required the introduction of a double

63

conformation. The temperature factors are also notably high for both conformers. These observations suggest that laminaribiose is not rigidly fixed in the active site. Indeed, the directional interactions, *i.e.* hydrogen bonds, are not enough to stabilize all the hydroxyl groups of the sugar and thus orient it in a unique fashion. As in the case of glucose-bound *Pc*GDH, the main forces responsible for the orientation of the two glucosyl units of laminaribiose are CH-π stackings with aromatic residues, whose chemistry ensures a precise orientation of the pyranose planes but does not provide a strict directionality to the hydroxyl groups. The high mobility can then be associated to an intrinsically loose substrate binding mechanism, which might not require the formation of highly specific interactions between the active site residues and laminaribiose.

Binding of both glucose and laminaribiose to *Pc*GDH does not cause any significant structural rearrangement in the protein environment close to the cofactor. The only remarkable conformational change is observed in the B13-B14 loop, which assumes a completely different conformation upon binding of both sugars. Since the type of reorganization observed for this loop is the same for both substrates, it can be referred to as the "substrate recognition loop". The role of Phe421, primary involved in the substrate recognition process, is clarified with the structure of *Pc*GDH-LMB. In the case of laminaribiose this residue acts as a platform for the non-reducing glucosyl unit of the disaccharide rather than providing a path connecting the solvent to the active site, as pointed out in *Pc*GDH-GLC. This type of stabilization is found also in *Mt*CDH[60], where a tryptophan residue interacts with the non-reducing glucosyl unit of cellobiose but no significant conformational change in the protein backbone is observed.

Structural changes due to binding of ligands are not common in GMC oxidoreductases, which usually display a pre-organized structure, with the exception of P2O from *Trametes multicolor* (see par. 4.4.2).

4.3.2 *Specificity of the external sugar-binding sites*

As in the case of glucose, three laminaribiose molecules are found on the external surface of the protein, one for each of the three SBS identified in the structure of *Pc*GDH-GLC. In all cases the electron density map is very well-defined, especially for the glucosyl unit primarily interacting with the protein environment; this portion also shows lower B-factors than the other unit. Figure 4.13 shows the detail of the interactions between laminaribiose and the protein environment in these sites, which are very similar to those observed in *Pc*GDH-GLC.

Although, some additional interactions stabilizing the disaccharide can be identified. In SBS1 the O'2 hydroxyl group of laminaribiose is additionally stabilized by a hydrogen bond with the CO group of Ser141, while the CO group of Ala142 contacts both the O2 of the disaccharide. Even in SBS2 the additional interaction involves the two O2 groups, both interacting with the carboxyl group of Glu37. In SBS3 laminaribiose interacts with the hydroxyl group of Ser363 through the O2 atom of the reducing unit; this interaction is not found in *Pc*GDH-GLC.

Interestingly, all SBSs bind laminaribiose through the second glucosyl unit, *i.e.* the non-reducing end. This portion is anchored to the protein with its O'2, O'3 and O'4 hydroxyl group, with the same binding mode observed in *Pc*GDH-GLC. As a consequence, the non-reducing end is rigidly fixed and buried in the protein pocket, while the reducing end is exposed to the solvent and then more mobile compared to the non-reducing unit.
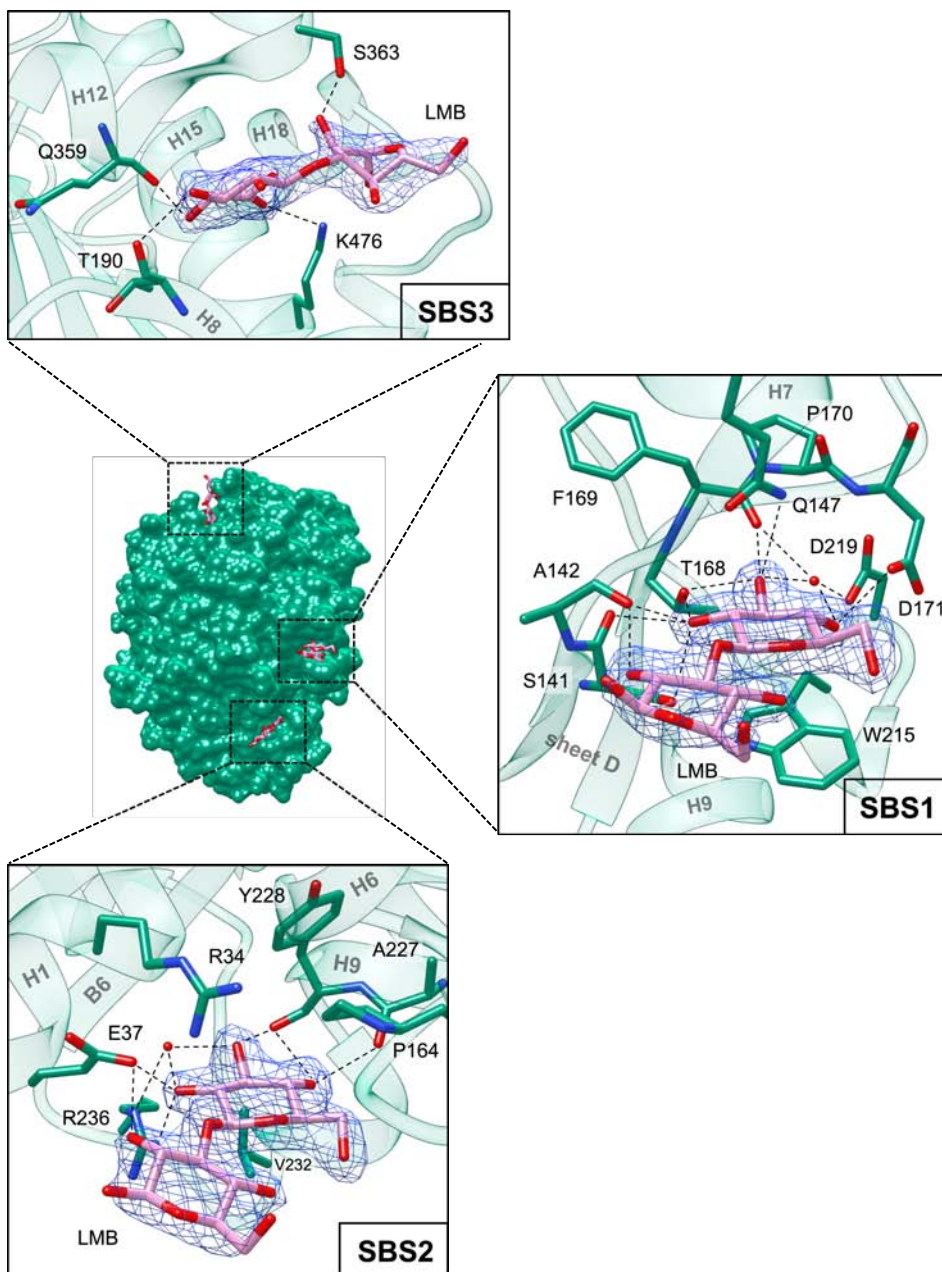
65

Figure 4.13*. External sugar-binding sites (SBSs) found in PcGDH-LMB. All three sites are located on the convex face of PcGDH, whose surface is represented in the center and colored in green. Laminaribiose molecules are represented in sticks and colored in pink, by heteroatom. The three sites are numbered according to their proximity to the active site, located on the concave face of the protein (not shown). The three boxes show the detail of the interactions between the protein and the ligands found in the three sites. Water molecules are depicted as red spheres. Hydrogen bonds (distance <3.2 Å) are represented with a dotted line. The $2F_o$-F_c electron density map is also depicted for the glucose molecules, contoured at 1σ.*

66

The observation that all SBSs bind both glucose and laminaribiose in the same specific fashion strengthens the hypothesis that recognition of sugars by remote sites located on the external protein surface might be of importance for *Pc*GDH enzymatic activity in the physiological context.

4.4 Comparison with *Af*GDH, *An*GOX and other GMC oxidoreductases

4.4.1 *A conserved overall fold*

*Pc*GDH displays the characteristic fold of GMC oxidoreductases. In terms of sequence and structure similarity, GDH from *Aspergillus flavus* (*Af*GDH) and GOX from *Aspergillus niger* (*An*GOX) are closely related to *Pc*GDH. The structure-based sequence alignment gave a 36.4% identity with *Af*GDH and 34.3% identity with *An*GOX (Fig. 4.1). Cα atoms superposition yielded a 1.2 Å r.m.s.d. in the case of *Af*GDH (PDB entry 4YNT[54]) and 1.3 Å for *An*GOX (PDB entry 1CF3[53]). In terms of oligomeric state, the two GDHs are monomers while *An*GOX is a dimer. *Pc*GDH and *An*GOX structures have been modelled with an oxidized FAD showing a distortion from planarity; in *Af*GDH structures the cofactor is found in its reduced state.

*Pc*GDH and *Af*GDH share a very similar overall architecture. *Pc*GDH has an 11 amino acids insertion at the N-terminus and a 5 amino acids insertion folding into H5, absent in *Af*GDH. A 7 amino acids depletion in the loop region between H2 and H3 is structurally compensated by 5 additional amino acids located at the end of H18 in *Pc*GDH (H16 in *Af*GDH).

*Pc*GDH and *An*GOX also share a common backbone architecture. In terms of overall fold, *An*GOX has a 3 amino acids longer N terminus compared to *Pc*GDH, and a 6 amino acid insertion in the loop region between H2 and H3, like *Af*GDH. The helix H5 found in *Pc*GDH, caused by a 5 amino acids insertion, is not conserved in *An*GOX. *Pc*GDH shows an additional 5 amino acids insertion at the end of H18 (H12 in *An*GOX), like in comparison to *Af*GDH. In both cases, this insertion has the effect of slightly narrowing the

substrate-binding cavity of *Pc*GDH in comparison to *Af*GDH and *An*GOX. The maximal width of the cavity measured in *Pc*GDH is ~28 Å, while is ~31 Å for *Af*GDH and ~34 Å for *An*GOX.

4.4.2 *The peculiar substrate binding mode of PcGDH*

The structure of *Af*GDH is available in the ligand-free form (PDB entry 4YNT) and in complex with the oxidation product D-glucono-δ-lactone (LGC - PDB entry 4YNU) and it shows a substrate recognition mechanism completely different from the one observed in *Pc*GDH. In *Af*GDH-LGC the ligand is tightly bound in the active site through hydrogen bonds involving all the hydroxyl groups of glucose except O6. Figure 4.14 shows the superposition of the residues involved in interactions with ligands in the active sites of *Pc*GDH-GLC and *Af*GDH-LGC.
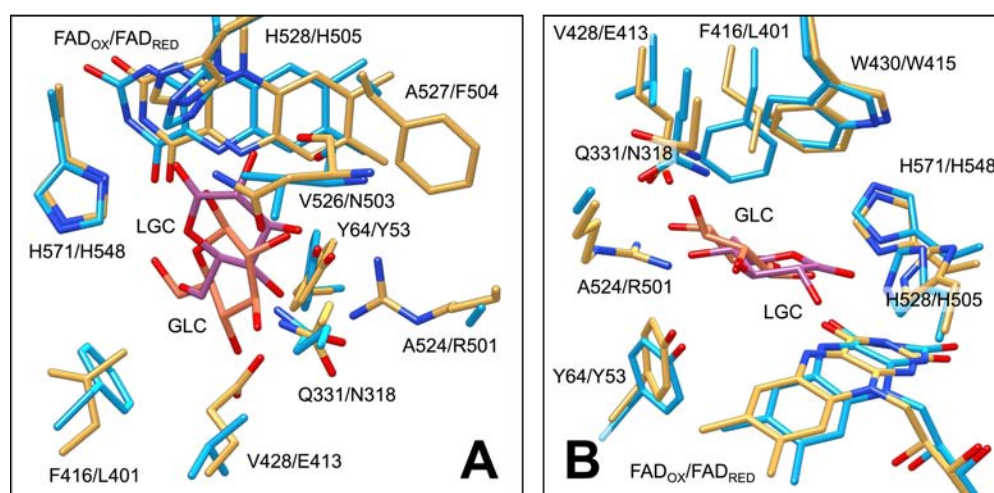

Figure 4.14. *Structural comparison between PcGDH-GLC and AfGDH-LGC active sites - panel A: top view; panel B: side view. PcGDH is shown in light blue, AfGDH in ocher, GLC in orange and LGC in purple, colored by heteroatom. The first residue specified belongs to PcGDH, the second to AfGDH.*

69

Compared to the O1 hydroxyl group of GLC, the carbonyl group of LGC is ~1.34 Å closer to the cofactor, which results in hydrogen bonding with both catalytic histidines (His505/His548). The other residues involved in stabilizing LGC are: Asn503, interacting with both its side chain CO and its backbone CO groups; Arg501, contacting both O3 and O4 hydroxyl group through its terminal nitrogen atoms; Tyr53 and Glu413, interacting with both the O5 hydroxyl group; Leu401 and Trp415 providing an hydrophobic environment around the bound LGC on the *si*-face, which corresponds to the β-face of glucose.

*Pc*GDH displays a very different substrate binding mode, since both glucose and laminaribiose mainly interact with the active site residues through CH-π interactions. As an example, the conserved Tyr64 in *Pc*GDH (Tyr53 in *Af*GDH) interacts with sugars through aromatic stacking rather than hydrogen bonding. Glu413, Arg501 and Asn503 in *Af*GDH correspond to Val428, Ala524, and Val526 in *Pc*GDH, unable to interact through hydrogen bonds with the ligand. Gln331, which is involved in a hydrogen bond with glucose in *Pc*GDH, corresponds to Asn318 in *Af*GDH, not involved in any contact with LGC. The aromatic Trp430 found in *Pc*GDH is conserved in *Af*GDH (Trp415), but the proximal Phe416 corresponds to Leu401 in *Af*GDH, which lowers the aromatic contribution to CH-π interactions with the ligand.

The most evident difference between the two substrate-bound protein complexes is the absence of a second ligand molecule in the active site of *Af*GDH. Indeed, the substrate recognition loop, formed by residues 419-424 (SGFPDD) in *Pc*GDH, corresponds to residues 404-409 (PGGGNA) in *Af*GDH; this region lacks the aromatic residue responsible for the closure of the loop on the second glucose molecule found in the active site of *Pc*GDH. The different substrate binding modes are probably related to the different

physiological functions of the two proteins: the recognition of oligosaccharides containing a β(1→3) linkage might be a peculiar feature of *Pc*GDH, not shared with *Af*GDH. This explains why the binding of LGC does not trigger any significant conformational change in the loop connecting B15 and B16 of *Af*GDH. An additional ligand molecule is found at the edge of the substrate-binding cavity of *Af*GDH-LGC structure, in the same position occupied by GLC4 in *Pc*GDH-GLC structure (Fig. 4.10).

The crystallographic structure of *An*GOX is available only in the ligand-free form, but the authors propose an enzyme-substrate complex model based on molecular modeling. Figure 4.15 shows the superposition of the residues in the active sites of *Pc*GDH-GLC and *An*GOX. Most residues are conserved between *An*GOX and *Af*GDH, and so is the chemistry of the predicted protein-ligand interactions. Notably, the aromatic platform formed by Phe416/Trp430 in *Pc*GDH is conserved in *An*GOX (Phe414/Trp426).
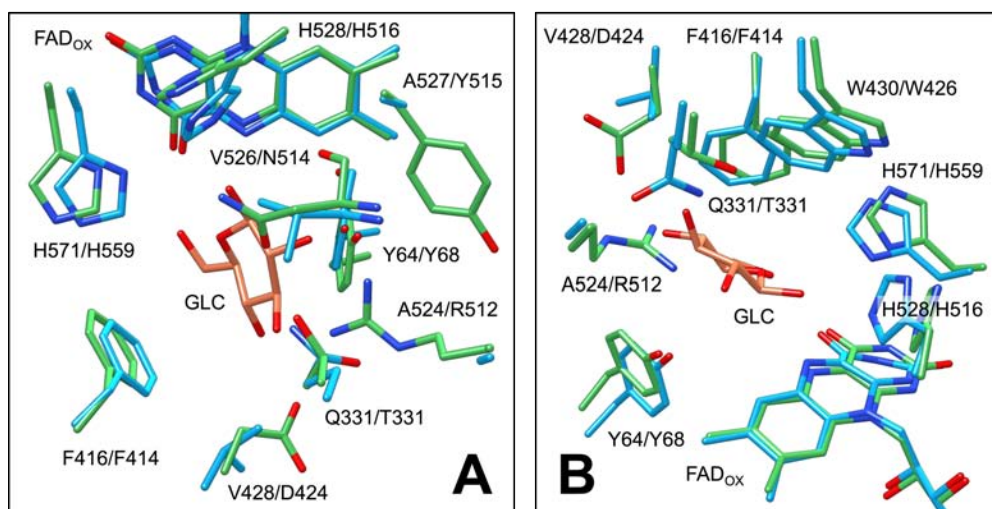

Figure 4.15. *Structural comparison between PcGDH-GLC and AnGOX active sites - panel A: top view; panel B: side view. PcGDH is shown in light blue, AnGOX in light green, GLC in orange, colored by heteroatom. The first residue specified belongs to PcGDH, the second to AnGOX.*

71

The region corresponding to the substrate recognition loop of *Pc*GDH, however, is not conserved in *An*GOX (residues 417-420, corresponding to the sequence TAGV).

An interesting feature observed in both *Af*GDH and *An*GOX is the presence of a bulky aromatic residue, Phe504 and Tyr515 respectively, in the proximity of the FAD pteridine moiety. In *Pc*GDH these residues correspond to the small Ala527 (Figg. 4.14A and 4.15A), with water molecules occupying the position of the aromatic rings. As a result, in *Pc*GDH there is a large unoccupied space close to the cofactor even in the sugar-bound structures. This space availability raises a point about *Pc*GDH ability to potentially bind even more sugar substrates or substrates with a different type of linkages. Alternatively, this additional space might be relevant in the secondary redox reaction, *i.e.* in the secondary electron acceptor recognition process.

Most of the available structures of GMC oxidoreductases in complex with substrates show a pre-organized active site, with no remarkable change in the architecture of the protein backbone. The only exception is P2O from *T. multicolor*: this protein is structurally very distant from *Pc*GDH, but they both share a substrate recognition mechanism that involves reorganization of a loop region (substrate recognition loop) upon binding of the pyranose ligand. In *Tm*P2O, however, the nature of this transition is very different, if not the right opposite, from what observed in *Pc*GDH: instead of the Phe421 recruitment in stabilizing the substrate, a Phe residue (Phe454) moves away from the flavin reactive center opening the active site and allowing the entrance of the ligand, since the chemistry of *Tm*P2O-ligand interactions is based on hydrogen bonding rather than CH-π electronic effects[74].

72

4.4.3 *External sugar-binding sites are not conserved in both AfGDH and AnGOX*

Among all the structurally characterized GMC oxidoreductases in complex with sugars, *Pc*GDH is the only member of the family displaying externally bound saccharides. The enzyme-substrate complex structures of *Af*GDH, *Mt*CDH and *Tm*P2O do not show any additional ligand on the external surface of the protein. To check if this structural feature might be conserved in *Af*GDH and *An*GOX, an analysis of the amino acids conservation in these regions was carried out.

Most of the residues forming SBS1 in *Pc*GDH are not conserved either in *Af*GDH or *An*GOX (Fig. 4.16). The backbone carbonyl groups involved in the interactions in *Pc*GDH are almost superimposable to those of *Af*GDH and *An*GOX, but the amino acid side chains differ significantly.
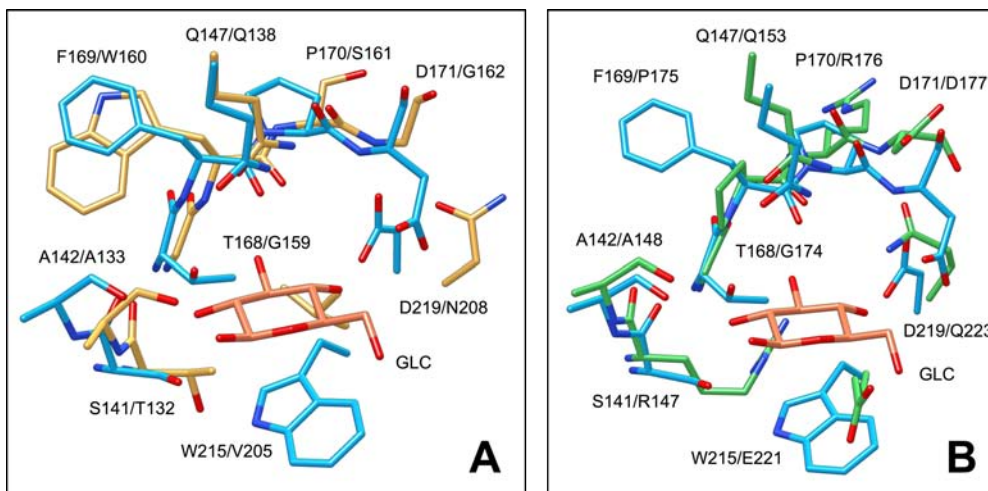

Figure 4.16. *Structural comparison between SBS1 in PcGDH-GLC and the corresponding residues in AfGDH (panel A) and in AnGOX (panel B). PcGDH is in light blue, AfGDH in ocher, AnGOX in light green, GLC in orange, colored by heteroatom. The first residue specified belongs to PcGDH, the second to AfGDH in panel A or to AnGOX in panel B.*

Thr168, which plays a key role in stabilizing the sugar O3 hydroxyl group, is substituted by Gly159 and Gly174 in *Af*GDH and *An*GOX respectively. Asp171 is not conserved in *Af*GDH (Gly162) but is conserved in *An*GOX (Asp177). The CH-π stabilization provided by Trp215 is not conserved, since Val205 is present in *Af*GDH and Glu221 in *An*GOX. Ser141 is not conserved in *An*GOX (Arg147) but a hydroxyl group able to interact through hydrogen bond is present in *Af*GDH (Thr132). Asp219, which interacts with sugar ligands bridged by a water molecule is not conserved in either proteins: Asn208 in *Af*GDH and Gln223 in *An*GOX; Asn208 has a very different orientation while the side chain of Gln223 might provide the same type of stabilization. Gln147 is conserved and occupies the same position in both the other proteins: Gln138 in *Af*GDH, Gln153 in *An*GOX.

Most residues found in SBS2 are conserved in *Af*GDH and *An*GOX (Fig. 4.17), and so are the positions of the backbone carbonyl groups involved in hydrogen bonds. Val232, however, is not conserved in *Af*GDH, where Tyr221 occupies most of the cavity that binds the sugar in *Pc*GDH. In *An*GOX Pro164, Tyr 228 and Val232 are not conserved (Thr170, Trp232 and Asn236 respectively).
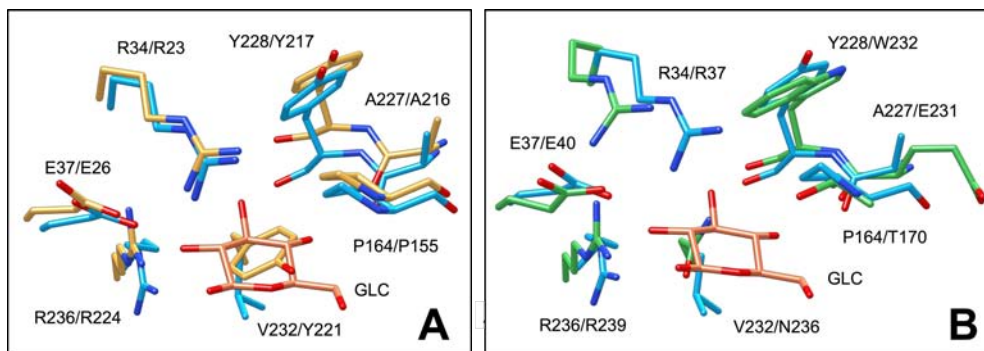

Figure 4.17. *Structural comparison between SB2 in PcGDH-GLC and the corresponding residues in AfGDH (panel A) and in AnGOX (panel B). PcGDH is in light blue, AfGDH in ocher, AnGOX in light green, GLC in orange, colored by heteroatom. The first residue specified belongs to PcGDH, the second to AfGDH in panel A or to AnGOX in panel B.*

Two out of the three interactions found in SBS3 involve the backbone carbonyl groups, which in *Af*GDH and *An*GOX occupy similar positions in comparison to *Pc*GDH (Fig. 4.18). The amino group stabilizing the O6 hydroxyl group of the sugar is conserved in *Af*GDH (Lys461) but not in *An*GOX, where the side chain of Asn473 is too short to interact with a potential ligand.

To sum up, the residues found in *Pc*GDH SBS1 are not conserved in *Af*GDH and *An*GOX, especially Thr168 and Trp215 that play an essential role in sugar stabilization. Many residues forming SBS2 are conserved, but the bulky Tyr221 occupies the pocket almost fully in *Af*GDH. Regarding SBS3, the limited number of interactions are conserved, although *An*GOX lacks the stabilizing lysine side chain. It is not possible to definitively assess the relevance of the SBSs found in *Pc*GDH just on the basis of structural comparison. Further experiments have to be carried out to clarify the physiological relevance of the external SBS in solution, which might be a peculiar feature of *Pc*GDH.
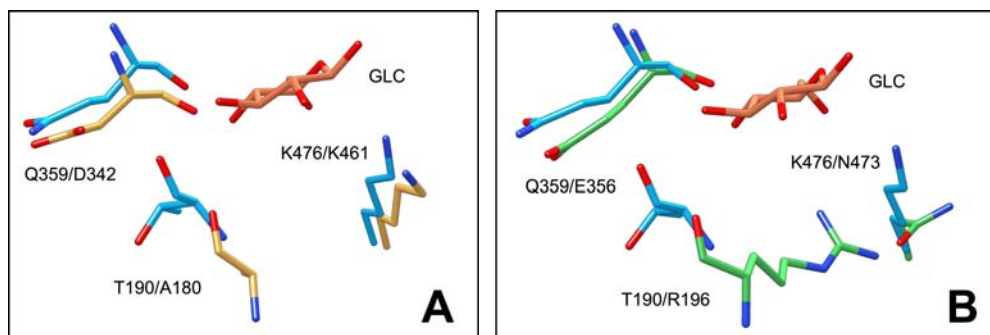

Figure 4.18. *Structural comparison between SBS3 in PcGDH-GLC and the corresponding residues in AfGDH (panel A) and in AnGOX (panel B). PcGDH is in light blue, AfGDH in ocher, AnGOX in light green, GLC in orange, colored by heteroatom. The first residue specified belongs to PcGDH, the second to AfGDH in panel A or to AnGOX in panel B.*

75

4.5 Structure-function analysis: evaluation of substrate specificity

In order to evaluate the possibility for *Pc*GDH to work as an oxidase, a preliminary standard ABTS test[90] was carried out. This assay allows to monitor the ability of the enzyme to use molecular oxygen as a secondary electron acceptor. The resulting production of hydrogen peroxide is coupled to the ABTS oxidation catalyzed by a peroxidase. A change in the optical signal of ABTS due to the redox reaction reveals the oxidase nature of the enzyme of interest. The results obtained by testing *Pc*GDH clearly indicate that the enzyme is not able to consume molecular oxygen, confirming the strict dehydrogenase nature of *Pc*GDH (data not shown).
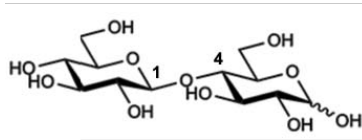
Analysis of the substrate specificity of *Pc*GDH was carried out using a set of five different substrates, shown in Fig. 4.19: D-glucose (GLC), cellobiose (CLB), laminaribiose (LMB), and two different glucose-based trisaccharides (GTA, GTB). Initially, the enzymatic activity of *Pc*GDH was tested only using glucose and laminaribiose, through the enzymatic assay based on the DCIP method described in par. 3.5. Consistently with the structural data, a major conversion of DCIP was observed when laminaribiose was used as a substrate (Fig. 4.20). Therefore, a more extensive enzymatic characterization using these two substrates was carried out; the resulting kinetic constants are shown in Table 4.2. *Pc*GDH showed a typical Michaelis-Menten kinetics over the substrate concentration range of 0.12-2 M for glucose and 0-115.5 mM for laminaribiose (Fig. 4.21). The measured K_M for D-glucose is 516 ± 176 mM while for laminaribiose is 44 ± 17 mM, suggesting that *Pc*GDH binds laminaribiose with more than a ten-fold higher affinity than glucose. The enzyme displays a comparable maximum turnover rate (k_{cat}) for the two substrates, although a slightly higher value was measured for laminaribiose. As a result, the overall catalytic efficiency (k_{cat}/K_M) is higher for this sugar.
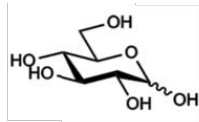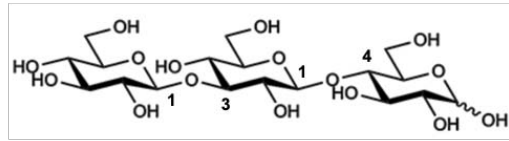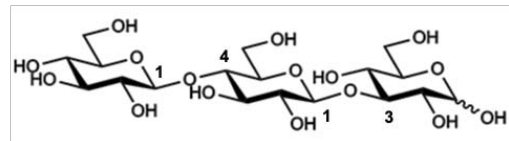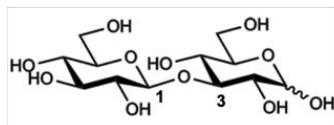
Figure 4.19. *Chemical structures of the five sugars employed to test the substrate specificity of PcGDH. Carbon atoms involved in glycosidic bonds are labelled.*

The apparent values of the kinetic parameters suggest that the enzyme rate of consumption of the two substrate is very similar, but *Pc*GDH preferentially binds laminaribiose.

The preferential activity of *Pc*GDH on disaccharides raised the question about a potential specificity toward the type of connection between the two glucosyl units. Therefore, in order to assess the enzyme ability to discriminate between $\beta(1\rightarrow3)$ and $\beta(1\rightarrow4)$ glycosidic bonds, *Pc*GDH activity was tested toward cellobiose, a disaccharide with a $\beta(1\rightarrow4)$ linkage: in this case no activity was detected, indicating that the enzyme preferentially recognizes the $\beta(1\rightarrow3)$ glycosidic bonds.

77

To confirm such specificity and to investigate the dependence on the substrate oligomerization state, PcGDH activity was tested toward the trisaccharides 1,3;1,4-β-glucotriose A (GTA) and B (GTB), shown in Figure 4.19. These glucose-based trisaccharides are structural isomers with opposite glycosidic linkages. In GTA the reducing glucosyl unit is linked through a β(1→4) bond to the second unit, which is involved in a β(1→3) bond with the terminal unit. These linkages are inverted in GTB, where the reducing unit is engaged in a β(1→3) bond, while a β(1→4) connection links the second and third units. As reported in Figure 4.20, only GTB acts as a substrate of PcGDH, while no enzymatic activity was detected toward GTA. Finally, the comparison of the relative activities revealed that among all the tested sugars laminaribiose is the preferred substrate of PcGDH (Fig. 4.20).
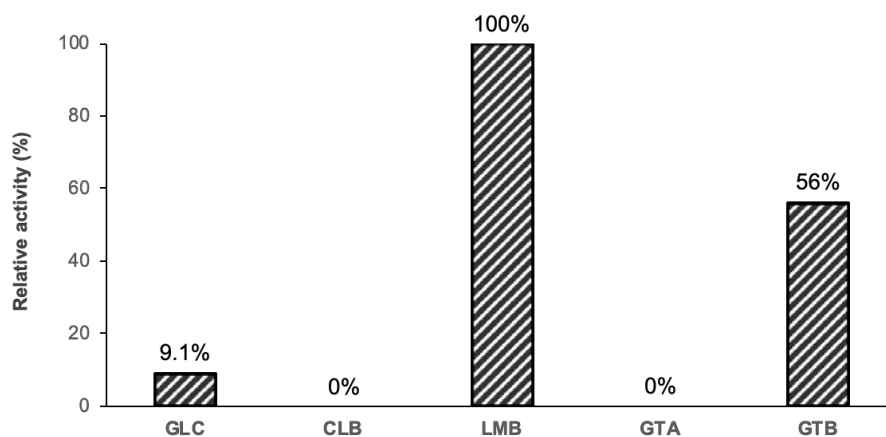

Figure 4.20. Histogram of PcGDH relative activity tested toward the five sugars shown in Fig. 4.19, using DCIP as secondary electron acceptor. Activity values are given in percentage relative to laminaribiose (LMB).

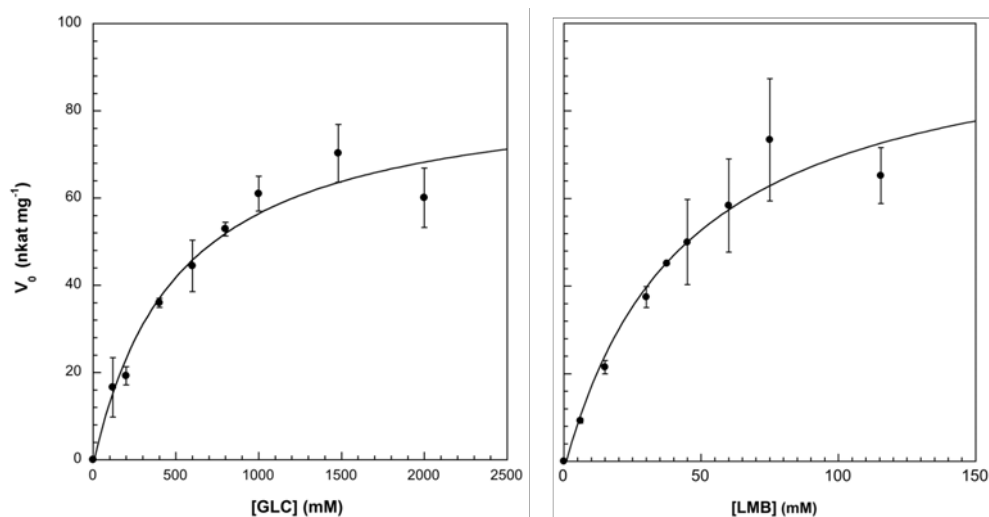|	K_M (mM)	V_{max} (nkat mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
D-glucose	516 ± 176	87 ± 9	6.1 ± 0.6	12 ± 2
Laminaribiose	44 ± 17	103 ± 14	7.2 ± 1.0	163 ± 35

Figure 4.21. *Initial velocity of the reaction as a function of the substrate concentration for glucose (GLC, left panel) and laminaribiose (LMB, right panel). Initial velocity points were determined applying a linear fit to the linear region of the time trace. Data were fit to the Michaelis-Menten hyperbolic equation (see par. 3.5). Error bars represent the standard deviation of initial velocities in three independent experiments.*

This finding can be interpreted considering the role of *Pc*GDH in its physiological context. The measured enzymatic activity is higher toward the disaccharide laminaribiose than toward the monosaccharide glucose, showing that a sugar possessing a β(1→3) type of connection is a preferential substrate of *Pc*GDH. The observation that also GTB can be oxidized by *Pc*GDH confirms that the catalytic activity of the enzyme depends on the type of glycosidic bond in which the reducing sugar monomer takes part; the enzyme retains activity also if an additional third terminal glucose molecule is linked to the laminaribiosyl portion through a β(1→4) bond, as in GTB.

79

Mixed β(1→3, 1→4) glucans are typical of the hemicellulose matrix and particularly abundant at the growth stage of the plant cell. Among all the possible products released during the cleavage of mixed β(1→3, 1→4) glucans, *Pc*GDH is thus able to oxidize glucose, laminaribiose and GTB; cellobiose and GTA instead cannot be oxidized by *Pc*GDH (Fig. 4.22).
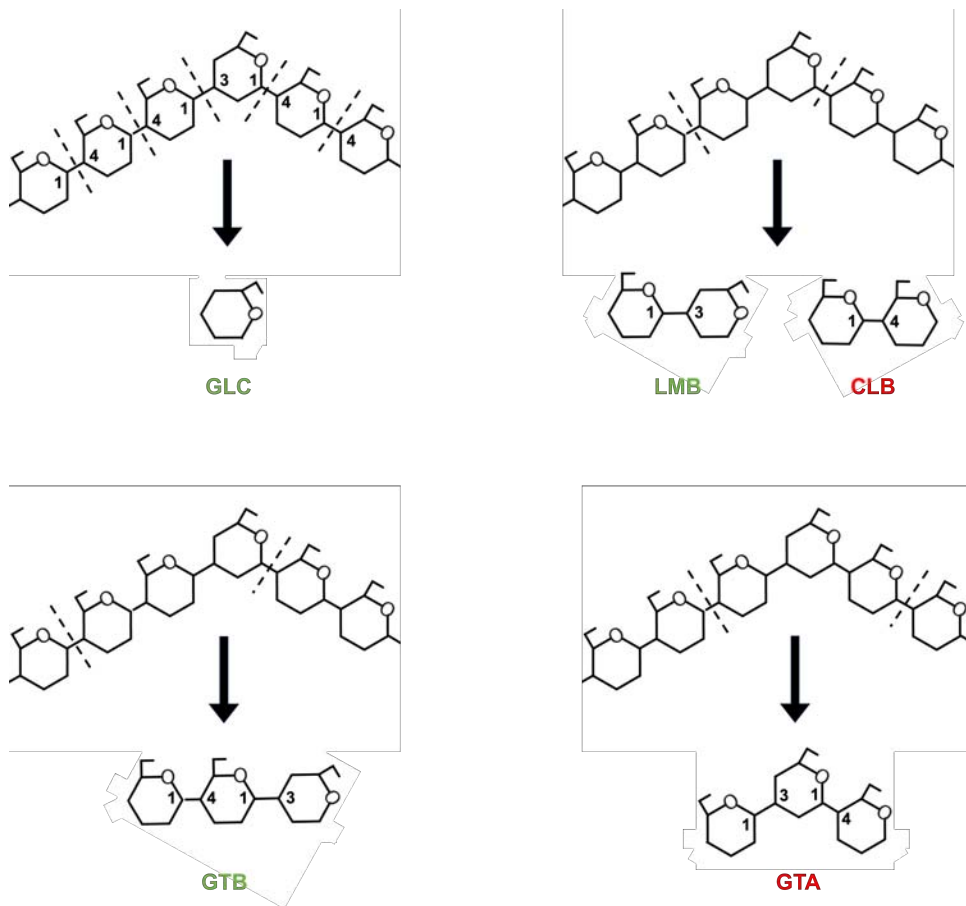
Figure 4.22. *Schematic representation of the possible cleavage products of mixed β(1→3, 1→4) glucans typical of hemicellulose. Depending on the type of cleavage performed by hydrolases or LPMOs, saccharides with different types of connection can be released. PcGDH is able to oxidize GLC, LMB and GTB (written in green), but it is not active on CLB and GTA (written in red). Carbon atoms of a typical polysaccharide and of the possible cleavage products are numbered.*

This might not constitute a problem for the fungus: other sugar dehydrogenases, such as CDH[91], have been identified in the genome of *P. cinnabarinus*, so the organism might have developed a different system to oxidize these sugars. Previous studies carried out by Piumi et al.[17] already demonstrated that *Pc*GDH is able to oxidize different sugar substrates, including mono and disaccharides, with glucose being the preferred one with a K_M compatible with our determination (272 mM). In this study, the structural analysis of glucose-bound *Pc*GDH suggested to try laminaribiose as an alternative substrate, and *in vitro* enzymatic assays revealed it to be the preferred one.

This substrate promiscuity can be structurally justified noticing that the primary interactions involved in the enzyme-substrate complex formation are CH-π interactions, responsible for orienting the substrate pyranose rings but unable to lock the hydroxyl groups in a unique fashion. In addition to that, we proved the enzyme ability to exploit the reducing potential of carbohydrates with different degrees of polymerization. The lack of specificity in terms of substrate oligomerization state allows *Pc*GDH activity to be fueled by a variety of saccharide chains deriving from the preliminary degrading action of hydrolases or LPMOs on larger polysaccharides, with no strict need for a complete cleavage, *i.e.* production of glucose monomers.

This can be interpreted as an aspect of the evolutionary optimization of the coupling between the oxidation of sugars and the reduction of radicals produced by laccases, resulting in a very efficient *Pc*GDH auxiliary activity in the overall process of lignin degradation. The preferential recognition of β(1→3) bonds might represent a peculiar feature of *Pc*GDH activity and a fungal adaptation to the substrate availability, since hemicellulose is much more accessible to lytic enzymes than cellulose, which forms crystalline fibers.

81

This allows the enzyme to be readily reduced in the polysaccharide rich environment where its activity is required.

These observations are consistent with the macroscopic evidence that *P. cinnabarinus* is a white-rot fungus able to perform a complete lignin degradation. The enzyme promiscuity and the preferential activity toward glycosidic bonds typical of hemicellulose can provide a molecular explanation for both the very efficient lignin degrading properties and the presence of residual white cellulose on the plant wood attacked by the fungus.

Chapter 5

Conclusions and future perspectives

5.1 *Pc*GDH has the features of an efficient auxiliary enzyme

The structural and functional characterization of *Pc*GDH described in this study provided an explanation for the enzyme promiscuity pointed out by previous work and allowed us to identify a preferred substrate.

This enzyme was identified as a glucose dehydrogenase by sequence homology with GOX and CDH and functional characterization, but it displays residual activity toward sugar substrates other than glucose. The structural analysis of the glucose-bound enzyme shed light on the molecular basis of substrate recognition and suggested the ability of *Pc*GDH to bind a disaccharide. Functional data on glucose, laminaribiose and glucose-based trisaccharides demonstrated that the enzyme activity is higher toward oligosaccharides containing a $\beta(1\rightarrow3)$ type of connection in the reducing unit. The three-dimensional structure of *Pc*GDH in complex with laminaribiose confirms the substrate binding mechanism suggested by the glucose-bound structure.

Mixed $\beta(1\rightarrow3, 1\rightarrow4)$ glucans are part of the hemicellulose matrix of the plant cell wall, where they associate with cellulose microfibrils during cell growth. The ability to oxidize sugar substrates with different stereochemistry and different polymerization degree strongly supports an auxiliary activity for *Pc*GDH. Indeed, both an easily reducible FAD cofactor and a substrate promiscuity confer to *Pc*GDH the ability to readily reduce lignin-derived radicals produced by the action of laccases, which is essential for a powerful lignin degrading system.

83

Some aspects of *Pc*GDH activity, however, still need to be clarified. The unoccupied space found in the active site suggests that the enzyme might be able to recognize sugar substrates with alternative types of linkages, such as $\beta(1\rightarrow2)$ or $\beta(1\rightarrow6)$ glycosidic bonds, or with a lower number of carbon atoms in the ring, *i.e.* pentose sugars. Moreover, since *Pc*GDH belongs to the secretome of the fungus, we propose that the presence of external sugar binding sites could play a role in sensing the substrate availability and directing the enzymatic activity toward the lignocellulose matrix.

5.2 Future perspectives

The study described in this thesis unveils novel, peculiar features of *Pc*GDH activity, but it also raises some questions.

For example, the substrate promiscuity and its molecular basis are not fully explored; we plan to obtain the structure of *Pc*GDH bound to different sugar substrates, with a focus on the unoccupied space in proximity to the FAD cofactor. Determining the structure with monosaccharides, like xylose, might allow us to identify other sugars or glycosidic bonds able to better fit within the active site, since a structural approach has already been successful for the identification of laminaribiose as the preferred substrate. It is not excluded that the *Pc*GDH active site might be able to accept alternative sugar substrates with an even higher affinity.

Structural analysis with disaccharides that already proved to act as substrates of *Pc*GDH, like lactose and maltose, would complement the information about the substrate binding mode and would provide more detail on the specificity of the enzyme-substrate complex.

Another structural feature of *Pc*GDH requiring further investigation is the presence of external sugar-binding sites. Equilibrium binding experiments with sugar ligands that do not act as substrates of *Pc*GDH would allow a deeper understanding of the physiological relevance of these sites; isothermal titration calorimetry coupled to mutagenesis experiments might provide a molecular rationale for the process of sugar recognition, highlighting a potential role of the external sites in allosteric regulation of the binding events. Co-crystallization experiments might be useful to reveal structural changes induced by binding at those sites, whose effects might not be detectable in enzyme-ligand structures obtained through soaking experiments.

Lastly, a more detailed understanding of *Pc*GDH mechanism of action would pave the way for i) the improvement of existing biotechnological applications, such as the design of a glucose-specific dehydrogenase for the development of alternative blood glucose monitoring systems and ii) the design and production of an engineered *Pc*GDH to be utilized in the modification of compounds deriving from plant biomass, in the ambitious pursuit of mastering both the detoxification of lignin-derived compounds and the production of precursor molecules for applications in industry and pharmaceutics.

References

[1] Ferri, S., Kojima, K., and Sode, K. (2011) Review of Glucose Oxidases and Glucose Dehydrogenases: A Bird's View of Glucose Sensing Enzymes, *Journal of Diabetes Science and Technology 5*, 1068-1076.

[2] Cavener, D. R. (1992) GMC Oxidoreductases. A Newly Defined Family of Homologous Proteins with Diverse Catalytic Activities, *Journal of Molecular Biology 223*, 811-814.

[3] Sützl, L., Foley, G., Gillam, E. M. J., Bodén, M., and Haltrich, D. (2019) The GMC superfamily of oxidoreductases revisited: analysis and evolution of fungal GMC oxidoreductases, *Biotechnology for Biofuels 12*, 1-18.

[4] Pazur, J. H., and Kleppe, K. (1964) The Oxidation of Glucose and Related Compounds by Glucose Oxidase from *Aspergillus niger*, *Biochemistry 3*, 578-583.

[5] Adams, E. C., Mast, R. L., and Free, A. H. (1960) Specificity of glucose oxidase, *Archives of Biochemistry and Biophysics 91*, 230-234.

[6] Bright, H. J., and Appleby, M. (1968) The pH Dependence of the Individual Steps in the Glucose Oxidase Reaction, *The Journal of Biological Chemistry 244*, 3625-3634.

[7] Wong, C. M., Wong, K. H., and Chen, X. D. (2008) Glucose oxidase: natural occurrence, function, properties and industrial applications, *Applied Microbiology and Biotechnology 78*, 927-938.

[8] Coulthard, C. E., Michaelis, R., Short, W. F., Sykes, G., Skrimshire, G. E. H., Standfast, A. F. B., Birkinshaw, J. H., and Raistrick, H. (1942) Notatin: an Anti-bacterial Glucose-aerodehydrogenase from *Penicillium notatum* Westling, *Nature 150*, 634-635.

[9] Ohashi, K., Natori, S., and Kubo, T. (1999) Expression of amylase and glucose oxidase in the hypopharyngeal gland with an age-dependent role change of the worker honeybee (Apis mellifera L.), *European Journal of Biochemistry 265*, 127-133.

[10] Shin, K., Youn, H., Han, Y., Kang, S., and Hah, Y. (1993) Purification and characterisation of D-glucose oxidase from white-rot fungus *Pleurotus ostreatus*, *European Journal of Biochemistry 215*, 747-752.

[11] Kang, S., Shin, K., Han, Y., Youn, H., and Hah, Y. (1993) Purification and characterisation of an extracellular peroxidase from white-rot fungus *Pleurotus ostreatus*, *Biochimica et Biophysica Acta 1163*, 158-164.

[12] Ogura Y., N. M. (1937) Untersuchungen über die Atmung und die Dehydrasesysteme von *Aspergillus oryzae*, *Bot. Mag. Tokyo 51*, 597-612.

[13] Bak, T. (1967) Studies on glucose dehydrogenase of *Aspergillus oryzae*. Purification and physical and chemical properties, *Biochimica et Biophysica Acta 139*, 277-293.

[14] Bak, T. (1967) Studies on glucose dehydrogenase of *Aspergillus oryzae*. General enzymatic properties, *Biochimica et Biophysica Acta 146*, 317-327.

[15] Yang, Y., Huang, L., Wang, J., and Xu, Z. (2014) Expression, characterization and mutagenesis of an FAD-dependent glucose dehydrogenase from *Aspergillus terreus*, *Enzyme and Microbial Technology 68*, 2015.

[16] Sygmund, C., Staudigl, P., Klausberger, M., Pinotsis, N., Djinović-Carugo, K., Gorton, L., Haltrich, D., and Ludwig, R. (2011) Heterologous overexpression of *Glomerella cingulata* FAD-

dependent glucose dehydrogenase in *Escherichia coli* and *Pichia pastoris*, *Microbial Cell Factories 10*, 1-9.

[17] Piumi, F., Levasseur, A., Navarro, D., Zhou, S., Mathieu, Y., Ropartz, D., Ludwig, R., Faulds, C. B., and Record, E. (2014) A novel glucose dehydrogenase from the white-rot fungus *Pycnoporus cinnabarinus*: production in *Aspergillus niger* and physicochemical characterization of the recombinant enzyme, *Applied Microbiology and Biotechnology 98*, 10105-10118.

[18] Sode, K., Loew, N., Ohnishi, Y., Tsuruta, H., Mori, K., Kojima, K., Tsugawa, W., La Belle, J. T., and Klonoff, D. C. (2017) Novel fungal FAD glucose dehydrogenase derived from *Aspergillus niger* for glucose enzyme sensor strips, *Biosensors and Bioelectronics 87*, 305-311.

[19] Mori, K., Nakajima, M., Kojima, K., Murakami, K., Ferri, S., and Sode, K. (2011) Screening of *Aspergillus*-derived FAD-glucose dehydrogenases from fungal genome database, *Biotechnology Letters 33*, 2255-2263.

[20] Sakai, G., Kojima, K., Mori, K., Oonishi, Y., and Sode, K. (2015) Stabilization of fungi-derived recombinant FAD-dependent glucose dehydrogenase by introducing a disulfide bond, *Biotechnology Letters 37*, 1091-1099.

[21] Sygmund, C., Klausberger, M., Felice, A. K., and Ludwig, R. (2011) Reduction of quinones and phenoxy radicals by extracellular glucose dehydrogenase from *Glomerella cingulata* suggests a role in plant pathogenicity, *Microbiology 157*, 3203-3212.

[22] Levasseur, A., Lomascolo, A., and et al. (2014) The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model

with a versatile arsenal for lignocellulosic biomass breakdown, *BMC Genomics 15*, 1-24.

[23] Keegstra, K. (2010) Plant Cell Walls, *Plant Physiology 154*, 483-486.

[24] Somerville, C. (2006) Cellulose Synthesis in Higher Plants, *Annual Review of Cell and Developmental Biology 22*, 53-78.

[25] Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses, *Annual Review of Plant Biology 61*, 263-289.

[26] Li, X., and Chapple, C. (2010) Understanding Lignification: Challenges Beyond Monolignol Biosynthesis, *Plant Physiology 154*, 449-452.

[27] Janusz, G., Pawlik, A., Sulej, J., Swiderska-Burek, U., Jarosz-Wilkołazka, A., and Paszczynski, A. (2017) Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution, *FEMS Microbiology Reviews 41*, 941-962.

[28] Payne, C. P., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal Cellulases, *Chemical Reviews 115*, 1308-1448.

[29] Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V. G. H. (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides, *Science 330*, 219-222.

[30] Kracher, D., Scheiblbrandner, S., Felice, A. K. G., Breslmayr, E., Preims, M., Ludwicka, K., Haltrich, D., Eijsink, V. G. H., and Ludwig, R. (2016) Extracellular electron transfer systems fuel cellulose oxidative degradation, *Science 352*, 1098-1101.

[31] Shraddha, Shekher, R., Sehgal, S., Kamthania, M., and Kumar, A. (2011) Laccase: microbial sources, production, purification, and potential biotechnological applications, *Enzyme Research 2011*, 1-11.

89

[32] Pollegioni, L., Tonin, F., and Rosini, E. (2015) Lignin-degrading enzymes, *FEBS Journal 282*, 1190-1213.

[33] Sützl, L., Laurent, C. V. F. P., Abrera, A. T., Schütz, G., Ludwig, R., and Haltrich, D. (2018) Multiplicity of enzymatic functions in the CAZy AA3 family, *Applied Microbiology and Biotechnology 102*, 2477-2492.

[34] Warburg, O., and Christian, W. (1933) Über das gelbe Ferment und seine Wirkungen, *Biochem. Zeit. 266*, 377-411.

[35] Walsh, C. (1980) Flavin Coenzymes: At the Crossroads of Biological Redox Chemistry, *Accounts of Chemical Research 13*, 148-155.

[36] Beinert, H. (1957) Evidence for an intermediate in the oxidation-reduction of flavoproteins, *Journal of Biological Chemistry 225*, 465-478.

[37] Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics - a genomic and structural view of flavin-dependent proteins, *FEBS Journal 278*, 2625-2634.

[38] Lomascolo, A., Uzan-Boukhris, E., Herpoël-Gimbert, I., Sigoillot, J., and Lesage-Meessen, L. (2011) Peculiarities of *Pycnoporus* species for applications in biotechnology, *Applied Microbiology and Biotechnology 92*, 1129-1149.

[39] Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic Acids Research 37*, D233-D238.

[40] Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes, *Biotechnology for Biofuels 6*, 1-14.

[41] Syed, K., Shale, K., Pagadala, N. S., and Tuszynski, J. (2014) Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi, *PLoS ONE 9*, e86683.

[42] Salame, T. M., Knop, D., Levinson, D., Yarden, O., and Hadar, Y. (2013) Redundancy among manganese peroxidases in *Pleurotus ostreatus*, *Applied and Environmental Microbiology 79*, 2405-2415.

[43] Ramos, J. A., Barends, S., Verhaert, R. M., and de Graaff, L. H. (2011) The *Aspergillus niger* multicopper oxidase family: analysis and overexpression of laccase-like encoding genes, *Microbial Cell Factories 10*, 1-11.

[44] Hawksworth, D. L., and Lucking, R. (2017) Fungal diversity revisited: 2.2 to 3.8 million species, *Microbiological Spectrum 5*, 79-95.

[45] Yoo, E. H., and Lee, S. Y. (2010) Glucose Biosensors: An Overview of Use in Clinical Practice, *Sensors 10*, 4558-4576.

[46] Tsujimura, S., Kojima, S., Kano, K., Ikeda, T., Sato, M., Sanada, H., and Omura, H. (2006) Novel FAD-dependent glucose dehydrogenase for a dioxygen-insensitive glucose biosensor, *Bioscience, Biotechnology, and Biochemistry 70*, 654-659.

[47] Chaubey, A., and Malhotra, B. D. (2002) Mediated biosensors, *Biosensors and Bioelectronics 17*, 441-456.

[48] Okurita, M., Suzuki, N., Loewb, N., Yoshida, H., Tsugawa, W., Mori, K., Kojima, K., Klonoff, D. C., and Sode, K. (2018) Engineered fungus derived FAD-dependent glucose dehydrogenase with acquired ability to utilize hexaammineruthenium(III) as an electron acceptor, *Bioelectrochemistry 123*, 62-69.

[49] Hatada, M., Loewb, N., Inose-Takahashi, Y., Okuda-Shimazaki, J., Tsugawa, W., Mulchandani, A., and Sode, K. (2018) Development of

91

a glucose sensor employing quick and easy modification method with mediator for altering electron acceptor preference, *Bioelectrochemistry 121*, 185-190.

[50] Ito, K., Okuda-Shimazaki, J., Mori, K., Kojima, K., Tsugawa, W., Ikebukuro, K., Lin, C., La Belle, J., Yoshida, H., and Sode, K. (2019) Designer fungus FAD glucose dehydrogenase capable of direct electron transfer, *Biosensors and Bioelectronics 123*, 114-123.

[51] Hecht, H. J., Kalisz, H. M., Hendle, J., Schmid, R. D., and Schomburg, D. (1993) Crystal Structure of Glucose Oxidase from *Aspergillus niger* Refined at 2.3 Å Resolution, *Journal of Molecular Biology 229*, 153-172.

[52] Fernandez, I. S., Ruiz-Duenas, F. J., Santillana, E., Ferreira, P., MartInez, M. J., MartInez, A. T., and Romero, A. (2009) Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase, *Acta Crystallographica Section D Biological Crystallography 65*, 1196-1205.

[53] Wohlfahrt, G., Witt, S., Hendle, J., Schomburg, D., Kalisz, H. M., and Hecht, H. J. (1999) 1.8 and 1.9 Å resolution structures of the *Penicillium amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modelling substrate complexes, *Acta Crystallographica Section D Biological Crystallography 55*, 969–977.

[54] Yoshida, H., Sakai, G., Mori, K., Kojima, K., Kamitori, S., and Sode, K. (2015) Structural analysis of fungus-derived FAD glucose dehydrogenase, *Scientific Reports 5*, 1-13.

[55] Quaye, O., Lountos, G. T., Fan, F., Orville, A. M., and Gadda, G. (2008) Role of Glu312 in Binding and Positioning of the Substrate for the

Hydride Transfer Reaction in Choline Oxidase, *Biochemistry 47*, 243-256.

[56] Tan, T. C., Spadiut, O., Wongnate, T., Sucharitakul, J., Krondorfer, I., Sygmund, C., Haltrich, D., Chaiyen, P., Peterbauer, C. K., and Divne, C. (2013) The 1.6 Å Crystal Structure of Pyranose Dehydrogenase from *Agaricus meleagris* Rationalizes Substrate Specificity and Reveals a Flavin Intermediate, *PLoS ONE 8*, e53567.

[57] Hallberg, B. M., Leitner, C., Haltrich, D., and Divne, C. (2004) Crystal Structure of the 270 kDa Homotetrameric Lignin-degrading Enzyme Pyranose 2-Oxidase, *Journal of Molecular Biology 341*, 781-796.

[58] Swoboda, B. E. P. (1969) The mechanism of binding of flavin-adenine dinucleotide to the apoenzyme of glucose oxidase and evidence for the involvement of multiple bonds, *Biochimica et Biophysica Acta 175*, 380-387.

[59] Hol, W. G. J., van Duijnen, P. T., and Berendsen, H. J. C. (1978) The alpha-helix dipole and the properties of proteins, *Nature 273*, 443-446.

[60] Tan, T. C., Kracher, D., Gandini, R., Sygmund, C., Kittl, R., Haltrich, D., Hällberg, B. M., Ludwig, R., and Divne, C. (2015) Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation, *Nature Communications 6*, 1-11.

[61] Wongnate, T., and Chaiyen, P. (2013) The substrate oxidation mechanism of pyranose 2-oxidase and other related enzymes in the glucose-methanol-choline superfamily, *FEBS Journal 280*, 3009-3027.

[62] Kujawa, M., Ebner, H., Leitner, C., Hallberg, B. M., Prongjit, M., Sucharitakul, J., Ludwig, R., Rudsander, U., Peterbauer, C., Chaiyen, P., Haltrich, D., and Divne, C. (2006) Structural Basis for Substrate

93

Binding and Regioselective Oxidation of Monosaccharides at C3 by Pyranose 2-Oxidase, *Journal of Biological Chemistry 281*, 35104-35115.

[63] Gadda, G. (2008) Hydride Transfer Made Easy in the Reaction of Alcohol Oxidation Catalyzed by Flavin-dependent Oxidases, *Biochemistry 47*, 13745-13753.

[64] Hallberg, B. M., Henriksson, G., Pettersson, G., Vasella, A., and Divne, C. (2003) Mechanism of the Reductive Half-reaction in Cellobiose Dehydrogenase, *Journal of Biological Chemistry 278*, 7160-7166.

[65] Lario, P. I., Sampson, N., and Vrielink, A. (2003) Sub-atomic Resolution Crystal Structure of Cholesterol Oxidase: What Atomic Resolution Crystallography Reveals about Enzyme Mechanism and the Role of the FAD Cofactor in Redox Activity, *Journal of Molecular Biology 326*, 1635-1650.

[66] Lyubimov, A. Y., Lario, P. I., Moustafa, I., and Vrielink, A. (2006) Atomic resolution crystallography reveals how changes in pH shape the protein microenvironment, *Nature Chemical Biology 2*, 259-264.

[67] Kommoju, P. R., Chen, Z. W., Bruckner, R. C., Mathews, F. S., and Schuman Jorns, M. (2011) Probing Oxygen Activation Sites in Two Flavoprotein Oxidases Using Chloride as an Oxygen Surrogate, *Biochemistry 50*, 5521-5534.

[68] Mattevi, A. (2006) To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes, *Trends in Biochemical Sciences 31*, 276-283.

[69] Lyubimov, A. Y., Heard, K., Tang, H., Sampson, N. S., and Vrielink, A. (2007) Distortion of flavin geometry is linked to ligand binding in cholesterol oxidase, *Protein Science 16*, 2647–2656.

[70] Carugo, O., and Djinovic-Carugo, K. (2005) When X-rays modify the protein structure: radiation damage at work, *Trends in Biochemical Sciences 30*, 213-219.

[71] Orville, A. M., Lountos, G. T., Finnegan, S., Gadda, G., and Prabhakar, R. (2009) Crystallographic, spectroscopic, and computational analysis of a flavin C4a-oxygen adduct in choline oxidase, *Biochemistry 48*, 720-728.

[72] Sucharitakul, J., Prongjit, M., Haltrich, D., and Chaiyen, P. (2008) Detection of a C4a-Hydroperoxyflavin Intermediate in the Reaction of a Flavoprotein Oxidase, *Biochemistry 47*, 8485-8490.

[73] Carro, J., Martínez-Júlvez, M., Medina, M., Martínez, A. T., and Ferreira, P. (2017) Protein dynamics promote hydride tunnelling in substrate oxidation by aryl-alcohol oxidase, *Physical Chemistry Chemical Physics 19*, 28666-28675.

[74] Tan, T. C., Haltrich, D., and Divne, C. (2011) Regioselective Control of β-D-Glucose Oxidation by Pyranose 2-Oxidase Is Intimately Coupled to Conformational Degeneracy, *Journal of Molecular Biology 409*, 588-600.

[75] Ruiz-Dueñas, F. J., Ferreira, P., Martínez, M. J., and Martínez, A. T. (2006) In vitro activation, purification, and characterization of *Escherichia coli* expressed aryl-alcohol oxidase, a unique H2O2-producing enzyme, *Protein Expression and Purification 45*, 191-199.

[76] Shental-Bechor, D., and Levy, Y. (2008) Effect of glycosylation on protein folding: A close look at thermodynamic stabilization, *PNAS 105*, 8256-8261.

[77] Courjean, O., Gao, F., and Mano, N. (2009) Deglycosylation of Glucose Oxidase for Direct and Efficient Glucose Electrooxidation on a Glassy Carbon Electrode, *Angewandte Chemie 121*, 6011-6013.

95

[78] van Hartingsveldt, W., Mattern, I. E., van Zeijl, C. M., Pouwells, P. H., and van den Hondel, C. A. (1987) Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene, *Molecular and General Genetics 206*, 71-75.

[79] Punt, P. J., and van den Hondel, C. A. (1992) Transformation of filamentous fungi based on hygromycin B and phleomycine resistance markers, *Methods in Enzymology 216*, 447-457.

[80] Kabsch, W. (2010) XDS, *Acta Crystallographica Section D Biological Crystallography 66*, 125-132.

[81] Vagin, A., and Teplyakov, A. (1998) A translation-function approach for heavy-atom location in macromolecular crystallography, *Acta Crystallographica Section D Biological Crystallography 54*, 400-402.

[82] Collaborative Computational Project 4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallographica Section D Biological Crystallography 50*, 7690-7763.

[83] Pannu, N., Murshudov, G., Dodson, E., and Read, R. J. (1998) Incorporation of prior phase information strengthens maximum-likelihood structure refinement, *Acta Crystallographica Section D Biological Crystallography 54*, 1285-1294.

[84] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D Biological Crystallography 66*, 486-501.

[85] Winn, M., Isupov, M., and Murshudov, G. N. (2000) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta Crystallographica Section D Biological Crystallography 57*, 122-133.

[86] Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.

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

[88] Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallographica Section D Biological Crystallography 60*, 2256-2268.

[89] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis, *Computational Chemistry 25*, 1605-1612.

[90] Werner, W., Rey, H. G., and Wielinger, H. (1970) Über die Eigenschaften eines neuen Chromogens für die Blutzuckerbestimmung nach der GOD/POD-Methode, *Journal of Analitical Chemistry 252*, 224-228.

[91] Moukha, S. M., Dumonceaux, T. J., Record, E., and Archibald, F. S. (1999) Cloning and analysis of *Pycnoporus cinnabarinus* cellobiose dehydrogenase, *Gene 234*, 23-33.

Acknowledgments

I would like to express my gratitude to all the people who contributed to the production of this thesis and were part of my PhD experience.

Beatrice Vallone has been an amazing supervisor and the most important personality during my PhD. Thank you for introducing me to structural biology and for giving me a taste of the scientist I want to become one day. You have believed in my potential when the problems I had to solve seemed so much bigger than me. Your suggestions always led to success and guided me in understanding the nature of things from both a scientific and a human perspective.

Carmelinda Savino taught me X-ray crystallography and as if it was not enough, has been extremely supportive during the thesis development. Thank you for your supervision and your precious help. Also, a special thanks for your patience in dealing with an awfully exuberant PhD student.

Giuliano Sciara has given me the opportunity to work on GDH, introducing me to the fascinating world of fungal enzymes. Thank you for your help and your dedication, which made this thesis possible.

Filippo Mancia hosted me for one and a half year in New York and gave me the opportunity to learn the secrets of membrane protein biochemistry. Thank you for teaching me that being stubborn is a prerequisite for aspiring scientists.

Alberto Boffi allowed me to be part of the hemoglobin and ferritin projects. Thank you for giving me the opportunity to solve my first protein structure (Hb-DFO) and to explore the cryo-EM world.

Amédée Des Georges introduced me to cryo-EM at CUNY. Thank you for teaching me how to freeze grids and for sharing your precious cryo-EM tricks.

Alberto Macone taught me GC-MS. Thank you for your precious help and you support in SRD5α functional characterization.

Francesco Malatesta and Stefano Gianni were the coordinators during the PhD course. Thank you for training us PhD students and for the many opportunities you gave us to improve our speaking and writing skills in science in a multidisciplinary and international context.

Linda Celeste Montemiglio has been extremely supportive from every point of view, from purification issues to desperate phone calls. Thank you for helping me in a lot of different ways, and in making me realize how to overcome my own limits. Even if we fight sometimes, you know it is sugar sweet.

Cécile Exertier has been my partner in crime for the last three years. Thank you for handling my mood swings, sometimes successfully, sometimes not. I am glad I have not set your lab on fire, for now.

Giacomo Parisi has been my mentor at the very early stage of my experience in structural biology. Thank you for your support and for being my SRD5α buddy, through good and bad.

Antonella Scaglione has been a very precious friend, especially in the last year. Thank you for your optimism and your energy, in which I see my whole self.

Claudia Testi shared with me the stress of being a student abroad and helped me with cryo-EM issues. Thank you for our good chats and our desperate moments.

Elena Gugole helped me with all the functional experiments on GDH, so her work has been absolutely essential for this thesis. Thank you for being patient with me and good luck with your PhD.

Ida Freda has been the chill counterpart of the lab, an example of calm and respect, thank you and good luck with your PhD.

Ida Sannino had the misfortune of working with me for her Bachelor thesis, and I am glad she has been my Ida for three months.

Rhett Millsaps II helped me with the English grammar and taught me patience and meditation. Thank you for being awesome. One day I will be able to correctly pronounce a priest climbing up the stairs.

All my friends and roommates who lived in Rector Place, with whom I had so much fun in New York City.

H2CU College Italia has been extremely supportive, hosting me in Rector Place. Many thanks to all the people contributing to this initiative, who provide a cultural exchange platform of extraordinary level to young Italian students.

People from the Department of Biochemical Science "A. Rossi Fanelli", who shared with me synchrotron shifts, journal clubs and conferences.

People from Columbia University, including Wayne Hendrickson for his amazing crystallography classes and all the members of the Mancia Lab. A special thank you to Brianna, Jonathan and Sabrina, who will be in Rome for my defense.

All my closest friends: Alessandra, Antonello, Francesca, Giulia, Ilaria, Marta, Roberto and Valeria. Thank you for being the best thing about me.

My family, thank you for being always by my side.

In the end, thank you to everyone who has been part of my PhD experience for the last three years who, one way or another, consciously or not, contributed to who I am now.

Appendix

During the PhD course (2016-2019) the following manuscripts have been published:

1) Scaglione, A., Montemiglio, L.C., Parisi, G., Asteriti, I.A., Bruni, R., Cerutti, G., Testi, C., Savino, C., Mancia, F., Lavia, P. and Vallone, B. (2017) Subcellular localization of the five members of the human steroid 5α-reductase family, *Biochimie Open 4*, 99-106.

2) Parisi, G., Montemiglio, L.C., Giuffrè, A., Macone, A., Scaglione, A., Cerutti, G., Exertier, C., Savino, C. and Vallone, B. (2018) Substrate induced conformational change in cytochrome P450 OleP, *Faseb Journal 33*, 1787-1800.

3) Exertier, C., Freda, I., Montemiglio, L.C., Scaglione, A., Cerutti, G., Parisi, G., Anselmi, M., Savino, C. and Vallone, B. (2019) Proximal and distal control for ligand binding in neuroglobin: role of the CD loop and evidence for His64 gating, *Scientific Reports 9*, 5326, 1-9.

The following manuscript has been submitted to a scientific journal:

4) Kiraga, Ł.*, Cerutti, G.*, Braniewska, A., Strzemecki, D., Sas, Z., Boffi, A., Savino, C., Montemiglio, L.C., Turnham, D., Seaton, G., Bonamore, A., Clarkson, R., Dabkowski, A., Paisey, S., Rygiel, T. and Krol, M. (2019) Pharmacokinetics of human hemoglobin-DFO ^{89}Zr complex in mice determined by PET.

101

The following manuscript is currently in preparation:

5) Cerutti, G., Montemiglio, L.C., Gugole, E., Savino, C., Vallone, B. and Sciara, G. (2019) Identification of a preferential substrate in glucose dehydrogenase from *Pycnoporus cinnabarinus*: a "sugar sweet" structural approach.

Research paper

Subcellular localization of the five members of the human steroid 5α-reductase family

Antonella Scaglione [a,1], Linda Celeste Montemiglio [a,f,1], Giacomo Parisi [a,1], Italia Anna Asteriti [b], Renato Bruni [c], Gabriele Cerutti [a], Claudia Testi [d], Carmelinda Savino [b], Filippo Mancia [e], Patrizia Lavia [b], Beatrice Vallone [a,f,g,*]

[a] *Dept. of Biochemical Sciences, Sapienza University of Rome, P.le A.Moro 5, 00185, Rome, Italy*
[b] *CNR-IBPM, P.le A.Moro 5, 00185, Rome, Italy*
[c] *Center on Membrane Protein Production and Analysis (COMPPÅ), New York, NY, 10027, USA*
[d] *Center for Life Nano Science@Sapienza, IIT, V.le Regina Elena 291, Rome, I-00185, Italy*
[e] *Department of Physiology and Cellular Biophysics, Columbia University, New York, NY, 10032, USA*
[f] *Istituto Pasteur−Fondazione Cenci Bolognetti, Dept. of Biochemical Sciences, Sapienza University of Rome, Italy*
[g] *Italian Academy for Advanced Studies in America at Columbia University, USA*

Abstract

In humans the steroid 5α-reductase (SRD5A) family comprises five integral membrane enzymes that carry out reduction of a double bond in lipidic substrates: Δ^4-3-keto steroids, polyprenol and trans-enoyl CoA. The best-characterized reaction is the conversion of testosterone into the more potent dihydrotestosterone carried out by SRD5A1-2. Some controversy exists on their possible nuclear or endoplasmic reticulum localization.

We report the cloning and transient expression in HeLa cells of the five members of the human steroid 5α-reductase family as both N- and C-terminus green fluorescent protein tagged protein constructs. Following the intrinsic fluorescence of the tag, we have determined that the subcellular localization of these enzymes is in the endoplasmic reticulum, upon expression in HeLa cells. The presence of the tag at either end of the polypeptide chain can affect protein expression and, in the case of trans enoyl-CoA reductase, it induces the formation of protein aggregates.
© 2017 Published by Elsevier B.V. on behalf of Société Française de Biochimie et Biologie Moléculaire (SFBBM). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Steroid 5α-reductase; Polyprenol reductase; Trans-enoyl-CoA reductase; Subcellular localization

1. Introduction

In humans the steroid 5α-reductase family (SRD5A) comprises five membrane embedded enzymes that carry out the reduction of a double bond of lipidic substrates. They can be traced back to a common ancestor that split in three subfamilies (SRD5A1-2, SRD5A3 and TECR-TECRL). Members of the three subfamilies are found in all primary eukaryotic species from plant, amoeba and yeast up to vertebrates, indicating that the split arose in early eukaryota [1]. The duplication of an ancestral gene led to substrate diversification and acquisition of a central role in vertebrates in the context of the complex physiological roles of steroids.

SRD5A1 and SRD5A2 carry out the NADPH-dependent reduction of the Δ^4 group (double bond) of C-19 and C-21 steroids into 5α-stereoisomers [2]. This reaction yields dihydrotestosterone, a more potent derivative of testosterone [3]. The first identified role of dihydrotestosterone is to trigger the development of external genitalia in the male fetus, as

---

* Corresponding author. Dept. of Biochemical Sciences, Sapienza University of Rome, P.le A.Moro 5, 00185, Rome, Italy.

E-mail address: beatrice.vallone@uniroma1.it (B. Vallone).
[1] These authors equally contributed to this work.

dramatically brought to evidence by the identification of familial incomplete male pseudo-hermaphroditism, a congenital condition due to mutations of the SRD5A2 gene [4]. Due to their activity on testosterone, both SRD5A1 and SRD5A2 can be viewed as pivotal factors in several physio-pathological conditions, from benign prostatic hyperplasia and prostatic cancer to acne and male pattern baldness [5,6]. Two commercial drugs, finasteride and dutasteride, are used in therapy for benign prostatic hyperplasia and, in the case of finasteride, also for male pattern baldness.

The action of SRD5A1 and SRD5A2 on progesterone, testosterone and other related steroids is also a committing step for the production of neurosteroids. These are among the strongest modulators of GABA(A) receptors with anticonvulsant, antidepressant and anxiolytic effects [7].

The expression of SRD5A1 and SRD5A2 was studied in several organs and tissues, with results depending on the methodology (antibody, mRNA analysis, *in situ*, after protein extraction, etc.) and on tissue preparation and origin (for a review see Ref. [8]). Expression of SRD5A1 is consistently reported in the brain, liver and sebaceous glands, whereas SRD5A2 is mainly found in the urogenital tract, genital skin, liver and transiently in the brain.

A third member, SRD5A3, has testosterone reductase activity, and seems to play a role in the onset of hormone refractory prostate cancer [9]. Surprisingly, inactivating mutations of the SRD5A3 gene cause a rare congenital neurodevelopmental disorder. This is possibly related to the fact that SRD5A3 is also endowed with polyprenol reductase activity, thus carrying out one of the earliest steps of protein N-linked glycosylation [10]. SRD5A3 was ubiquitously detected in human healthy tissues and is overexpressed in prostate and breast malignancies [11].

Trans-2-Enoyl-CoA reductase (TER or TECR) comes as a more recent addition to the SRD5A family. This gene, when mutated, causes a rare autosomic syndrome, nonsyndromic mental retardation, and the protein carries out the fourth step of very long-chain fatty acid synthesis and sphingosine degradation [12]. TECR expression seems ubiquitous in man and mouse organs by Northern blot analysis [13], with high expression in the nervous system [14] consistently with its involvement in the synthesis of membrane lipids and with congenital neurological impairment arising from mutations.

A fifth gene of unknown function, called TECR-like (TECRL), was identified in vertebrates through sequence homology with TECR [1].

During over 30 years of research on SRD5A enzymes, their purification in an active form has proven elusive. Nevertheless, biochemical studies on the microsomal fraction from liver and prostate led to the characterization of the main enzymatic parameters of SRD5A1, SRD5A2 [15] and, to a lesser extent, SRD5A3 proteins [8,16].

Sequence analysis and biochemical studies of the SRD5A family members did not highlight, to our knowledge, the presence of post-translational modifications, such as specific proteolytic signals, glycosylation, phosphorylation, acylation or isoprenylation. Similarly, no canonical signal for a specific subcellular localization was reported.

The subcellular localization of the SRD5A enzymes is significant to their functions since it might affect binding of reaction products to nuclear steroid receptors in the same cell or diffusion to other cells. We have undertaken a systematic analysis to address their localization. We have cloned all members of the human SRD5A family in vectors for transient expression in mammalian cells, fusing at either C- or N-terminus the enhanced green florescent protein (eGFP). We present a comprehensive assessment of their localization as a guideline for functional analysis in cells and for production and purification for structural and biochemical studies.

2. Material and methods

2.1. Cloning and plasmids

Plasmids containing the cDNA sequence for human SRD5A1, SRD5A2, SRD5A3 and TECR were purchased from the IMAGE Consortium (www.imageconsortium.org), the sequences of the human TECR and TECRL genes were synthesized.

The five genes were cloned into the pTriEx modifed vectors pNYCOMPS_RB_30_GFP and pNYCOMPS_RB_33_GFP, using the ligase independent cloning method described in Ref. [17]. These vectors are designed for mammalian, insect and bacterial protein expression and they add at the N-terminus of the protein coded by the cloned gene a cassette containing sequentially: ten histidines tag, FLAG tag, eGFP and TEV protease cleavage site at the amino terminus of the gene of interest for pNYCOMPS_RB_33_GFP. For the pNYCOMPS_RB_30_GFP the C-term is followed by: TEV protease cleavage site, eGFP, 10 histidine tag and FLAG-tag (see Fig. 1). These plasmids allow detection and expression of the cloned protein upon transient transfection of mammalian or insect cells, or expression in bacteria.

2.2. Cell culture and transfection

HEK293 and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, 2% L-glutamine, and 2% penicillin/streptomycin in humidified atmosphere at 37 °C in 5% CO_2. Confluent cell layers were trypsinized, and replated on poly-L-lysine coated coverslips in DMEM medium, which was replaced by serum-free media the next day.

Transfections were performed using Lipofectamine 2000, all controls were transfected with peGFP-N1 vector. Cells were harvested 48 h after transfection and analyzed by Western blot or immunofluorescence. The transfection rate was calculated after staining the cells on the coverslip with DAPI and counting eGFP-positive cells/DAPI-stained nuclei.
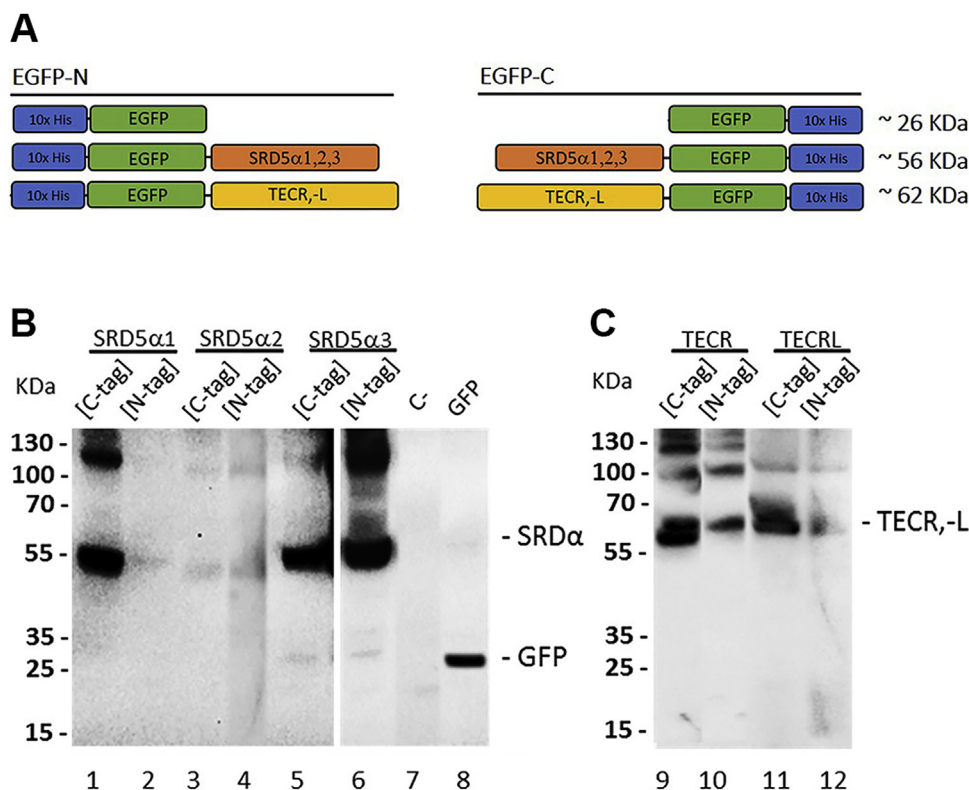
Fig. 1. **Schematic cartoon and expression of human SRD5As, TECR and TECRL eGFP-fusion proteins**. (A) Diagram of SRD5A constructs showing position of eGFP both at N- or C-term; the FLAG tag is omitted. (B) Extract from HeLa cells transfected with human SRD5A1, SRD5A2, SRD5A3 expression plasmids for C-term and N-term eGFP fusions, untransfected and eGFP-transfected cells. Different exposure times are shown (50 s for lanes 1–5, 25 s for lanes 6–8). (C) Extract from HeLa cells transfected with human TECR, TECRL (C-term and N-term eGFP fused). Proteins were visualized using anti-eGFP antibody after 10 s exposure (lanes 9–12).

2.3. Western blot

Total protein lysate was extracted with Normal Salt Buffer (20 mM Hepes, pH 7.0, 200 mM NaCl, 2 mM 2-Mercaptoethanol) supplemented with 1% β-D-Maltopyranoside and protease inhibitor cocktail (Roche). After incubation for 90 min at 4 °C, the lysates were clarified by centrifugation at 12,000×g at 4 °C for 20 min. Protein concentration was determined by the Bradford method. Equal amounts of proteins (40 μg) were loaded on 4–12% SDS-PAGE. Samples were electrotransferred onto nitrocellulose membranes. The membranes were blocked with 5% milk in TBS/0.1% Tween 20 (TBST) on an orbital shaker for 1 h and incubated for 1 h with rabbit monoclonal anti-GFP (1:1000, Roche). Following five TBST washes, the membranes were incubated for 1 h with horseradish peroxidase-conjugated anti-rabbit antibody in TBST. Membranes were developed by enhanced chemiluminescence, and detected proteins were identified by their molecular sizes.

2.4. Immunofluorescence and microscopy

Cells grown on coverslips were washed in PBS and fixed at room temperature by: (a) 3.7% PFA/30 mM sucrose in PBS, 10 min, (b) 5 min of membrane permeabilization in 0.1% TritonX-100 in PBS, and (c) 10 min in 0, 1 M Glycine. Blocking and incubations were performed in PBS/0.05% Tween 20/3% BSA. Incubation with primary antibodies was performed for 16 h at 4 °C. Secondary antibodies were incubated for 45 min at room temperature. Cells were counterstained with 4,6-diamidino-2-phenylindole (DAPI, 0.1 μg/ml) and mounted using Vectashield (Vector Laboratories).

Primary antibodies were: mouse anti-Erp57 (1:200, ab13506, Abcam), rabbit anti-Giantin (1:500, ab24586; Abcam), rabbit anti-pericentrin (2 μg/ml, ab4448; Abcam) and mouse anti-α-tubulin (1:2000, B-5-1-2, Sigma–Aldrich). Secondary antibodies were conjugated to Cy3, Texas red, FITC or Alexa647 (Vector Laboratories). Samples were analyzed using a Nikon Eclipse 90i microscope with a Qicam Fast 1394 CCD camera (QImaging). Image acquisition, deconvolution and Extended Depth of Focus on Z-serial optical sections were performed using Nis-Elements AR4.2 (Nikon); images were further processed with Adobe Photoshop CS 8.0.

The ProteoStat® aggresome detection kit was used to detect protein aggregates in HeLa cells transfected for SRD5A protein expression. Samples were analyzed using an Olympus FLUOVIEW FV10i-DOC Confocal Laser Scanning Microscope, equipped with a 60× phase contrast water-immersion objective (N.A.1.2) and an analog detector (PMT), using a

Texas Red filter set for the ProteoStat® dye, and an Midoriishi-Cyan filter set for the eGFP. Image acquisition and analysis were performed using an Image Processing and Analysis in Java software (ImageJ); images were further processed with Adobe Photoshop CS 8.0.

3. Results

3.1. Protein expression

Protein expression was assessed by immunoblot on detergent-solubilized whole lysates of HeLa cells transfected with the five members of the human SRD5A family. The antibody recognized eGFP fused either at the N- or at the C-term. We observed expression for all constructs, albeit with significant quantitative variations, as high as 100-fold (Fig. 1). SRD5A2, tagged at C- or N-term, and SRD5A1 when tagged at the N-term, show faint bands by immunoblot analysis. In parallel, we observed on duplicate slides that their transfection efficiency was only between 5% and 2%, thus explaining the weak signal detected in Western blots. The not transfected control shows no reaction with the anti-GFP antibody as expected. In extract from cells transfected with the vector for eGFP expression, the antibody recognizes a single polypeptide with the correct molecular weight (~27 kDa). Intensely stained bands, of apparent molecular weight corresponding to eGFP-fused SRD5A proteins (~56 kDa), are evident for cells transfected with C-term tagged eGFP-SRD5A1 (Fig. 1; lane 1), and with SRD5A3 eGFP-tagged at either end (Fig. 1; lanes 5 & 6). On the other hand, TECR and TECRL have a higher expected molecular weight (~62 kDa) and consistently their transient expression yielded slower-migrating bands on the immunoblot (Fig. 1; lanes 9−12).

In all samples, the anti-GFP antibody also recognized higher molecular weight species, which we attribute to an oligomeric state of SRD5A membrane proteins, which were not dissociated in SDS PAGE due to their very high hydrophobic character. Indeed, no high molecular weight signals were detected in control samples expressing eGFP not fused to SRD5A proteins.

3.2. Subcellular localization of eGFP tagged SRD5A proteins

The localization of each eGFP-tagged protein was examined relative to relevant markers of the membrane compartments, i.e. the ER resident protein Erp57 and the Golgi compartment resident protein Giantin [18−20]. Due to the fact that preliminary assays depicted diversified patterns for these markers, especially Giantin, in individual cells, we first evaluated the distribution of ERp57 and Giantin in co-staining with cell cycle markers, i.e. the centrosomal protein pericentrin (to identify interphase stages: one spot, G1 phase; two mature spots, late S and G2 phases) and the microtubule constituent α-tubulin (to identify mitosis). This allowed us to exclude that Erp57 and Golgi distribution patterns undergo significant remodeling during the cell cycle

and avoid inaccurate evaluation of SRD5As co-localization (Fig. S1).

Fig. 2A shows that N-term and C-term eGFP-tagged SRD5A1, SRD5A2 and SRD5A3 proteins expressed upon transfection reside in the ER, as indicated by co-localization with Erp57 and the overall distribution of the green fluorescence relative to Erp57. No nuclear localization was observed. On the other hand, eGFP alone localizes in the nucleus (Fig. S1). These observations are consistent with the ER localization of SRD5A proteins being genuine and not attributable to the presence of the eGFP tag.

Also for TECR and TECRL members, both the N- and C-term eGFP-tagged proteins, with one exception, showed a pattern and co-localization behavior corresponding to ER membrane insertion (Fig. 2B). A remarkable exception was observed in cells expressing the C-term eGFP-tagged TECR, where only sparse bright dots were detected, moreover the localization of the ER marker Erp57 was less diffuse than in non-transfected cells (Fig. S1). In some cases, there was some limited overlap between Erp57 and C-term eGFP-tagged TECR proteins (individual z-stack scanning are given in Fig. S2). On the other hand, the lack of co-localization with Giantin for TECR and TECRL, in both the C- and N-term eGFP-tagged proteins (Fig. 3A), ruled out the possibility that they might localize to the Golgi.

Staining with ProteoStat® that highlights the presence of aggresomes and protein aggregates [21], showed that the green fluorescent dots observed for C-term eGFP-tagged TECR form foci that are contained within larger aggregates, providing evidence consistent with the fact that the presence of eGFP at the C-term of TECR promotes protein misfolding, at least in part, resulting in aggregation (Fig. 3B).

4. Discussion

The subcellular localization of members of the SRD5A family in humans and also in mouse is controversial, especially as far as SRD5A1 and SRD5A2 are concerned. Russell and Wilson reported indications of a differential organ-specific subcellular localization, nuclear in the prostate and cytoplasmic in the liver, regardless of the enzyme variant (SRD5A1 or SRD5A2) [15,22], or that the localization for both is in the endoplasmic reticulum [23]. Early findings based mainly on cellular component fractionation [24,25], but later also on immunohistochemical studies [26,27], indicate that the SRD5A1 enzyme localization is mainly within nuclear or perinuclear membranes whereas the SRD5A2 protein is found in the cytoplasm, most likely in the ER.

Protein expression in mammalian (human HeLa and HEK293 and simian COS-M6), as well as yeast and insect cells, again yielded contradictory results. Some authors have reported that human SRD5A1 is associated with nuclear fractions upon expression in insect Sf9 cells [28,29] but later papers [30] have instead reported retrieval of enzyme activity from the microsomal fraction. Expression of SRD5A2 and SRD5A1 in HEK293 and in LNGK-9 PCa cells yielded activity in the nuclear and microsomal fractions [31,32]. Upon
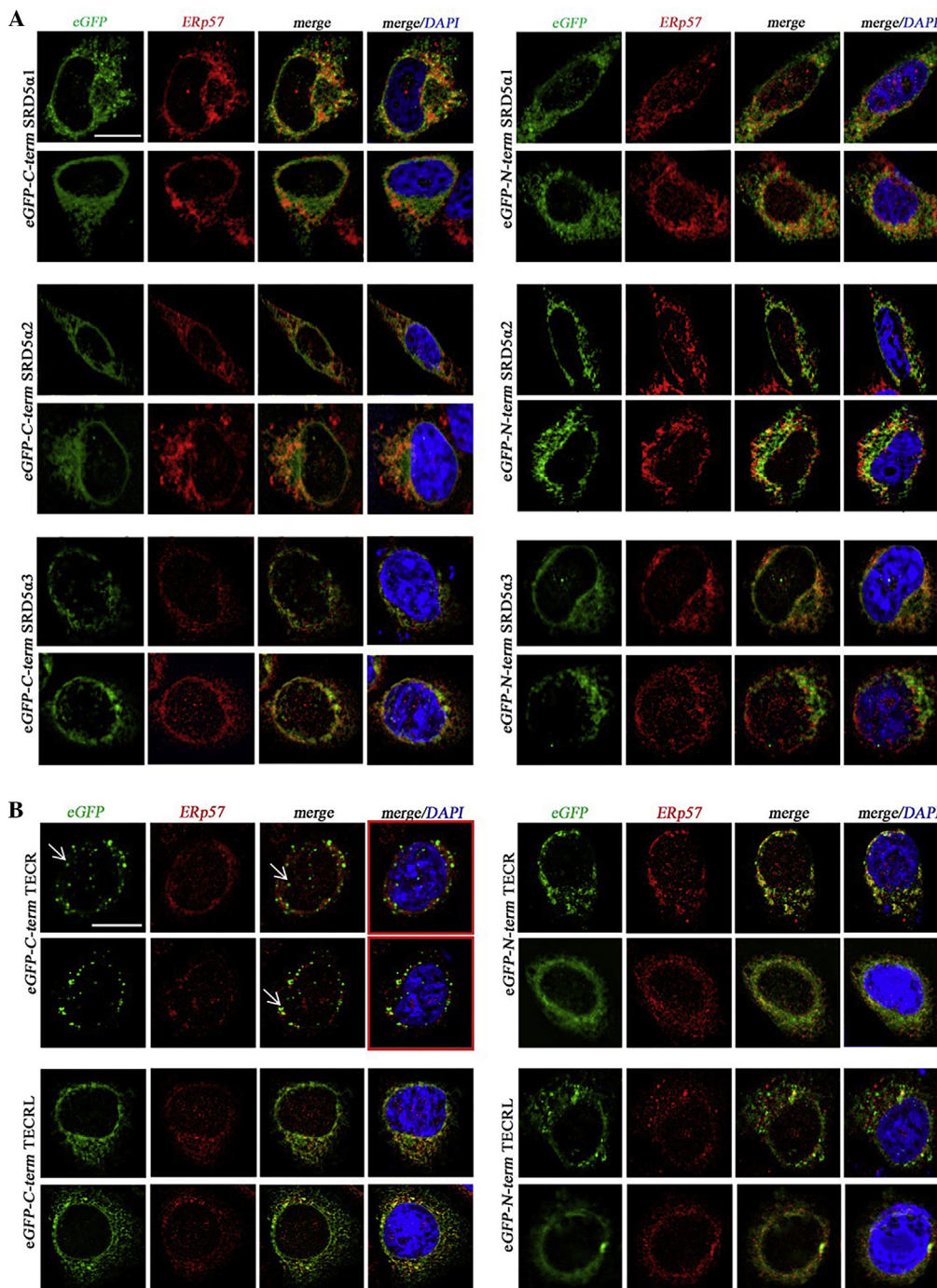
Fig. 2. **Fluorescence of human SRD5As (A) and TECR, TECRL (B) proteins in relation to endogenous ERp57.** SRD5A1, SRD5A2 and SRD5A3 proteins eGFP-tagged at C- or N-term (A), N-term eGFP-tagged TECR and N- and C-term eGFP-tagged TECRL proteins (B) reside in the ER, as indicated by co-localization with Erp57 and overall distribution of green fluorescence relative to Erp57. eGFP C-term TECR (B, left column) shows diffuse bright dots indicated by white arrows, with only limited overlap with ERp57. Deconvolutions plus Extended Depth of Focus (EDF) were performed on Z-stack series (13 frames-0, 6 mm each). The scanning of Z-stacks for red-framed images in B, with eGFP C-term TECR is given in supplementary Fig. 2. Scale bar = 10 μm, Magnification = 100×.

heterologous expression and purification in yeast, SRD5A1 appears to be associated to nuclei, whereas SRD5A2 is in the microsomal fraction [33]. Immunostaining for SRD5A2 showed presence of the enzyme in the cytoplasm throughout most regions of the brain [34]. Finally, it was reported that

glioma cells express SRD5A1 in the perinuclear space or in the cytoplasm, depending on culture conditions [35].

As far as the third isoform of testosterone α-reductase (SRD5A3, also known as polyprenol reductase), available data report recovery from the microsomal fraction [8] and transient

A

eGFP | Giantin | merge | merge/DAPI

eGFP-C-term TECR

eGFP-C-term TECRL

eGFP-N-term TECR

eGFP-N-term TECRL

B

eGFP C-term TECR. | Proteostat | merge
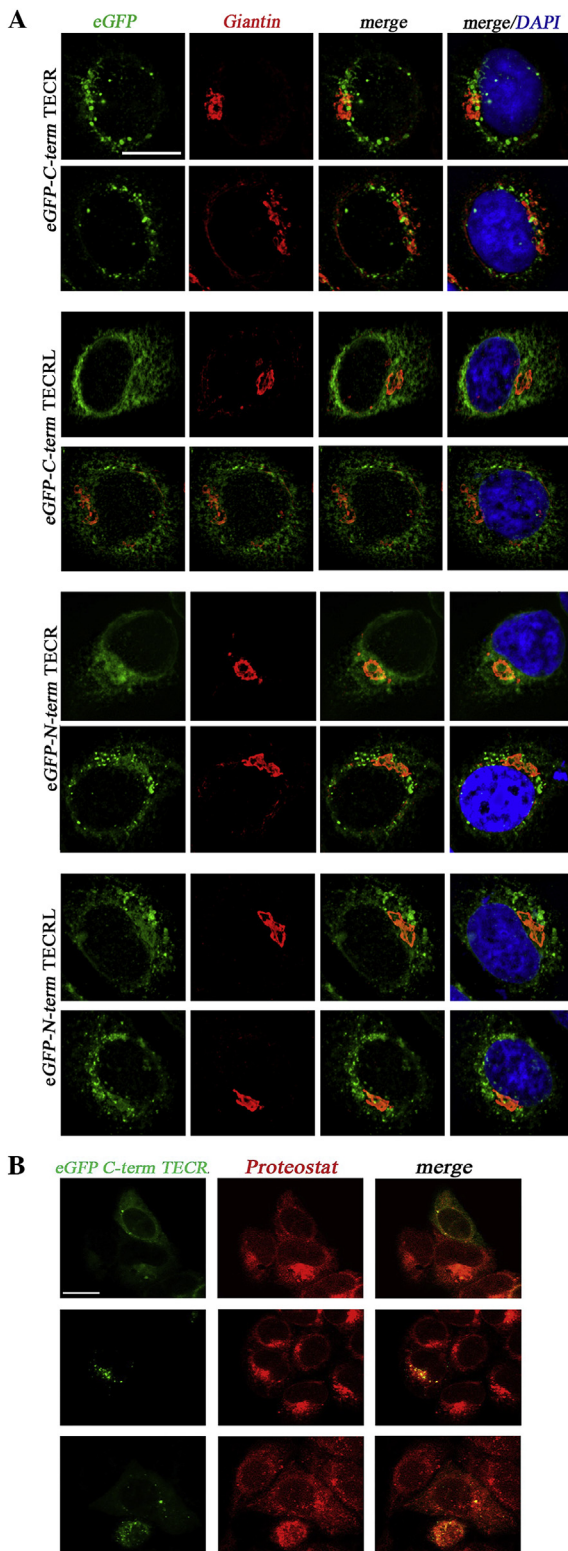
Fig. 3. **Fluorescence analysis of human TECR and TECRL relative to Giantin and fluorescence of eGFP C-term TECR and aggregates/aggresomes in the cytoplasm.** (A) TECR proteins do not localize in the Golgi apparatus, as associated eGFP fluorescence shows a broader distribution around, but not coinciding with, the Giantin signal. Scale bar = 10 μm, Magnification = 100×. (B) Localization of eGFP C-term TECR and aggresomes in HeLa cells after treatment with ProteoStat® dye for 30 min. Fluorescence of C-term eGFP-tagged TECR partially co-localizes with aggregated proteins. Scale bar = 10 μm, Magnification = 60×.

expression in HeLa cells yielded cytoplasmic localization, and no detectable activity [36]. For the TECR enzyme, immunohistochemistry has clearly suggested a localization in the ER [12]. No data is available for the TECRL that has been described only at the gene level.

SRD5A family members share significant sequence similarity, having likely evolved via duplication from an ancestral gene (sequence alignment is shown in Fig. S3). In localization studies, in order to eliminate any potential ambiguities due to antibody specificity, we directly examined each of the five human SRD5A family members by taking advantage of the fluorescent emission of the eGFP fusion tag in transfected HeLa cells. The caveat of aberrant subcellular localization due to the presence of the eGFP tag calls for careful consideration, and we therefore systematically compared the localization of C-term and N-term tagged paired constructs for each SRD5A member in parallel experiments.

The effect of a GFP tag on membrane protein localization and expression can vary, nevertheless this method is reported as a valid tool in subcellular localization studies [37–39]. The positioning on the C- or N-term of the polypeptide chain can affect protein expression and stability and its effect varies depending on the system under analysis. In an extended study on the effects on GFP tagging on subcellular localization [40] C-term tagging is reported to maintain physiological localization, whereas N-term affected it, contrarily to what we observed with human TECR.

The data we obtained with proteins fused with eGFP at the N-term indicate that all members of the SRD5A family reside in the endoplasmic reticulum, when transient expression is carried out in HeLa cells. This is also evident for protein constructs where the fusion with eGFP was at the C-term, with the exception of TECR. TECR was previously localized in the endoplasmic reticulum by immunofluorescence staining in transiently transfected HeLa cells [10]. This, and the patterns obtained here with the C-term chimaera, allowed us to conclude that the presence of eGFP at the C-term affects TECR folding, leading to the accumulation of aggregated protein.

In the SRD5A proteins, N-term eGFP-tagging reduces expression for SRD5A1, TECRL and TECR, and for the latter C-term eGFP-tagging promotes aggregation. On the other hand, for SRD5A3, expression is somewhat higher in the N-term eGFP-tagged form. Thus, even within a family of proteins, the effect of fusion tags on protein expression is not predictable and has to be assessed for each member. Our results indicate that membrane protein tagging with eGFP, which is commonly used to assess protein expression by fluorescence-coupled size-exclusion chromatography (FSEC) [41], can lead to very different expression yields, depending on the tag position, and that higher expression levels, as in the case of C-term tagged TECRL with respect to the N-term tagged form, can underlie unfolding and aberrant subcellular localization.

In summary, the present studies support the conclusion that isozymes 1 and 2 of testosterone 5α-reductase (SRD5A1 and SRD5A2) are both localized in the ER, upon transient expression in HeLa cells, our data are a valuable contribution to clarify the contradictory reports detailed above, although

any ultimate conclusion on the intracellular localization of SRD5As should be validated in other cell lines and complementary experimental settings.

Moreover, the subcellular localization in the ER for SRD5A3 is consistent with the reported microsomal subfractionation. Finally, we show that the previously uncharacterized TECRL gene can be expressed in human cells and that its product is a protein residing in the ER membrane.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biopen.2017.03.003.

References

[1] V.S. Langlois, D. Zhang, G.M. Cooke, V.L. Trudeau, Evolution of steroid-5alpha-reductases and comparison of their function with 5beta-reductase, Gen. Comp. Endocrinol. 166 (2010) 489−497, http://dx.doi.org/10.1016/j.ygcen.2009.08.004.

[2] R.I. Dorfman, E. Forchielli, Separation of delta 4-5 alpha-hydrogenases from rat liver homogenates, J. Biol. Chem. 223 (1956) 443−448, http://dx.doi.org/10.1155/2012/530121.

[3] T. Saartok, E. Dahlberg, J.A. Gustafsson, Relative binding affinity of anabolic-androgenic steroids: comparison of the binding to the androgen receptors in skeletal muscle and in prostate, as well as to sex hormone-binding globulin, Endocrinology 114 (1984) 2100−2106, http://dx.doi.org/10.1210/endo-114-6-2100.

[4] J. Imperato-McGinley, L. Guerrero, T. Gautier, R.E. Peterson, Steroid 5alpha-reductase deficiency in man: an inherited form of male pseudo-hermaphroditism, Science 186 (1974) 1213−1215. ISSN: 05476844.

[5] A. Cilotti, G. Danza, M. Serio, Clinical application of 5alpha-reductase inhibitors, J. Endocrinol. Investig. 24 (2001) 199−203.

[6] É. Lévesque, I. Laverdière, L. Lacombe, P. Caron, M. Rouleau, V. Turcotte, B. Têtu, Y. Fradet, C. Guillemette, Importance of 5α-reductase gene polymorphisms on circulating and intraprostatic androgens in prostate cancer, Clin. Cancer Res. 20 (2014) 576−584, http://dx.doi.org/10.1158/1078-0432.CCR-13-1100.

[7] B. Stoffel-Wagner, Neurosteroid biosynthesis in the human brain and its clinical implications, Ann. N. Y. Acad. Sci. 1007 (2003) 64−78, http://dx.doi.org/10.1196/annals.1286.007.

[8] F. Azzouni, A. Godoy, Y. Li, J. Mohler, The 5alpha-reductase isozyme family: a review of basic biology and their role in human diseases, Adv. Urol. 2012 (2012), http://dx.doi.org/10.1155/2012/530121. ID530121.

[9] M. Uemura, K. Tamura, S. Chung, et al., Novel 5alpha-steroid reductase (SRD5A3, type 3) is overexpressed in hormone-refractory prostate cancer, Cancer Sci. 99 (2008) 81−86, http://dx.doi.org/10.1111/j.1349-7006.2007.00656.x.

[10] V. Cantagrel, D.J. Lefeber, B.G. Ng, et al., SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder, Cell 142 (2010) 203−217, http://dx.doi.org/10.1016/j.cell.2010.06.001.

[11] K. Yamana, F. Labrie, V. Luu-The, Human type3 5alpha-reductase is expressed in peripheral tissues at higher levels than types1 and 2 and its activity is potently inhibited by finasteride and dutasteride, Horm. Mol. Biol. Clin. Investig. 2 (2010) 293−299, http://dx.doi.org/10.1515/HMBCI.2010.035.

[12] K. Abe, Y. Ohno, T. Sassa, et al., Mutation for nonsyndromic mental retardation in the trans-2-enoyl-CoA reductase TER gene involved in fatty acid elongation impairs enzyme activity and stability, leading to change in sphingolipid profile, J. Biol. Chem. 288 (2013) 36741−36749, http://dx.doi.org/10.1074/jbc.M113.493221.

[13] Y.A. Moon, J.D. Horton, Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade, J. Biol. Chem. 278 (2003) 7335−7343, http://dx.doi.org/10.1074/jbc.M211 84200.

[14] M. Çalışkan, J.X. Chong, L. Uricchio, R. Anderson, et al., Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13, Hum. Mol. Genet. 20 (2011) 1285−1289, http://dx.doi.org/10.1093/hmg/ddq569.

[15] D.W. Russell, J.D. Wilson, Steroid 5alpha-reductase: two genes/two enzymes, Annu. Rev. Biochem. 63 (1994) 25−61, http://dx.doi.org/10.1146/annurev.bi.63.070194.000325.

[16] K. Normington, D.W. Russell, Tissue distribution and kinetic characteristics of rat steroid 5alpha-reductase isozymes. Evidence for distinct physiological functions, J. Biol. Chem. 267 (1992) 19548−19554. ISSN: 00219258.

[17] R. Bruni, B. Kloss, High-throughput cloning and expression of integral membrane proteins in E.coli, Curr. Protoc. Protein Sci. 74 (2013), http://dx.doi.org/10.1002/0471140864.ps2906s74. Unit 29.6.

[18] N.I. Tarasova, R.H. Stauber, J.K. Choi, et al., Visualization of G protein-coupled receptor trafficking with the aid of the green fluorescent protein. Endocytosis and recycling of cholecystokinin receptor type A, J. Biol. Chem. 272 (1997) 14817−14824, http://dx.doi.org/10.1074/jbc.272.23.14817.

[19] M. Koreishi, T.J. Gniadek, S. Yu, et al., The golgin tether Giantin regulates the secretory pathway by controlling stack organization within Golgi apparatus, PLoS One 8 (2013) e59821, http://dx.doi.org/10.1371/journal.pone.0059821.

[20] A. Chaumet, G.D. Wright, S.H. Seet, et al., Nuclear envelope-associated endosomes deliver surface proteins to the nucleus, Nat. Commun. 6 (2015) 8218, http://dx.doi.org/10.1038/ncomms9218.

[21] Z.J. Guo, L.X. Tao, X.Y. Dong, et al., Characterization of aggregate/aggresome structures formed by polyhedrin of *Bombyx mori* nucleopolyhedrovirus, Sci. Rep. 5 (2015) 14601, http://dx.doi.org/10.1038/srep14601.

[22] R.I. Silver, E.L. Wiley, A.E. Thigpen, et al., Cell type specific expression of steroid 5alpha-reductase 2, J. Urol. 152 (1994) 438−442, http://dx.doi.org/10.1073/pnas.90.20.9359.

[23] A.E. Thigpen, K.M. Cala, D.W. Russell, Characterization of Chinese hamster ovary cell lines expressing human steroid 5alpha-reductase isozymes, J. Biol. Chem. 268 (1993) 17404−17412. ISSN: 00219258.

[24] R.W. Hudson, Studies of nuclear 5alpha-reductase of human hyperplastic prostatic tissue, J. Steroid Biochem. Mol. Biol. 14 (1981) 579−584.

[25] R.J. Moore, J.D. Wilson, Localization of the reduced nicotinamide adenine dinucleotide phosphate: 4-3-ketosteroid 5-oxidoreductase in the nuclear membrane of the rat ventral prostate, J. Biol. Chem. 247 (1972) 958−967. ISSN: 00219258.

[26] B. Houston, G.D. Chisholm, F.K. Habib, Evidence that human prostatic 5alpha-reductase is located exclusively in the nucleus, FEBS Lett. 185 (1985) 231—235, http://dx.doi.org/10.1016/0014-5793(85)80912-1.

[27] J.G. Savory, D. May, T. Reich, et al., 5-alpha-reductase type1 is localized to the outer nuclear membrane, Mol. Cell. Endocrinol. 110 (1995) 137—147.

[28] M.F. Taylor, M. Wang, A.K. Bhattacharyya, et al., Expression of rat steroid 5 alpha-reductase (isozyme-1) in *Spodoptera frugiperda*, SF21, insect cells: expression of rat steroid 5 alpha-reductase, Steroids 62 (1997) 373—378.

[29] S. Délos, C. Iehlé, P.M. Martin, J.P. Raynaud, Inhibition of the activity of 'basic' 5alpha-reductase (type 1) detected in DU 145 cells and expressed in insect cells, J. Steroid Biochem. Mol. Biol. 48 (1994) 347—352, http://dx.doi.org/10.1016/0960-0760(94)90074-4.

[30] C. Iehlé, S. Délos, O. Guirou, R. Tate, J.P. Raynaud, P.M. Martin, Human prostatic steroid 5alpha-reductase isoforms. A comparative study of selective inhibitors, J. Steroid Biochem. Mol. Biol. 54 (1995) 273—279, http://dx.doi.org/10.1016/0960-0760(95)00134-L.

[31] B.U. Panter, J. Jose, R.W. Hartmann, 5alpha-reductase in human embryonic kidney cell line HEK293: evidence for type II enzyme expression and activity, Mol. Cell Biochem. 270 (2005) 201—208, http://dx.doi.org/10.1007/s11010-005-4508-8.

[32] L.N. Thomas, R.C. Douglas, R.S. Rittmaster, C.K. Too, Overexpression of 5alpha-reductase type1 increases sensitivity of prostate cancer cells to low concentrations of testosterone, Prostate 69 (2009) 595—602, http://dx.doi.org/10.1002/pros.20911.

[33] A. Poletti, F. Celotti, M. Motta, L. Martini, Characterization of rat 5alpha-reductases type1 and type2 expressed in Saccharomyces cerevisiae, Biochem. J. 314 (1996) 1047—1052, http://dx.doi.org/10.1042/bj3141047.

[34] G. Pelletier, V. Luu-The, F. Labrie, Immunocytochemical localization of 5alpha-reductase in rat brain, Mol. Cell. Neurosci. 5 (1994) 394—399, http://dx.doi.org/10.1006/mcne.1994.1049.

[35] Y. Tsuruo, K. Ishimura, K. Morita, Influence of serum-free culture conditions on subcellular localization of steroid 5alpha-reductase in rat C6 glioma cells, Brain Res. 801 (1998) 130—136, http://dx.doi.org/10.1016/S0006-8993(98)00555-1.

[36] B. Chávez, L. Ramos, R. García-Becerra, F. Vilchis, Hamster SRD5A3 lacks steroid 5alpha-reductase activity in vitro, Steroids 94 (2015) 41—50, http://dx.doi.org/10.1016/j.steroids.2014.11.005.

[37] A. Limon, J.M. Reyes-Ruiz, F. Eusebi, R. Miledi, Properties of GluR3 receptors tagged with GFP at the amino or carboxyl terminus, Proc. Natl. Acad. Sci. U.S. A 104 (2007) (2007) 15526—15530, http://dx.doi.org/10.1073/pnas.0706773104.

[38] S.I. Hyun, L. Maruri-Avidal, B. Moss, Topology of endoplasmic reticulum-associated cellular and viral proteins determined with split-GFP, Traffic 16 (2015) 787—795, http://dx.doi.org/10.1111/tra.12281.

[39] L. Kallal, J.L. Benovic, Using green fluorescent proteins to study G-protein-coupled receptor localization and trafficking, Trends Pharmacol. Sci. 21 (2000) 175—180, http://dx.doi.org/10.1016/S0165-6147(00)01477-2.

[40] E. Palmer, T. Freeman, Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays, Comp. Funct. Genomics 5 (2004) 342—353, http://dx.doi.org/10.1002/cfg.405.

[41] T. Kawate, W. Gouax, Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins, Structure 14 (2006) 673—681, http://dx.doi.org/10.1016/j.str.2006.01.013.

Substrate-induced conformational change in cytochrome P450 OleP

Giacomo Parisi,*,†,1,2 Linda Celeste Montemiglio,*,†,‡,1 Alessandro Giuffrè,‡ Alberto Macone,†
Antonella Scaglione,*,†,3 Gabriele Cerutti,† Cécile Exertier,† Carmelinda Savino,‡,4 and Beatrice Vallone*,†,‡,5

*Istituto Pasteur-Fondazione Cenci Bolognetti and †Dipartimento di Scienze Biochimiche A. Rossi Fanelli, Sapienza Università di Roma, Rome, Italy; and ‡Institute of Molecular Biology and Pathology, National Research Council, Rome, Italy

ABSTRACT: The regulation of cytochrome P450 activity is often achieved by structural transitions induced by substrate binding. We describe the conformational transition experienced upon binding by the P450 OleP, an epoxygenase involved in oleandomycin biosynthesis. OleP bound to the substrate analog 6DEB crystallized in 2 forms: one with an ensemble of open and closed conformations in the asymmetric unit and another with only the closed conformation. Characterization of OleP-6DEB binding kinetics, also using the P450 inhibitor clotrimazole, unveiled a complex binding mechanism that involves slow conformational rearrangement with the accumulation of a spectroscopically detectable intermediate where 6DEB is bound to open OleP. Data reported herein provide structural snapshots of key precatalytic steps in the OleP reaction and explain how structural rearrangements induced by substrate binding regulate activity.—Parisi, G., Montemiglio, L. C., Giuffrè, A., Macone, A., Scaglione, A., Cerutti, G., Exertier, C., Savino, C., Vallone, B. Substrate-induced conformational change in cytochrome P450 OleP. FASEB J. 33, 000–000 (2019). www.fasebj.org

KEY WORDS: X-ray crystallography · multistep binding kinetics · 6DEB · structural transition · spectroscopic intermediate

The cytochrome P450 (P450) superfamily consists of a large number of heme-containing monooxygenases, distributed in virtually all organisms, catalyzing the oxidation of organic compounds by insertion of an oxygen atom derived from molecular oxygen, into nonactivated C—H or C—C bonds (1). They perform multiple functions on a multitude of structurally unrelated molecules. Many P450s catalyze the biosynthesis of endogenous compounds, such as steroids, vitamins, fatty acids, retinoids, and drugs, in reactions that require regio- and stereospecificity. Other

P450s are involved in the metabolism of xenobiotics and drugs, showing pronounced substrate promiscuity (2). Although P450s share common folding and catalytic mechanisms, the flexibility and the modularity of specific secondary structure elements surrounding the active site and differing in primary structure confer functional versatility to this superfamily of enzymes. These regions are responsible for substrate recognition and orientation in the active site, along with structural transitions (3).

Structural and functional data, molecular dynamics simulation, and multidimensional NMR have demonstrated that binding to P450s is often coupled to conformational changes of different extents that allow the enzyme to reach the optimal conformation to perform catalysis (4–11). Nevertheless, the extended versatility of reactions catalyzed by P450s, coupled to the variety of their substrates, makes the mechanism of binding difficult to generalize and only decipherable through a case-by-case analysis. However, for bacterial P450s involved in biosynthesis, a general picture is emerging on a role of the central portion of the I helix in sensing the presence of substrates and triggering major rearrangements (12).

The P450 8.8a-oleandolide epoxygenase (OleP) is a bacterial epoxidase produced by *Streptomyces antibioticus* and involved in the final steps of the biosynthesis of the 14-member polyketide antibiotic oleandomycin (13, 14).

During the first stage of the biosynthesis, a specific polyketide synthase, responsible for assembling an acetate

unit to 6 molecules of methylmalonyl coenzyme A (CoA), releases an aglycone intermediate, 8,8a-deoxyoleandolide (DEO). This compound then undergoes the following tailoring reactions: *1*) attachment of L-oleandrose and D-desosamine at C3 and C5, respectively; *2*) 3′-OH methylation of L-oleandrose by a methyl transferase, and *3*) regio- and stereospecific epoxidation at C8—C8a of the macrolactone ring catalyzed by OleP (**Fig. 1**) (15).

The precise timing of the OleP reaction has not yet been fully clarified, but some experimental evidence has shown that OleP has the potential to introduce the epoxide function both to the aglycone and the C3-monoglycosylated intermediates, suggesting the existence of parallel pathways in oleandomycin biosynthesis (13, 14, 16, 17). As far as the OleP substrates are concerned, both compounds are not commercially available and are hard to synthetize because of the presence of multiple chiral sites.

Recently, the structure of OleP in complex with clotrimazole (CLT) showed that the inhibitor bound to the heme iron locks the enzyme in an open conformation. The pre–steady-state kinetic characterization of CLT binding revealed that OleP follows a simple 1-step mechanism, without accumulation of detectable intermediates (18).

In this work, we present the structural and functional characterization of the binding of the substrate analog 6-deoxyerythronolide B (6DEB) to OleP. 6DEB is the aglycone intermediate of erythromycin biosynthesis. It differs marginally from DEO, the natural aglycone substrate of OleP, because only the methyl group at the C13 is replaced by an ethyl moiety (Fig. 1) and the activity of OleP on this compound has been reported *in vivo* (16). Structural analysis has revealed that OleP bound to 6DEB adopts both a closed and an open conformation in the asymmetric unit, whereas higher ionic strength crystallization conditions induce the complete shift of the OleP-6DEB complex to the closed conformation. The quasi-pseudo first-order kinetic characterization of OleP binding using 6DEB and CLT, in conjunction with structural data, allows us to propose a mechanism of complex formation. 6DEB binding to OleP proceeds through the accumulation of a spectroscopically detectable substrate-bound open intermediate, followed by a dramatic conformational transition upon repositioning of the substrate over the heme.

MATERIALS AND METHODS

Chemicals

6DEB was kindly provided by Barrie Wilkinson and Rachel Lill (Biotica Technology, Cambridge, United Kingdom). DMSO, spinach ferredoxin, spinach ferredoxin NADP⁺ reductase, NADPH, NADP⁺, glucose-6-phosphate, and glucose-6-phosphate
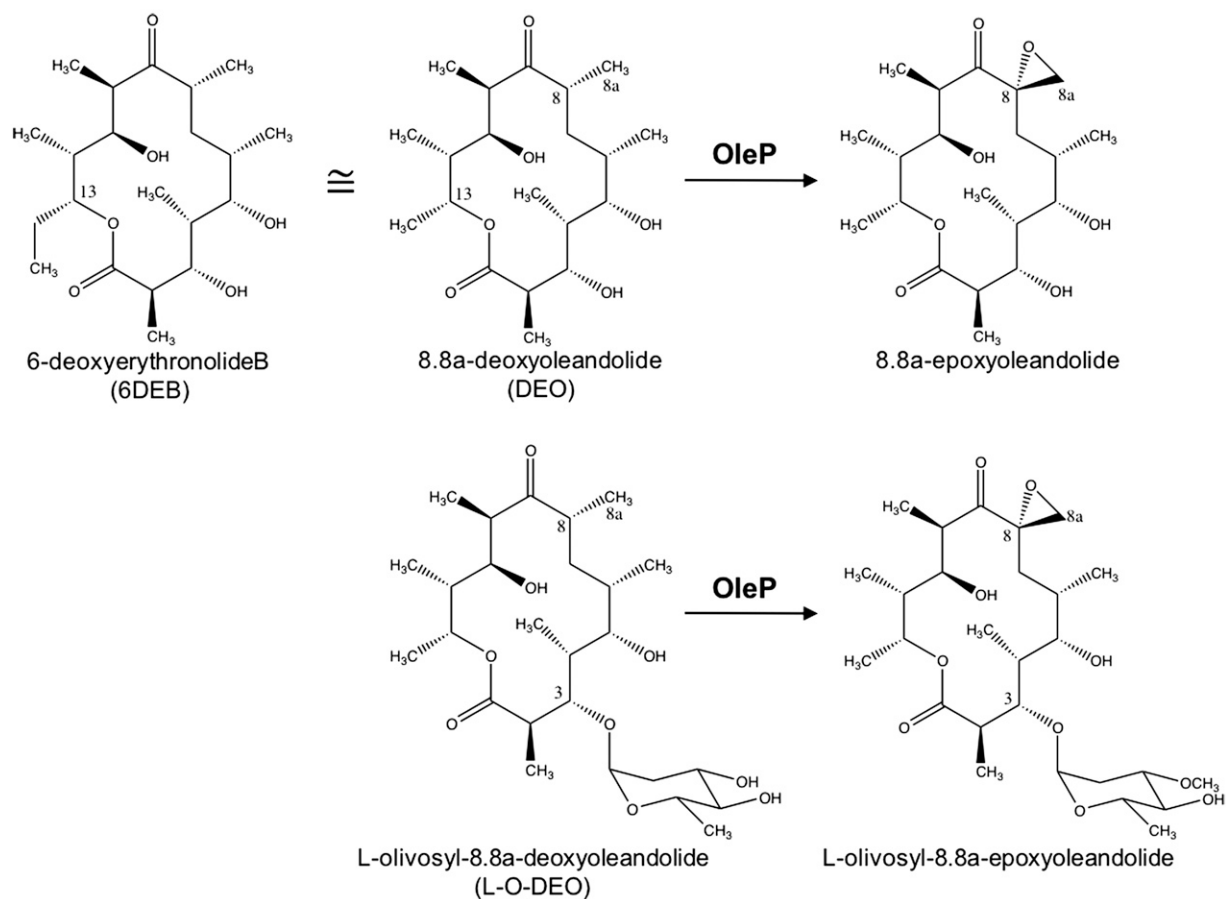

Figure 1. Scheme of the reactions catalyzed by OleP in oleandomycin biosynthesis. OleP introduces an epoxide function on both the macrolide DEO and the monoglycosylated intermediate L-olivosyl-8.8a-deoxyoleandolide. The substrate analog 6DEB (top left) differs from DEO only in the substituent group at C13, where it shows an ethyl moiety in the place of a methyl moiety.

dehydrogenase were purchased by MilliporeSigma (Burlington, MA, USA).

Protein expression, purification, and crystallization and X-ray data collection and analysis

Recombinant His-tagged OleP was overexpressed in the *Escherichia coli* BL21 (DE3) strain and purified as reported by our group (18). To check sample homogeneity and degree of monodispersion, protein solution was loaded into a BioFox 17/1200 SEC (Knauer, Berlin, Germany) gel permeation column and eluted at 293 K in isocratic mode in a buffer containing 20 mM Tris-HCl (pH 8.0) and 0.2 M NaCl. In this condition, OleP was found to be in a monomeric state (18). Cocrystallization of purified OleP in complex with the substrate analog 6DEB was performed by means of the hanging-drop vapor-diffusion method with purified OleP at 19 mg/ml and saturating concentrations of ligand (1 mM) at 293 K. Crystals from 2 different crystallization conditions gave high-quality X-ray diffraction data and were obtained as: *1*) 0.4 M NaCl, 10 mM Tris-HCl (pH 7), and 22% polyethylene glycol (PEG) 3350 (low salt, OleP-6DEB$_{LS}$) and *2*) 4 M sodium formate (high salt, OleP-6DEB$_{HS}$). Ten percent glycerol was used as a cryoprotectant for the latter crystals before freezing in liquid nitrogen. Diffraction data were collected at 100 K at the European Synchrotron Radiation Facility (ID14-01 beamline; Grenoble, France) and Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY; Helmholtz-Zentrum Berlin, Berlin, Germany), beamline ID-14-02, with a Pilatus detector (Dectris, Baden-Dättwil, Switzerland). Three data sets of OleP-6DEB$_{LS}$ from crystals grown in the same drop were merged, to increase the signal-to-noise ratio, and were subsequently indexed, scaled, and integrated by using the XDS package (19). HS condition crystals displayed diffraction patterns typical of multiple lattices. Spots belonging to different lattices were discriminated by using the XDS protocol for indexing of images from nonmerohedrally twinned crystals (Department of Biology University of Konstanz, Konstanz, Germany; *http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/indexing/*) and indexed. Data were finally scaled and integrated using the XDS package.

Crystallization conditions, data collection, and refinement statistics are summarized in **Table 1**.

Structure determination and refinement

For both structures, molecular replacement was performed to determine the initial crystallographic phases using MolRep (20) from the CCP4 suite (21). For the OleP-6DEB$_{LS}$ complex, coordinates of a single OleP monomer in complex with the inhibitor CLT (18) have been used as a search model [Protein Data Bank (PDB) Code 4XE3; Research Collaboratory for Structural Bioinformatics (RCSB), Rutgers University, New Brunswick, NJ USA; *http://www.rcsb.org*]. The coordinates of a single closed monomer of OleP from the latter structure have been used as a search model in molecular replacement of the OleP-6DEB$_{HS}$ complex. Nine monomers (A–I) and 6 monomers (A–F) were found in 1 asymmetric unit of OleP-6DEB$_{LS}$ and OleP-6DEB$_{HS}$, respectively. Iterative manual and automated refinement of the structures was performed with Refmac5 (22) in the CCP4 suite (21). Five percent of the reflections were excluded from refinement and used for the R_{free} calculation (23). Manual adjustment was performed using Coot (24) with an $F_{observed}$-$F_{calculated}$ (F_o-F_c) map contoured at 3σ and the 2Fo-Fc at 1σ. Water molecules were added to the F_o-F_c map, contoured at 3σ, using the Find Water tool of Coot. The geometric quality of the final models was assessed with Procheck (25). The first 11–13 N-terminal residues, depending on the monomer, are missing from the models because of insufficient electron density in this region. The quality of the electron density in both OleP-6DEB structures becomes weaker spanning from the first to the last monomer. This trend is particularly evident in the monomer named I in the OleP-6DEB$_{LS}$ structure that displays electron density accounting for only 70% of the protein (lack of density for all the β-sheet motif, B' helix and BC, FG, JK and KL loops). Nevertheless, the

TABLE 1. *Data collection, refinements, statistics, and validation*

Data collection	OleP-6DEB$_{LS}$, PDB 5MNV	OleP-6DEB$_{HS,}$ PDB 5MNS
Crystallization conditions	0.4 M NaCl, 10 mM Tris-HCl (pH 7), 22% PEG 3350	4 M sodium formate
Space group	P1	C2
Unit cell (Å, °)	a = 112.50, b = 116.70, c = 123.30, α = 103.90, β = 104.30, γ = 114.70	a = 247.60, b = 110.80, c = 159.10, β = 129.6
Resolution (Å)	121–2.97 (3.17–2.97)	120–2.62 (2.78–2.62)
Total measurements	356,364	599,678
Unique reflections	101,917	98,805
Completeness (%)	97.5 (91.8)	99.5 (97)
Redundancy	3.49 (3.54)	6.09 (6.8)
R_{merge}^{a} (%)	9.1 (48.6)	10.6 (83.3)
$I/σ$ (I)	11.92 (2.2)	13.63 (2.03)
Wilson B-value (Å^2)	54.4	42.5
Refinement		
Molecules per asymmetric unit	9	6
R_{work}/R_{free}^{b} (%)	23.8/31.1	20.8/25.8
RMSD		
Bond (Å)	0.007	0.009
Angles (°)	1.21	1.39
Ramachandran [(%) favored/allowed/outliers^c]	93.3/6.7/0.0	96.0/4.0/0.0

The range of the highest resolution shell is shown in parentheses, as well as the values for the corresponding refinement parameters. ^{a}R_{merge}=$\sum_i\sum_j|I_{i,j} - \langle I_j\rangle|/\sum_i\sum_j I_{i,j}$, where i runs over multiple observations of the same intensity, and j runs over all crystallographically unique intensities. ^{b}R_{work} = $\sum||F_{obs}|| - |F_{calc}|/\sum|F_{obs}|$, where $|F_{obs}| > 0$. R_{free} is based on 5% of the data randomly selected and is not used in the refinement. ^{c}No prolines or glycines.

differences between copies of OleP-6DEB adopting the same conformations are not significant [monomers in open conformation root mean square difference (RMSD) on C_α ranging from the minimum value of 0.475 for monomer D to the maximum value 0.681 of the monomer E, using monomer A as a reference in the OleP-6DEB$_{LS}$ structure; monomers in closed conformation RMSD ranging from the minimum value of 0.274 for monomer A in the OleP-6DEB$_{HS}$ structure to the maximum value 0.883 of the monomer I in the OleP-6DEB$_{LS}$ structure, using monomer C of OleP-6DEB$_{HS}$ as a reference] and are likely to reflect the quality of the individual electron density maps. Therefore, the better-defined molecule A from OleP-6DEB$_{LS}$ and monomer C from OleP-6DEB$_{HS}$ as representative of distinct conformations of the OleP-6DEB complex (open and closed) were used for the structural and comparative analyses discussed in the article.

Final crystallographic statistics for both structures are shown in Table 1; data extend to 2.9 Å resolution for OleP-6DEB$_{LS}$ and 2.6 Å for OleP-6DEB$_{HS}$. No outliers in the Ramachandran plot were identified in either structure. All figures were produced using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, New York, NY, USA; *http://pymol.sourceforge. net/*). The atomic coordinates and structure factors for the OleP-6DEB$_{LS}$ and OleP-6DEB$_{HS}$ structures have been deposited in the RCSB Protein Data Bank (Accession Codes 5MNV and 5MNS, respectively).

Equilibrium binding analysis

OleP binding affinity to 6DEB was determined at 288 K by titrating the enzyme (1.5 µM) with increasing concentration of the substrate analog (from 0 to 100 µM), in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 200 mM NaCl (pH 7.5). Because OleP was purified in the oxidized form and catalysis would require external electron donors, addition of 6DEB resulted in monitoring only the binding reaction. The stock solution of 6DEB was prepared at a concentration of 26 mM in DMSO. The final percentage of DMSO was maintained below 1%. The adsorption shift of the heme Soret peak of OleP from 417 to 390 nm was followed by collecting UV-visible spectra (200–800 nm) after each substrate addition, subtracting the appropriate blank (26). The dissociation constant (K_D) was estimated with Kaleidagraph software (Synergy Software, Reading, PA, USA). A nonlinear regression analysis was applied to the experimental data to estimate the binding affinity constant, K_D, of OleP to 6DEB using the hyperbolic equation $\Delta AU_{obs} = \Delta AU_{max}[L]/(K_D + [L])$, where ΔAU_{obs} is the absorbance difference, ΔAU_{max} is the maximum absorbance difference extrapolated to infinite ligand concentration, and [L] is the ligand analytical concentration. All data were also globally fitted with Prism (GraphPad Software, La Jolla, CA, USA).

Spectrophotometric assay for enzymatic activity and gas chromatography–mass spectrometry analyses of the epoxidation products

The epoxidase activity of OleP was tested with respect to the substrate-like compound 6DEB. The enzymatic assay to test the activity of OleP in response to 6DEB was performed with the system ferredoxin/ferredoxin NADP+ reductase as electron donor that uses the cofactor NADPH ($\varepsilon_{340\,nm} = 6220$ M/cm) as the source of electrons (27). In a total volume of 50 µl, 5 µM enzyme, 3.6 µM spinach ferredoxin, 1 µM spinach ferredoxin NADP$^+$ reductase, 50 mM Tris-HCl (pH 7.5), and substrate (from 0 to 600 µM) were mixed together in a 1-cm light path cuvette mounted on a constant-temperature cell holder equilibrated at 303 K. NADPH at a final concentration of 100 µM was added to start the reaction. Initial velocities were measured spectroscopically at different substrate concentrations by monitoring the rate of absorbance decay at 340 nm over time with a J650 spectrophotometer (Jasco, Easton, MD, USA) in kinetic mode, setting 600 s as run time. Initial rates were estimated by applying a linear fit to the linear region of the time trace, determined to be the first 30 s. Initial velocity data were fitted using the Kaleidagraph software package by following the hyperbolic equation of Michaelis–Menten:

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

where v is the initial velocity of the reaction, V_{max} is the maximum velocity of the reaction corresponding to substrate saturation of the enzyme, K_M is the Michaelis–Menten constant, and [S] is the substrate analytical concentration.

To analyze the product of the OleP reaction using 6DEB as a substrate, we incubated 1 µM of the enzyme with 100 µg of 6DEB in the presence of 20 µM spinach ferredoxin, 0.2 U of spinach NADP$^+$ reductase, 1 mM NADP$^+$, 7.5 mM glucose-6-phosphate, and 0.8 U glucose-6-phosphate dehydrogenase in a final volume of 1.86 ml of 50 mM Tris-HCl (pH 7.5) (27). Three aliquots were taken at 0, 60, and 180 min. Moreover, a negative control was obtained preparing the same mix without adding NADP$^+$. Six hundred and twenty microliter aliquots of each sample were dried under vacuum in a SpeedVac SC110 Savant (Thermo Fisher Scientific). Pellet was extracted 2 times with 100 µl MeOH and 200 µl EtOAc. The supernatants were dried under a stream of nitrogen and were resuspended in 30 µl CH$_2$Cl$_2$. Then, 15 µl trimethylsilane as a derivatization reagent were added. The reaction was heated for 45 min at 353 K, the solution was dried under a stream of nitrogen, and the residue was resuspended in 250 µl CH$_2$Cl$_2$, with 1 µl directly injected in gas chromatography–mass spectrometry (GC-MS). GC-MS analyses were performed on a 6850A gas chromatograph (Agilent Technologies, Santa Clara, CA, USA) coupled to a 5973N quadrupole mass selective detector (Agilent Technologies). Chromatographic separations were carried out on an HP-5ms fused silica capillary column (30 m × 0.25 mm inner diameter; Agilent Technologies) coated with 5% phenyl–95% dimethylpolysiloxane (film thickness, 0.25 µm) as the stationary phase. The injection mode was splitless at a temperature of 553 K. The column temperature program was 373 K (1 min), then changed to 573 K at a rate of 298 K/min and held for 5 min. The carrier gas was helium at a constant flow of 1.0 ml/min. The ionizing energy was 70 eV (ion source 553 K; ion source vacuum $10-5$ Torr). Mass spectra were collected in the full-scan mode within the m/z scan range of 50–700 atomic mass units.

The efficiency of OleP in NADPH consumption *vs.* 6DEB consumed was also assessed. In a total volume of 250 µl, 5 µM OleP, 3.6 µM spinach ferredoxin, 2.5 µM spinach NADP$^+$ reductase, 50 mM Tris-HCl, and 0.2 M NaCl (pH 7.5), and 100 µg 6DEB were mixed together in a 1 cm light-path cuvette mounted in a constant-temperature cell holder equilibrated at 303 K. After the addition of 100 µM NADPH, the reactions were run at 303 K. NADPH consumption was spectroscopically followed by monitoring the rate of absorbance decay at 340 nm over time with a J650 spectrophotometer (Jasco) in kinetic mode. The reactions were halted by adding 33 µl methanol after 300 s, when the initial NADPH was completely consumed. Then, each sample was dried under a stream of nitrogen, and pellets were treated as described above. Then, 15 µl trimethylsilane as a derivatization reagent was added. The reaction was heated for 45 min at 353 K, the solution was dried under a stream of nitrogen, the residue was resuspended in 250 µl CH$_2$Cl$_2$, and 1 µl was directly injected into the GC-MS column. A reliable estimation of the amount of substrate consumed was obtained by comparing the intensity of the peaks corresponding to 6DEB at the beginning and end of the reaction.

Stopped flow measurements

The time evolution of 6DEB binding and displacement experiments were followed by monitoring the change in absorbance on a stopped-flow apparatus. Measurements were performed on an SX18-MV stopped-flow instrument (Applied Photophysics, Leatherhead, United Kingdom) using symmetric 1:1 mixing mode. The instrument has a light path of 1 cm and can also work in a sequential mixing mode, allowing 2 sequential mixing events with a preset delay time in-between, and is equipped with either a photomultiplier or a diode-array for single- and multiwavelength acquisition, respectively.

Ligand-binding kinetics

The time evolution of 6DEB binding was followed by monitoring the change in absorbance at 390 nm on a stopped-flow apparatus. OleP, at a final constant concentration of 2.5 μM, was mixed with 6DEB diluted to the desired final concentration, varying in a range from 0 to 75 μM. The traces were recorded in 50 mM Tris-HCl (pH 7.5) and NaCl 0.2 M at 288 K. Control experiments indicated that the presence of 1% DMSO had no significant effects on kinetics. The kinetic traces, obtained by averaging at least 3 experiments, were analyzed by applying a double exponential fit with the software provided by both Applied Photophysics and Kaleidagraph.

The kinetics of binding of OleP to 6DEB at 288 K was also monitored with the stopped-flow apparatus in in the multiwavelength mode. OleP (5 μM) was symmetrically mixed with 10 μM 6DEB in 50 mM Tris-HCl, NaCl 0.2 M, and 1% DMSO. In this experiment, the instrument was set to acquire, over 20 s, 500 absorption spectra in logarithmic time-scale fashion, with a 1-ms acquisition time for each spectrum. To avoid light-induced artifacts, a filter cutting UV light at $\lambda < 360$ nm was used. Spectral deconvolution was performed by global analysis. When used, global fit analysis of spectra sets was performed with Matlab software (Mathworks, Northampton, MA, USA) by combining curve fit and singular value decomposition analysis, according to Henry and Hofrichter (28).

Displacement and double jump measurements

Displacement measurements were performed by pre-incubating a solution containing 5 μM OleP with 100 μM 6DEB in 50 mM Tris-HCl, 0.2 M NaCl (pH 7.5), and 1% DMSO. The complex was rapidly mixed with a solution of the same buffer containing different concentrations of CLT (final concentration, 6–25 μM) at 288 K. Upon CLT binding, the Soret peak shifts from 392 to 425 nm, as related to the displacement of the type I ligand 6DEB by the type II ligand CLT (Supplemental Fig. S1), showing a maximum differential absorbance at 433 nm, where we chose to monitor the displacement event. Kinetic traces collected at 433 nm showed a double transition. The rates for the 2 phases were estimated applying a double exponential fit to each kinetic trace, obtained by averaging at least 3 experiments, with Kaleidagraph software.

Double-mixing measurements in 50 mM Tris-HCl and 0.2 M NaCl (pH 7.5) were also performed by using an SX18-MV (Applied Photophysics) stopped-flow instrument at 288 K. OleP (10 μM) was premixed 1:1 with 200 μM 6DEB for different delay times (0.08–4 s) and then symmetrically mixed with CLT at a final concentration of 25 μM in 50 mM Tris-HCl, 0.2 M NaCl, and 1% DMSO (pH 7.5). The traces, obtained by averaging at least 3 experiments, were analyzed with Kaleidagraph software. The observed amplitudes were plotted as a function of delay times and fitted to a single exponential function.

Reduction kinetics

The electron transfer rate from reduced ferredoxin (from *Spinacia oleracea*) to ferric OleP bound to 6DEB was measured by following the time evolution of OleP reduction in anaerobic conditions, monitoring the change in absorbance at 450 nm upon CO binding to the ferrous OleP on a stopped-flow apparatus (SX18-MV; Applied Photophysics). A solution containing 7.2 μM spinach ferredoxin, 2 μM spinach ferredoxin NADP+ reductase, 200 μM NADPH in 50 mM Tris-HCl, and 0.2 M NaCl (pH 7.5) was symmetrically mixed with a solution containing 10 μM OleP and 200 μM 6DEB in 50 mM Tris-HCl and 0.2 M NaCl (pH 7.5) at 303 K. All solutions were previously degassed and saturated with CO (final concentration ~1 mM). Solutions were loaded into gastight syringes, and absorbance change was monitored at 450 nm to follow formation of the ferrous-CO complex. The kinetic traces, obtained by averaging at least 3 experiments, were analyzed by applying a single exponential fit using the software provided by Applied Photophysics and Kaleidagraph.

RESULTS

6DEB behaves as a substrate of the P450 OleP

The OleP activity against 6DEB was assessed *in vivo* by Gaisser *et al.* in 2002 (16). In the current study, we performed an *in vitro* characterization of the affinity and activity with this substrate analog, using the recombinant and isolated enzyme.

Upon 6DEB binding, the type I spectral transition typical of P450 substrates was observed, resulting in a shift of the Soret peak from 417 to 390 nm (**Fig. 2A**) (26). Equilibrium-binding experiments were performed at 288 K by measuring the absorbance spectra of OleP, at varied 6DEB concentrations. The titration profile observed at 417 and 390 nm was satisfactorily fitted to a hyperbolic function as expected from a simple 2-state equilibrium binding (Fig. 2A, inset). The K_D, as estimated from global analysis of 50 different wavelengths, was 5.1 ± 0.2 μM. The extent of the spin state shift is small; this is not unusual for bacterial macrolide acting P450s (27, 29–31). However, we cannot exclude a more pronounced effect on binding of the physiologic substrates.

To measure the steady-state constants for OleP activity on 6DEB, we performed an NADPH oxidation assay, using the ferredoxin/ferredoxin NADP$^+$ reductase system as an electron delivery shuttle. Initial rates, as estimated by following over time the absorbance decay at 340 nm related to NADPH consumption, were plotted against 6DEB concentration (Fig. 2B) and analyzed with the Michaelis-Menten equation, yielding $K_M = 8.7 \pm 1.6$ μM, $V_{max} = 0.077 \pm 0.004$ μM/s, corresponding to a catalytic constant of $k_{cat} = 0.015 \pm 0.001$/s (Fig. 2B). The productive usage of NADPH was assessed by measuring substrate consumption and product formation through a GC-MS analysis of the reaction mixture (Supplemental Fig. S2A, B).

We assessed the electron entry rate from ferredoxin to OleP-6DEB (Supplemental Fig. S2C), yielding a value of $(12.3 \pm 0.1) \cdot 10^{-3}$/s, very close to the rate of consumption of the substrate. One might expect a velocity of electron entry faster than the apparent rate of reaction, given that 2 electrons are necessary for catalysis and we have
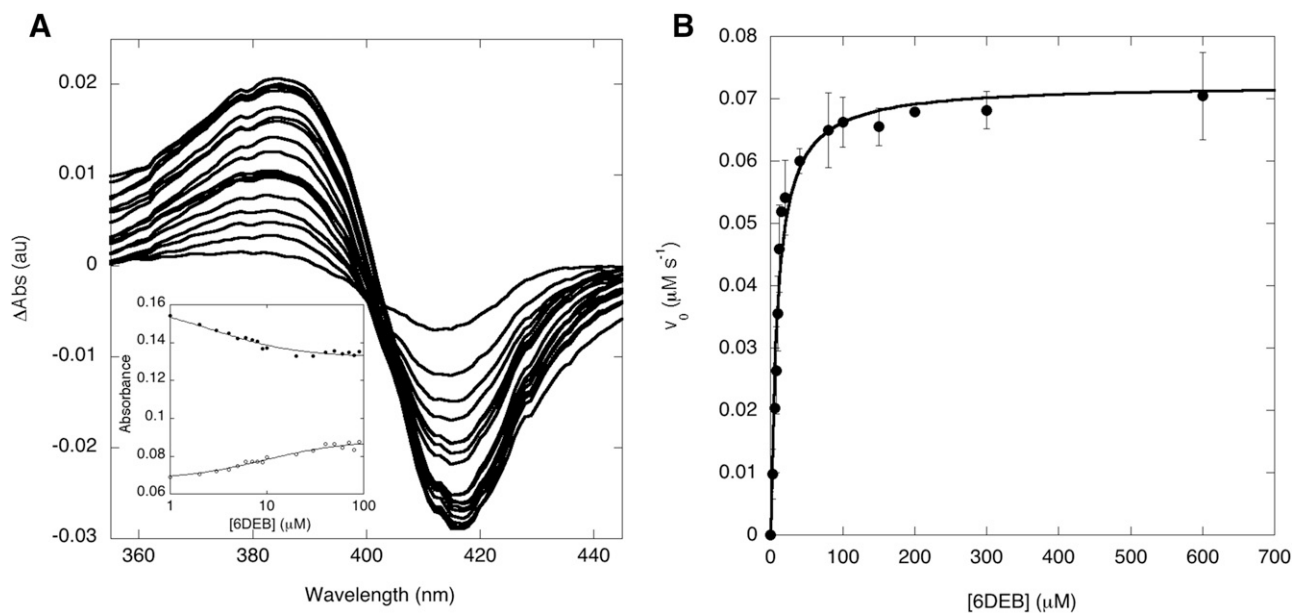
Figure 2. OleP binding affinity and epoxygenase activity on 6DEB. *A*) Equilibrium binding analysis. Differences in spectral changes observed upon 6DEB binding are reported. Substrate binding to OleP induces the typical type I spectral shift of the Soret band, with a peak at 390 nm and a trough at 417 nm. Inset: the absorbance changes at 417 (full dots) and 390 nm (open dots) as a function of substrate concentration on a logarithmic scale. Data were fitted to a hyperbolic equation (fits are shown as solid lines) to determine the K_D = 5.1 ± 0.2 µM at 288 K. *B*) Michaelis-Menten plot of 6DEB epoxygenation catalyzed by OleP. The initial rates were determined by applying a linear fit to the linear region of the time trace (first 30 s). In each experiment, absorbance was monitored at 340 nm by following the NADPH consumption over time at 303 K in 50 mM Tris-HCl and 0.2 M NaCl (pH 7.5). Error bars represent the SD of initial rates in 3 independent experiments.

measured the entrance of the first electron, since we followed CO binding to OleP(Fe^{2+}). Considering that the 2 electron transfer reactions between reduced ferredoxin and ferric OleP and ferrous and oxygen bound OleP may differ in affinity, binding site, and reaction mechanism, we hypothesize that the entrance of the second electron is considerably faster than the one of first electron, as reported for putidaredoxin/P450 monooxygenase from *Pseudomonas putida* system (32). This suggests that the first electron transfer reaction is likely to limit the catalytic turnover.

Moreover, the analysis of the total efficiency of NADPH oxidation and 6DEB consumption resulted in a 1:1 stoichiometry. Therefore, the data exclude uncoupling and suggest that the k_{cat} is affected by suboptimal electron transfer from spinach ferredoxin to OleP.

Nevertheless, these data confirm that OleP binds and is active against 6DEB, as suggested by *in vivo* studies (16).

OleP-6DEB structure

Crystals of the OleP-6DEB complex were obtained in 2 different conditions: *1*) 0.4 M NaCl and 22% PEG3350 (low salt, OleP-6DEB$_{LS}$) and *2*) 4 M HCOONa (high salt, OleP-6DEB$_{HS}$).

OleP-6DEB$_{LS}$ crystallized in P1 space group and the structure was refined to 2.97 Å resolution (R_{work} = 23.8% and R_{free} = 31.1%). Nine monomers were identified in the asymmetric unit adopting 2 different conformations (**Fig. 3***A*). In more detail, 6 copies of OleP were found in an open state corresponding to the one previously observed when

CLT (Supplemental Fig. S3*A*) occupies the active site (RMSD on C$_\alpha$ ≈1.3 Å) (18). Significant differences (RMSD > 2 Å) occur at the level of 1) the BC-loop (Arg82-Pro85), 2) the C terminus (C term) of the F helix and the N terminus (N term) of the G helix (Ser175-Glu188) and the loop between, 3) the C term of the H helix and the HI-loop (Leu219-His228), and 4) the β-hairpin β4 (Gly394-Met395). All these displacements result in a slight closure of the active site when 6DEB is bound with respect to CLT-bound OleP. Phe84 on the BC loop rotates of almost 90° after the entrance of 6DEB into the active site, driven by a hydrophobic interaction that the residue establishes with the methyl group at the C2 of the aglycone (Fig. 3*B*). The BC loop follows the Phe84 rotation and moves of ≈2.5 Å.

The remaining 3 monomers of OleP-6DEB$_{LS}$ adopt a conformation in which the accessibility to the active site is strongly hindered (Fig. 3*A*), which we consequently indicate as closed. A nearly identical closed conformation (RMSD on C$_\alpha$ ≈ 0.4 Å) was observed in the 6 monomers identified in the asymmetric unit of the OleP-6DEB$_{HS}$ structure, where the complex crystallized in the C2 space group (R_{work} = 20.8% and R_{free} = 25.8%). Two main differences between the closed OleP-6DEB$_{LS}$ monomers were observed for some external amino acids [*i.e.*, Arg208 (RMSD = 2.83 Å) at the C term of the G helix and in the segment between Val136 and Gly139, with a maximum RMSD for His138 of ≈ 2.2 Å], most likely related to the different packing in an asymmetric unit.

Closure of the active site is achieved mainly by a large (≈10–12 Å) movement of the FG portion (helices F and G
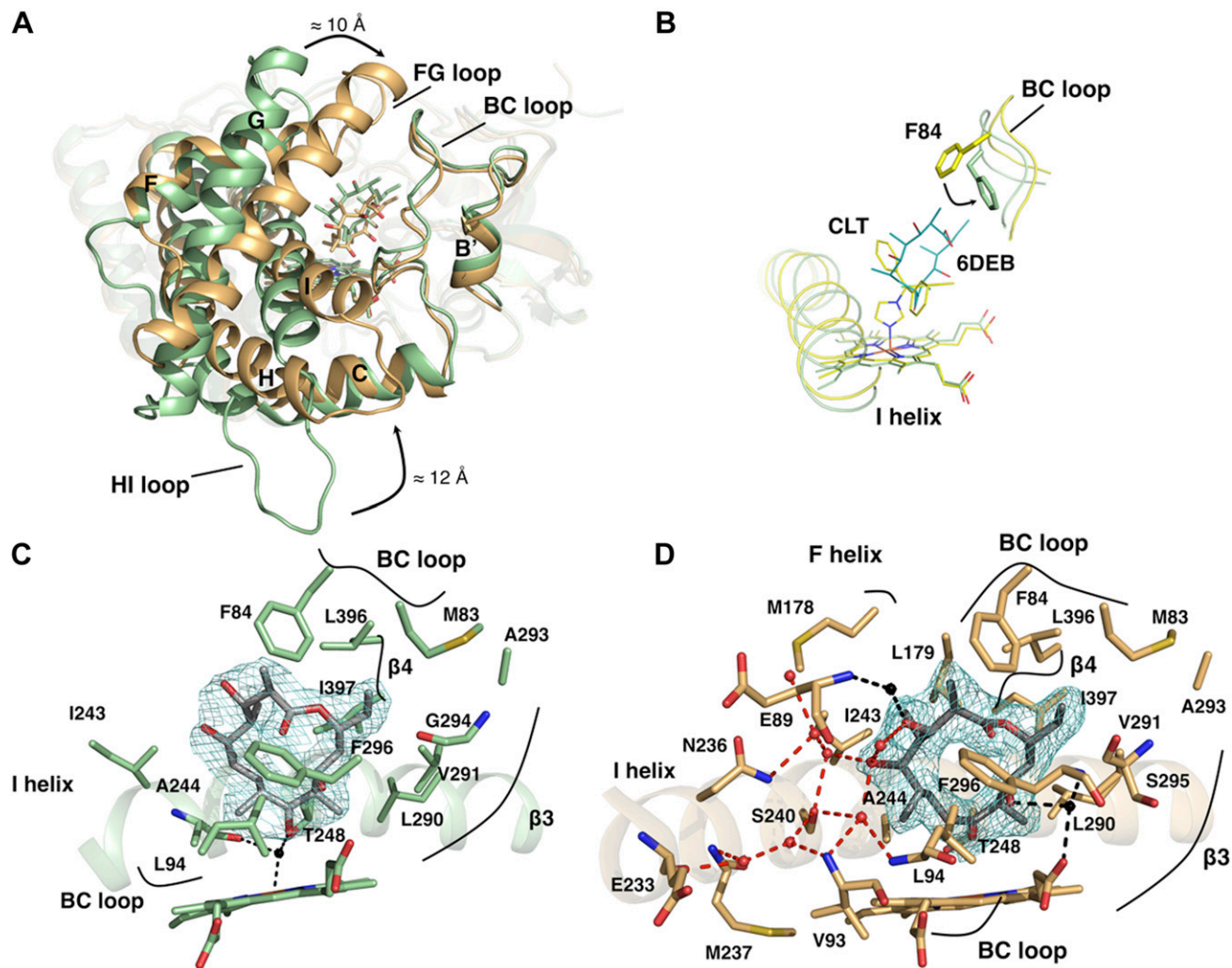
Figure 3. OleP-6DEB complex. *A*) Superposition of the open (green) and closed (wheat) conformations of OleP-6DEB complex, shown in a nonstandard orientation for P450s that allows clear visualization of the structural transition. Secondary structure elements mainly involved in the conformational change are emphasized. *B*) Superposition of the open OleP-CLT [yellow, PDB 4XE3 (16)] and the open OleP-DEB complex (green) that details the movement of Phe84 that occurs upon 6DEB binding. *C, D*) Close-up views of the active sites of open (*C*; green) and closed (*D*; wheat) OleP bound to 6DEB. Residues and water molecules within 5 Å from the substrate are displayed. Red spheres: water; dashed lines: hydrogen bonds. Gray sphere: water that mediates interaction between protein and substrate. Secondary structural elements and amino acids are labeled. In both panels, the electron density map ($2F_o$–F_c) contoured at 1σ around 6DEB is also shown.

and the FG loop) and of the HI loop respectively at the top and bottom of the active site (Fig. 3*A*).

The initial omit maps, calculated without inclusion of 6DEB in the model, revealed clear electron density in the active site of all OleP molecules from both crystal forms, corresponding to a single molecule of the macrolide. When the protein is in the open conformation, 6DEB is nearly perpendicular to the protoporphyrin IX ring, assuming slightly different positions, depending on the monomer. This mobility appears to be a consequence of the absence of specific interactions within the active site residues that could constrain the substrate in a unique orientation. In fact, 6DEB establishes only van der Waals interactions with hydrophobic residues belonging to the BC loop (Met83, Phe84, and Leu94), the I helix (Ile243, Ala244, and Thr248), the β-hairpin β3 (Leu290, Val291, Ala293, Gly294, and Phe296), and the β-hairpin β4 (Leu396 and Ile397)

(Fig. 3*C*). Moreover, in open OleP, 6DEB is located at a distance of ~6 Å from the catalytic Fe, without displacing the water molecule in the sixth coordination position, typical of the resting state of the enzyme, that provides a hydrogen bond to the C9—OH of the substrate and to the carbonyl group of Ala244 on the I helix (Fig. 3*C*).

The electron density of 6DEB in the closed monomers unequivocally locates the ligand deeper in the active site, ~2 Å closer to the heme iron, without variability among the 6 OleP copies in the asymmetric unit. The closure of the active site creates a chamber that extends the network of van der Waals interactions with 6DEB, recruiting residues belonging to the F helix (Met178 and Leu179) and favoring the formation of the following H-bonds: 1) C3—OH of the substrate and a water molecule H-bonded with the amine group of Glu89 and 2) the C11-OH of the substrate and a

water molecule, H-bonded with the amine group of Phe296 on β-hairpin β3 and with the C2A-propionate of the heme (Fig. 3D; the 2 water molecules are highlighted in gray).

A closer look at the regions of OleP that contact the ethyl moiety at the C13 of 6DEB (methyl in DEO) to analyze possible structural effects on OleP related to the use of a nonphysiologic substrate shows that this group is located in a small hydrophobic cavity where no specific interactions or clashes with Met83 and Phe84 on the BC loop and Leu396 on the β-hairpin β4 at ~4 Å are formed (Supplemental Fig. S3B). As a possible aspect of the structural effect of utilization of the alternative substrate, because the methyl group at C13 of DEO would be placed even farther from those residues, we may hypothesize that Met83 side chain would assume an alternative conformation, in place of the 6DEB C13 ethyl C2 carbon.

Moreover, different from the open OleP-6DEB, a set of water molecules detectable only in the OleP-6DEB$_{HS}$ closed structures was identified inside a small pocket lined by the I helix, the BC loop, and the substrate. These molecules form a network of H-bonds that connects the 2 hydroxyl moieties at C3 and C5 of the macrolactone ring of 6DEB with residues Glu233, Asn236, Met237, and Ser240 on the I helix and residues Glu89, Val93, and Leu94 on the BC loop (Fig. 3D). These water-mediated connections further reduce the mobility of the substrate and constrain it to adopt a unique position.

The slight tilt of 6DEB toward the BC loop observed when OleP adopts the closed conformation reduces the distance between the C8—C8a bond and the heme iron to ~4 Å and induces the displacement of the sixth heme-coordinating water molecule, thus breaking the H-bond contacts with Ala244 and the substrate, which rearranges in a catalytically competent location and orientation (12).

Because the differences between copies of OleP-6DEB adopting the same conformation (open and closed) are not significant and are likely to reflect the quality of the individual electron density maps (Supplemental Fig. S4), throughout this article we have referred to the better defined monomer A from OleP-6DEB$_{LS}$ and monomer C from OleP-6DEB$_{HS}$ as representative of the open and closed state, respectively. We have thoroughly checked that the individual structural features discussed are present in all independent monomers, when allowed by the quality of the respective electron density map.

Structural determinants for the open-to-closed transition

The crystallization of the OleP-6DEB complex in the open and closed conformations offers 2 snapshots along the dynamics of the P450 OleP, unveiling the existence of distinct states explored by the epoxygenase in the substrate-bound form that are associated with the position of the macrolide in the active site. This large structural rearrangement mainly involves secondary structural elements that define the substrate binding region (i.e., the F and G helices; the FG, HI, and BC loops; and the internal I helix; **Fig. 4**).

We performed an analysis of the open and closed OleP-6DEB structures in conjunction with open CLT-bound OleP (18) to shed light on the structural determinants and mechanism of the open-to-closed transition. The comparison between open OleP-CLT and open OleP-6DEB$_{LS}$ shows that the specific bulk of the macrolide substrate triggers a contraction of the active site in addition to the notable downward shift of the BC loop, driven by the rotation of Phe84 which establishes a van der Waals interaction with 6DEB (Fig. 3B).

Searching for specific interactions between the substrate and the I helix, which induce a productive structural transition in some P450s (10, 33), we found that when OleP is open, 6DEB mostly contacts the I helix at the level of Ile243, Ala244, and Thr248 (Fig. 4A). The latter residue represents in OleP the conserved threonine responsible for proton delivery during the catalytic cycle, corresponding to Thr252 in P450$_{cam}$ (34), which is commonly flanked by the acidic residue, namely Glu247 in OleP. In this state, these interactions do not induce the 1-turn distortion responsible for the formation of the typical catalytic cleft used for proton delivery in P450s (Supplemental Fig. S5A) (34). Unaltered geometry of the I helix has also been observed when CLT coordinates the heme iron (18).

On the other hand, in closed OleP-6DEB, both the displacement of the sixth coordinating water molecule and the approach of the substrate to the heme iron reduce the distance between 6DEB and the I helix, inducing a change in the conformation of the main chain of Ile243, Ala244, and Thr248 to compensate for the larger steric hindrance of the ligand (Fig. 4A). As a result, the carbonyl group of Ala244 rotates and moves toward the I helix tunnel, establishing an H-bond with the side chain of Thr248. As a consequence, the helical organization of the main chain in the Ala244–Thr248 turn is lost, leading to: 1) distortion of the I helix, 2) change in conformation of residues belonging to the helical turn, and 3) formation of the canonical P450 catalytic cleft (Supplemental Fig. S5A). The carbonyl group of Gly245, which in the open OleP-6DEB$_{LS}$ and OleP-CLT structures forms the intrahelical hydrogen bond with the amine of Ser249, leaves the helical geometry, and establishes an H-bond with a water molecule, itself H-bonded to the side chain of Glu365 on the L helix (Fig. 4B). This network of interactions, starting from Thr248 and ending with Glu365, both conserved among P450s, defines the internal solvent channel that allows the access of protons necessary for the catalytic cycle in OleP.

The effect of the I helix distortion induced by productive binding of the substrate is not limited to this local environment. By introducing a kink in the middle of this long helix, the helical distortion represents the epicenter of a dramatic rearrangement of secondary structure elements that expands over the vault of the binding site, leading to its sealing by means of the collapse of the FG loop over the BC loop (Supplemental Movie S1). We detail below the main structural interactions that connect and amplify the distortion of the 244–248 turn of the I helix, leading to OleP closure.

The kink of the helical backbone at the level of the 244–248 segment does not affect the C term portion of the I helix, but induces a swinging movement and ≈15°
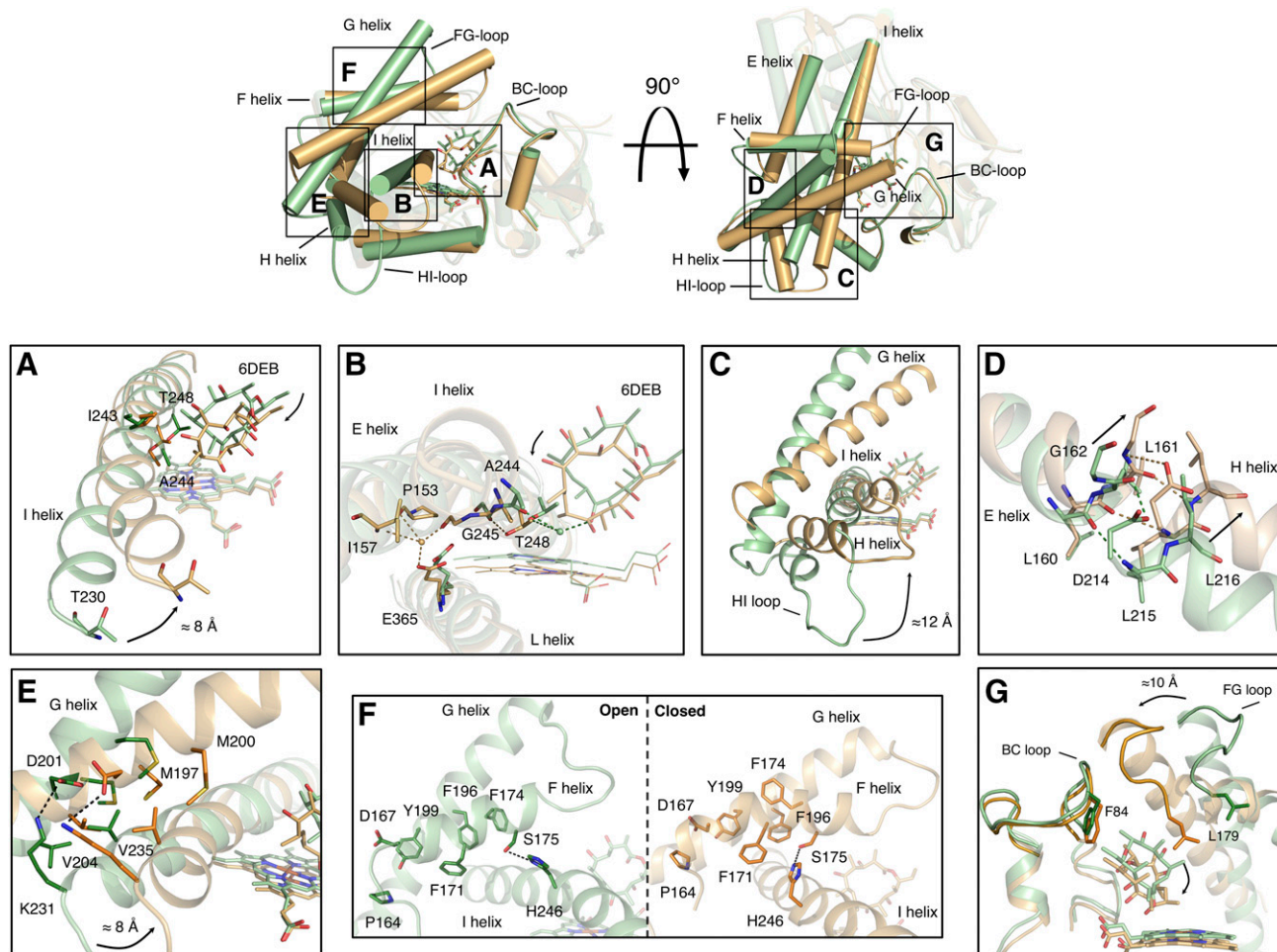
Figure 4. Structural determinants of the open-to-closed transition of OleP. At the top, open (green) and closed (wheat) structures of OleP-6DEB complex are superposed and shown in orientations that emphasize the structural transition, highlighting the secondary structure elements mainly involved in the conformational change. Helices are represented as cylinders. Detail of individual regions and amino acids involved in the transition: I helix distortion (*A*), functional water channel (*B*), HI-loop shortening and H helix fold (*C*), connections between E and H helices (*D*), interactions between G and I helices (*E*), π–π chain connecting F and G helices (*F*), collapse of the FG-loop (*G*). Red spheres: waters; dashed lines: hydrogen bonds.

rotation of the I helix N term portion (231–248), with a maximum 8.6 Å displacement of the α carbon of Thr230, the first I helix amino acid (Fig. 4*A*).

The H helix, coupled to the moving N term half of the I helix *via* the HI loop and interhelical hydrophobic contacts, follows the I helix swing and causes the reshaping of the HI loop, that folds into an additional helical turn at the C term of the H helix. As a consequence of the HI loop shortening, the H helix is lifted and pushes the C term of the G helix upward (Fig. 4*C*).

The E helix also follows the I helix swing, as a result of the head-to-tail interaction with the H helix *via* H-bonds between the main chains (Leu160–Leu215 and Leu161–Leu216) and main chain–side chain (Gly162–Asp214), as if the E helix behaves like an N term extension of the H helix itself (Fig. 4*D*).

Moreover, the G helix is coupled to the swinging of the N term half of the I helix by the Asp201–Lys231 (mid G helix–N term of I helix) salt bridge and by Val235 on the I helix, which is hooked to the G helix within a niche formed by Met197, Met200, Val204, and Asp201. All the aforementioned interactions are conserved in the open and closed structure, indicating that the Val235 hook couples secondary elements during the open-to-closed transition (Fig. 4*E*).

The combined action of swing and rotation of the I helix, the lever exerted by the H helix, and the shift of the last turn of the E helix have a dramatic effect on the F helix–FG loop–G helix segment, which moves as a rigid body (Fig. 4, top). The H-bond between Ser175 (F helix) and His246 (I helix) acts as an additional link that drives the FG segment rearrangement, as His246 is rotated toward the active site (Fig. 4*F*).

The repositioning of the FG segment acts as a lid over the active site entrance and seals it by means of the contact between the BC loop and Leu179. The latter residue is exposed to the solvent when OleP is open, and, upon transition to the closed conformation, enters into the enzyme cavity to interact directly with the C4-methyl group of 6DEB, thus participating in the ensemble of interactions that univocally define the position of 6DEB into the distal pocket upon OleP closure (Fig. 4*G*).

Moreover, the shift of the FG segment determines the disruption of the network of π–π stacking interactions involving Phe171 and Phe174 (F helix) and Phe196 and Tyr199 (G helix) packed onto Pro164 in the EF loop (Fig. 4*F*), previously identified in the open OleP-CLT structure (18). The entry and the initial positioning of 6DEB into the OleP active site, which leaves the protein in an open conformation (OleP-6DEB$_{LS}$), has no effect on the arrangement of the above-mentioned π-chain (Supplemental Fig. S5*B*), whereas, upon the open-to-closed transition, a marked movement is observed at the expenses of the aromatic residues involved in the interaction, thus altering its organization while maintaining the cohesion between helices (Fig. 4*F*).

To summarize, closure of the active site is observed only when 6DEB is deeply bound to the OleP active site in a position suitable for catalysis. Such transition appears to originate from the interaction of the substrate with the central residues of the I helix, and it extends from the I helix to the external HI and FG units *via* a complex ensemble of interactions. The pronounced movement of the phenyl moiety of Phe84 on the BC loop that follows the entrance of the substrate in the open OleP-6DEB structure allows the closure by the FG loop, which brings Leu179 onto the substrate. Both Phe84 and Leu179 constrain the position of 6DEB into the active site.

We checked crystal contacts in the mobile FG and HI loops, and we observed that crystal interactions are absent in the open copies or different among the closed copies, indicating that the crystal lattice does not play a major role in inducing the open-to-closed transition.

Therefore, we hypothesize that 6DEB-bound OleP exists in solution in an equilibrium between 2 conformations, whereby the open form represents an initial binding intermediate that evolves into the closed form, adopting the catalytically competent binding organization.

Substrate binding to OleP proceeds through multistep kinetics

The structures of the OleP–6DEB complex revealed an open and a closed conformation for the substrate-bound form. It is then reasonable to hypothesize the existence of an equilibrium between the relative conformers in solution. In an effort to validate this hypothesis and to investigate the mechanism underlying substrate binding to OleP, we monitored binding during the pre–steady state using a stopped-flow apparatus, after the heme absorbance increase at 390 nm occurring upon 6DEB binding. Experiments were performed under pseudo first-order conditions in symmetric mixing mode, by challenging OleP with increasing concentrations of 6DEB.

For all explored ligand concentrations, single exponential fits proved unsatisfactory, suggesting at least biphasic kinetics. The kinetic traces were indeed satisfactorily fitted to a double exponential equation, as judged by the residual analyses (**Fig. 5*A***). The rate constants of the fast phase correspond to $k_{on} = (2.3 \pm 0.1) \times 10^6$ M/s and $^{app}k_{off} = 21.3 \pm 3.8$/s, whereas for the slow phase, essentially independent on 6DEB concentration, the rate constant was estimated to be 1.59 ± 0.06/s. The sum of amplitudes, characteristic of the fast and the slow phase, changes hyperbolically with 6DEB concentration (Supplemental Fig. S6*A*) and replicates the behavior observed in equilibrium experiments. Analysis of this curve allows us to estimate the apparent K_D, yielding a value of 2.0 ± 0.9 μM, which is in good agreement with equilibrium
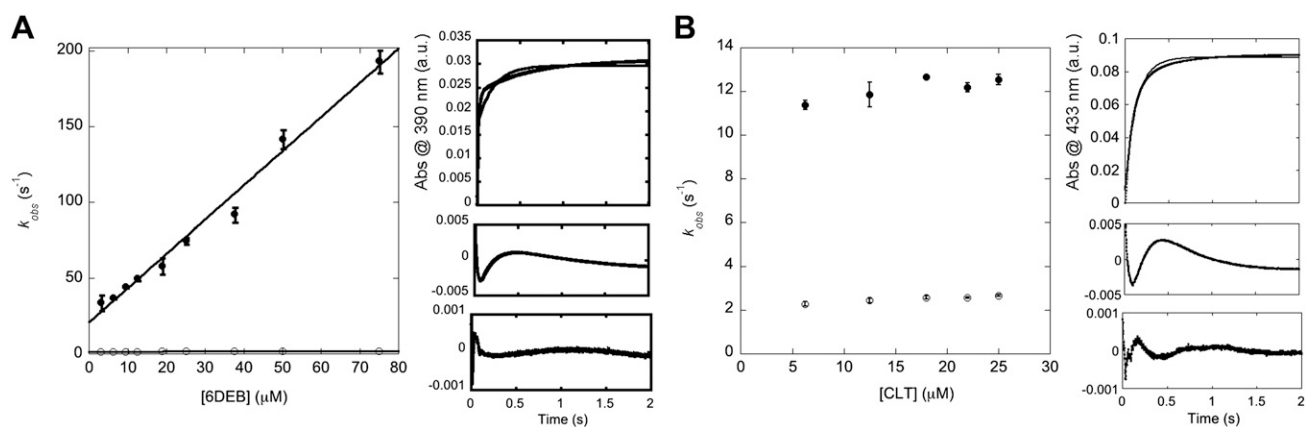

Figure 5. Pseudo first-order kinetics of the 6DEB binding to OleP and CLT displacement kinetics. *A*) The k_{obs} of the fast phase (filled circles) and the slow phase (open circles) are reported as a function of the substrate concentration. The former increases linearly with the concentration of 6DEB, whereas the latter is essentially independent from substrate concentration. Error bars indicate the SD on the k_{obs} obtained by averaging 3 traces for each ligand concentration. Error bars that are not visible are smaller than the size of the symbol used. On the right, the trace as the averages of 3 independent acquisitions obtained at 19 μM 6DEB is shown. The residuals to a single (top) and double (bottom) exponential fit are reported in panels below the trace. Thin line: the fit to a single exponential decay. *B*) The k_{obs} of the fast phase (filled circles) and the slow phase (open circles) of the displacement reaction of the OleP-6DEB complex with CLT are reported as a function of CLT concentration. Error bars representing the SD on the k_{obs} are shown. Right: trace as the averages of 3 independent acquisitions obtained at 25 μM CLT is shown. The residuals to a single (top) and double (bottom) exponential fit are reported in panels below the trace. Thin line: the fit to a single exponential decay.

titration (K_D = 5.1 ± 0.2 µM), suggesting the absence of additional phases in the binding time course.

Previous work on other P450s showed that complex substrate binding kinetics, as observed for 6DEB binding to OleP, can be ascribed to the binding of multiple substrate molecules to a P450 enzyme (26, 35–37), rearrangement of the protein into oligomeric structures (38, 39), or an event of conformational transition (10). Inspection of the substrate-bound structures revealed that the active site of the enzyme in both OleP conformers accommodates only 1 molecule of 6DEB, even when it is present in large excess. Furthermore, OleP is monomeric in our experimental conditions, as previously assessed by size exclusion chromatography (18).

Therefore, the complex kinetics observed for OleP-6DEB binding could arise from an event of conformational transition, assuming either the *a priori* existence of a heterogeneous ensemble of conformations that the enzyme explores in solution (conformational selection model), or a mechanism where substrate binding triggers the structural reorganization (10, 39–41).

To further investigate the substrate binding mechanism, we followed the displacement kinetics of 6DEB using CLT as a binding competitor. Previous structural and functional data showed that CLT stably locks OleP in an open conformation by following a simple 1-step binding mechanism (18). Moreover, compared to 6DEB, CLT showed a significantly higher affinity (2 orders of magnitude, $K_D \approx$ 30 nM) toward OleP, likely due to the direct coordination that it establishes with the heme iron (42).

Therefore, the tight binding capacity of CLT to the enzyme with an extremely slow k_{off}, which can be considered an irreversible process on the time scale of the stopped-flow experiment, was exploited to perform displacement kinetic measurements. OleP was preincubated with saturating concentrations of 6DEB (0.1 mM), according to the K_D value obtained from equilibrium binding analysis. The OleP-6DEB complex was then rapidly mixed with a solution containing a saturating concentration of CLT and inhibitor binding was monitored at 433 nm over time. In this condition, if the reaction was only limited by the dissociation rate of 6DEB, CLT displacement should result in a single exponential with a rate constant that is the k_{off} of 6DEB. Nevertheless, the kinetic traces of CLT binding were biphasic, suggesting a more complex mechanism for CLT binding to the OleP-6DEB complex with respect to what observed in binding to ligand-free OleP (Fig. 5B) (18). Different concentrations of CLT were used in the experiment, ranging from 6 to 25 µM after mixing, but the observed rate constants for both phases were found to be independent of the inhibitor concentration, given the CLT saturating conditions. In this experiment, the fast phase would account for the event of substrate displacement, yielding $^{fast}k_{obs}$ = k_{off} = 12.5 ± 2.8/s, which is not inconsistent, yet is more accurate, with that estimated by extrapolation ($^{app}k_{off}$ = 21.3 ± 3.8/s). The additional slow phase ($^{slow}k_{obs}$ = 2.7/s), not observed when CLT binds unliganded OleP (18), could arise from a conformational transition, suggesting the existence of a heterogeneous ensemble of populations that forms when OleP is bound to 6DEB.

This result, combined to the complex kinetics observed for 6DEB binding and to the structures herein reported, suggests that the mechanism of 6DEB binding to OleP proceeds through at least 3 steps that involve the accumulation of an intermediate where 6DEB is bound and OleP is in an open conformation. In substrate-binding kinetic experiments, the fast phase would be related to the entrance of 6DEB in the OleP active site, whereas the slow phase would account for the conformational transition of OleP induced by 6DEB, resulting in the displacement of the heme sixth coordinating water molecule and in the closure of the active site, as observed in the crystallographic structure, which would induce a further adjustment of the iron spin equilibrium.

In an effort to measure the absorption changes associated to each kinetic phase, 6DEB binding to OleP was investigated by multiwavelength UV-visible absorption stopped-flow spectroscopy, and the set of time-resolved spectra acquired after rapid mixing of the protein with 6DEB was analyzed by global fit analysis. According to this analysis, most of absorption changes occurred during the fast phase of the reaction, and the 2 kinetic phases observed upon 6DEB binding proved to be associated with optical transitions somewhat similar to each other, yet distinct in shape (**Fig. 6**). The latter observation is in line with the hypothesis that the slow phase accounts for a conformational transition in the protein that takes place after 6DEB binding to open OleP, associated with completion of the heme spectroscopic transition.

As further confirmation of our hypothesis and with the purpose of directly measuring the rate constant for the structural rearrangement that OleP undergoes in the presence of 6DEB, a double-jump mixing experiment was designed to alter the relative populations of the 2 conformers. The protein was initially mixed with a saturating concentration of 6DEB (premix), and, after a variable delay time, the premix was mixed with CLT at a saturating concentration. At all different delay times, the observed time courses were fitted to double exponentials. As expected, the observed rate constants for both phases were independent of time delays between the first and the second mix (Supplemental Fig. S6B). On the other hand, the amplitudes of the 2 phases displayed a single exponential dependence on the delay times, yielding an apparent rate constant for the conformational transition of $k_1 + k_{-1}$ = 2.8 ± 1.0/s (**Fig. 7A**). Whereas the amplitude of the fast phase, accounting for the OleP-6DEB population of molecules assigned to open conformers, decreased with increasing delay times, the amplitude of the slow phase, which accounts for the population of OleP-6DEB molecules assigned to closed conformers, increased. The dependence on the delay times of the amplitudes of the slow and fast phases of the displacement reaction also allows estimation of the fraction of open and closed OleP-6DEB at equilibrium. The $A_{slow}/(A_{slow} + A_{fast})$ ratio accounts for the fraction of closed OleP-6DEB, whereas $A_{fast}/(A_{slow} + A_{fast})$ represents the fraction of open OleP-6DEB, where A_{slow} = 0.026 and A_{fast} = 0.044. From this analysis, the closed conformation resulted to be almost as populated (40%) as the open one (60%) at the equilibrium (*i.e.*, at infinite delay time).
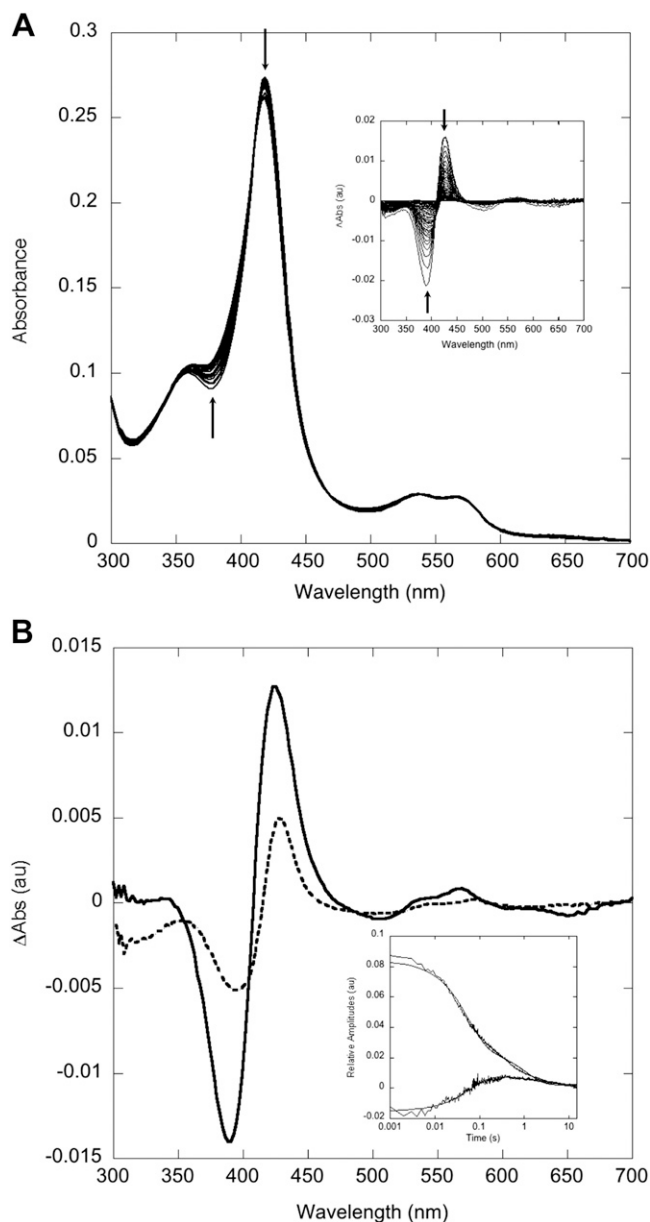
Figure 6. Binding of 6DEB to OleP investigated by multiwavelength stopped-flow spectroscopy. *A*) UV-visible absorption spectra collected over 20 s after stopped-flow mixing OleP with 6DEB at 288 K. Arrows: the direction of absorption changes. Inset: absorption changes relative to the last spectrum acquired at 20 s after mixing. *B*) Optical species relative to the fast (solid line) and slow (dotted line) phases observed on 6DEB binding, as obtained by global fit analysis of the time-resolved spectra shown in *A*. Inset: best fit to double exponentials of the first 2 V columns of the SVD output scaled for their corresponding singular values. Fitted rate constants: $k_1 = 19.3/s$ and $k_2 = 0.9/s$.

Notably, the observed rate constant for the equilibrium between the open and closed conformation of OleP-6DEB obtained by the double-mixing experiment ($k_{obs} = 2.8 \pm 1.0/s$) is consistent with the rate constant value determined in the displacement kinetics ($^{slow}k_{obs} = 2.7/s$) and it is of the same order of magnitude as the slow phase of the OleP–6DEB binding kinetics ($^{app}k_{obs} = 1.59 \pm 0.06/s$). An overall interpretation of crystallographic and

kinetic data of the binding of OleP to 6DEB is given in the energy diagram in Fig. 7*B*.

DISCUSSION

P450 enzymes endowed with high substrate specificity evolved a mechanism that leads to confined oxygen activation and to the entrapment of target molecules by means of active site closure with transitions of variable extent, depending on the size of the ligand. Electron entry in P450s is often coupled to substrate binding that also induces the formation of a catalytic cleft on the I helix to generate a proton relay network and the displacement of the H_2O molecule coordinated in the heme axial position (12).

In the case of OleP, we describe how a limited set of short-range interactions established by 6DEB on the I helix leads to the appearance of catalytic activation structural signatures and expands within the protein fold, inducing the active site closure by extensive repositioning of secondary structure elements. Entry of 6DEB into the binding cavity, as observed in the OleP–6DEB_LS open structure, induces the flipping of Phe84 on the BC loop and a compacting of the active site in a conformation that we interpret as a preactivation complex and that we assign to the kinetic intermediate detected by the multiwavelength stopped-flow experiment. This assignment relies also on the observation that this intermediate allows rapid 6DEB displacement by CLT. Although it is not possible to directly assess the redox activity of the open-bound state of OleP, the presence of water bound to the heme iron in its crystal structure is a diagnostic feature of low spin/low redox potential. Therefore, we infer that this form of OleP is not reducible by ferredoxin, and thus it would not contribute to catalysis or to uncoupling.

In summary, crystallization allowed us to capture the OleP-6DEB complex in open and closed states. To our knowledge this is the first time that a complex of a P450 with a substrate has been described in the early (open) and late (closed) stages of binding. Structural data allowed us to mechanistically describe the cascade of interactions responsible for the conformational change and to propose an overall model for the multistep process of substrate binding, which is consistent with an induced-fit scenario.

Based on our data on OleP and on structures of other bacterial P450s such as P450_cam, P450 monooxygenase from *Streptomyces venezuelae* (PikC), P450 C12-hydroxylase from *Streptomyces erythraeus* (EyrK), P450 monooxygenase from *Bacillus megaterium* (P450BM3), P450 monooxygenase from *Sulfolobus solfataricus* (CYP119A), and mycinamicin monooxygenase from *Micromonospora griseorubida* (MycG) (8, 10, 11, 43–47), we support the mechanism of substrate sensing by the conserved A/G_{n-1}-G_n-XX-T_{n+3} motif in the center of the I helix (12) which allosterically drives transitions within the protein fold as a general feature of biosynthetic bacterial P450s and as the basis of the regulation of potentially detrimental oxygen activation. Consequently, substrate-bound bacterial P450 structures where catalytic competence signatures are missing (*i.e.*, water channel and cleft on the I helix, heme pentacoordination), should be
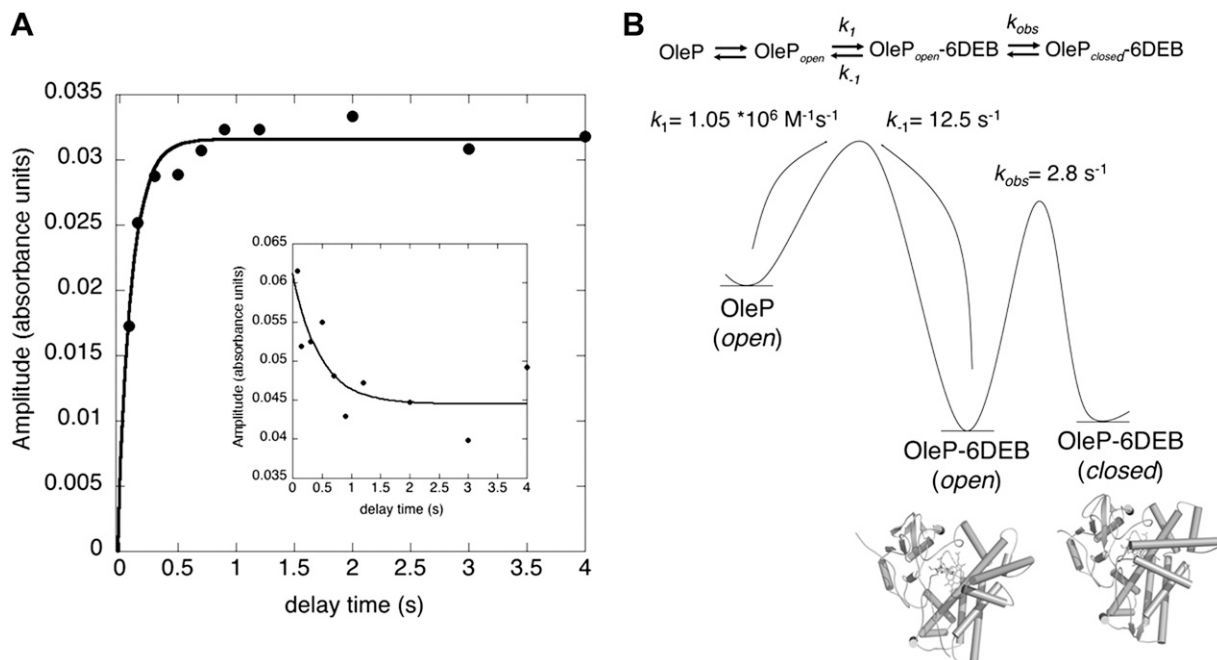
A

B

$$\text{OleP} \rightleftharpoons \text{OleP}_{open} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{OleP}_{open}\text{-6DEB} \overset{k_{obs}}{\rightleftharpoons} \text{OleP}_{closed}\text{-6DEB}$$

$k_1 = 1.05 *10^6 \text{ M}^{-1}\text{s}^{-1}$ $k_{-1} = 12.5 \text{ s}^{-1}$

$k_{obs} = 2.8 \text{ s}^{-1}$

OleP *(open)*

OleP-6DEB *(open)*

OleP-6DEB *(closed)*

Figure 7. Double-jump mixing experiment and schematic energy diagram of substrate binding to OleP. *A*) The amplitudes of the 2 phases are reported as a function of the delay time between the first and second mix. The amplitudes of the slow phase and of the fast phase (shown in the inset) were globally fitted to a single exponential function, yielding a value of $k_{obs} = 2.8 \pm 1.0/\text{s}$. Since the fast phase of 6DEB binding is complete after 0.08 s, the delay times allow following only the build-up of the OleP–6DEB species that binds CLT with slow kinetics. *B*) Schematic energy diagram showing the binding reaction of OleP to 6DEB. Representative structure of open OleP-6DEB and closed OleP-6DEB are given respectively in gray ribbon representation.

considered as preactivation complexes requiring further rearrangements to reach the competent conformation for catalysis (9).

[FJ]

AUTHOR CONTRIBUTIONS

C. Savino and B. Vallone conceived the study; G. Parisi, L. C. Montemiglio, C. Savino, and B. Vallone planned the experiments, interpreted the data, and wrote the manuscript; G. Parisi, A. Scaglione, G. Cerutti, and C. Exertier performed the expression and purification of the P450 OleP; G. Parisi, L. C. Montemiglio, and A. Macone performed the enzymatic assays and GC-MS measurements; A. Giuffrè performed the diode array experiment and data analysis; G. Parisi, L. C. Montemiglio, A. Scaglione, C. Savino, and B. Vallone crystallized the OleP-6DEB and determined the X-ray structures; and G. Parisi, L. C. Montemiglio, and B. Vallone designed and performed equilibrium and kinetic measurements of OleP-6DEB binding.

REFERENCES

1. Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005) Structure and chemistry of cytochrome P450. *Chem. Rev.* **105**, 2253–2277

2. Bernhardt, R. (2006) Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* **124**, 128–145

3. Poulos, T. L., and Johnson, E. F. (2015) Structures of cytochrome P450 enzymes. In *Cytochrome P450: Structure Mechanism and Biochemistry.* (Ortiz de Montellano, P. R., ed.), 4th ed., pp. 3–32, Kluwer Academic/Plenum Publisher, New York

4. Balding, P. R., Porro, C. S., McLean, K. J., Sutcliffe, M. J., Maréchal, J. D., Munro, A. W., and de Visser, S. P. (2008) How do azoles inhibit cytochrome P450 enzymes? A density functional study. *J. Phys. Chem. A* **112**, 12911–12918

5. Ekroos, M., and Sjögren, T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. USA* **103**, 13682–13687

6. Isin, E. M., and Guengerich, F. P. (2006) Kinetics and thermodynamics of ligand binding by cytochrome P450 3A4. *J. Biol. Chem.* **281**, 9127–9136

7. Lee, Y. T., Wilson, R. F., Rupniewski, I., and Goodin, D. B. (2010) P450cam visits an open conformation in the absence of substrate. *Biochemistry* **49**, 3412–3419

8. Li, H., and Poulos, T. L. (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Biol.* **4**, 140–146

9. Pochapsky, T. C., Kazanis, S., and Dang, M. (2010) Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid. Redox Signal.* **13**, 1273–1296

10. Savino, C., Montemiglio, L. C., Sciara, G., Miele, A. E., Kendrew, S. G., Jemth, P., Gianni, S., and Vallone, B. (2009) Investigating the structural plasticity of a cytochrome P450: three-dimensional structures of P450 EryK and binding to its physiological substrate. *J. Biol. Chem.* **284**, 29170–29179

11. Sherman, D. H., Li, S., Yermalitskaya, L. V., Kim, Y., Smith, J. A., Waterman, M. R., and Podust, L. M. (2006) The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from Streptomyces venezuelae. *J. Biol. Chem.* **281**, 26289–26297

12. Podust, L. M., and Sherman, D. H. (2012) Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep.* **29**, 1251–1266

13. Shah, S., Xue, Q., Tang, L., Carney, J. R., Betlach, M., and McDaniel, R. (2000) Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. *J. Antibiot. (Tokyo)* **53**, 502–508

14. Tatsuta, K., Gunji, H., Tajima, S., Ishiyama, T., Imai, S., Okuyama, S., and Fukatsu, S. (1990) Biosynthetic studies on oleandomycin by incorporation of the chemically synthesized aglycones. *J. Antibiot. (Tokyo)* **43**, 909–911

15. Quirós, L. M., and Salas, J. A. (1995) Biosynthesis of the macrolide oleandomycin by Streptomyces antibioticus: purification and kinetic characterization of an oleandomycin glucosyltransferase. *J. Biol. Chem.* **270**, 18234–18239

16. Gaisser, S., Lill, R., Staunton, J., Méndez, C., Salas, J., and Leadlay, P. F. (2002) Parallel pathways for oxidation of 14-membered polyketide macrolactones in Saccharopolyspora erythraea. *Mol. Microbiol.* **44**, 771–781

17. Rodriguez, A. M., Olano, C., Méndez, C., Hutchinson, C. R., and Salas, J. A. (1995) A cytochrome P450-like gene possibly involved in oleandomycin biosynthesis by Streptomyces antibioticus. *FEMS Microbiol. Lett.* **127**, 117–120

18. Montemiglio, L. C., Parisi, G., Scaglione, A., Sciara, G., Savino, C., and Vallone, B. (2016) Functional analysis and crystallographic structure of clotrimazole bound OleP, a cytochrome P450 epoxidase from Streptomyces antibioticus involved in oleandomycin biosynthesis. *Biochim. Biophys. Acta* **1860**, 465–475

19. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132

20. Vagin, A., and Teplyakov, A. (1998) A translation-function approach for heavy-atom location in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 400–402

21. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763

22. Pannu, N. S., Murshudov, G. N., Dodson, E. J., and Read, R. J. (1998) Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 1285–1294

23. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475

24. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501

25. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067

26. Isin, E. M., and Guengerich, F. P. (2008) Substrate binding to cytochromes P450. *Anal. Bioanal. Chem.* **392**, 1019–1030

27. Montemiglio, L. C., Macone, A., Ardiccioni, C., Avella, G., Vallone, B., and Savino, C. (2013) Redirecting P450 EryK specificity by rational site-directed mutagenesis. *Biochemistry* **52**, 3678–3687

28. Henry, E. R., and Hofrichter, J. (1992) Singular value decomposition: application to analysis of experimental data. *Methods Enzymol.* **210**, 129–192

29. Cupp-Vickery, J. R., Han, O., Hutchinson, C. R., and Poulos, T. L. (1996) Substrate-assisted catalysis in cytochrome P450eryF. *Nat. Struct. Biol.* **3**, 632–637

30. Cupp-Vickery, J. R., Li, H., and Poulos, T. L. (1994) Preliminary crystallographic analysis of an enzyme involved in erythromycin biosynthesis: cytochrome P450eryF. *Proteins* **20**, 197–201

31. Ogura, H., Nishida, C. R., Hoch, U. R., Perera, R., Dawson, J. H., and Ortiz de Montellano, P. R. (2004) EpoK, a cytochrome P450 involved in biosynthesis of the anticancer agents epothilones A and B: substrate-mediated rescue of a P450 enzyme. *Biochemistry* **43**, 14712–14721

32. Kuznetsov, V. Y., Poulos, T. L., and Sevrioukova, I. F. (2006) Putidaredoxin-to-cytochrome P450cam electron transfer: differences between the two reductive steps required for catalysis. *Biochemistry* **45**, 11934–11944

33. Montemiglio, L. C., Gianni, S., Vallone, B., and Savino, C. (2010) Azole drugs trap cytochrome P450 EryK in alternative conformational states. *Biochemistry* **49**, 9199–9206

34. Haniu, M., Armes, L. G., Tanaka, M., Yasunobu, K. T., Shastry, B. S., Wagner, G. C., and Gunsalus, I. C. (1982) The primary structure of the monoxygenase cytochrome P450CAM. *Biochem. Biophys. Res. Commun.* **105**, 889–894

35. Domanski, T. L., He, Y. A., Khan, K. K., Roussel, F., Wang, Q., and Halpert, J. R. (2001) Phenylalanine and tryptophan scanning mutagenesis of CYP3A4 substrate recognition site residues and effect on substrate oxidation and cooperativity. *Biochemistry* **40**, 10150–10160

36. Korzekwa, K. R., Krishnamachary, N., Shou, M., Ogai, A., Parise, R. A., Rettie, A. E., Gonzalez, F. J., and Tracy, T. S. (1998) Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry* **37**, 4137–4147

37. Shou, M., Grogan, J., Mancewicz, J. A., Krausz, K. W., Gonzalez, F. J., Gelboin, H. V., and Korzekwa, K. R. (1994) Activation of CYP3A4: evidence for the simultaneous binding of two substrates in a cytochrome P450 active site. *Biochemistry* **33**, 6450–6455

38. Boehr, D. D., and Wright, P. E. (2008) Biochemistry: how do proteins interact? *Science* **320**, 1429–1430

39. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118

40. Bah, A., Garvey, L. C., Ge, J., and Di Cera, E. (2006) Rapid kinetics of Na+ binding to thrombin. *J. Biol. Chem.* **281**, 40049–40056

41. Gianni, S., Dogan, J., and Jemth, P. (2014) Distinguishing induced fit from conformational selection. *Biophys. Chem.* **189**, 33–39

42. Poulos, T. L., and Johnson, E. F. (2015) Inhibition of cytochrome P450 enzymes. In *Cytochrome P450: Structure Mechanism and Biochemistry.* (Ortiz de Montellano, P. R., ed.), 4th ed., pp. 177–260, Kluwer Academic/Plenum Publisher, New York

43. Basudhar, D., Madrona, Y., Kandel, S., Lampe, J. N., Nishida, C. R., and de Montellano, P. R. (2015) Analysis of cytochrome P450 CYP119 ligand-dependent conformational dynamics by two-dimensional NMR and X-ray crystallography. *J. Biol. Chem.* **290**, 10000–10017

44. Colthart, A. M., Tietz, D. R., Ni, Y., Friedman, J. L., Dang, M., and Pochapsky, T. C. (2016) Detection of substrate-dependent conformational changes in the P450 fold by nuclear magnetic resonance. *Sci. Rep.* **6**, 22035

45. Hays, A. M., Dunn, A. R., Chiu, R., Gray, H. B., Stout, C. D., and Goodin, D. B. (2004) Conformational states of cytochrome P450cam revealed by trapping of synthetic molecular wires. *J. Mol. Biol.* **344**, 455–469

46. Lampe, J. N., Brandman, R., Sivaramakrishnan, S., and de Montellano, P. R. (2010) Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. *J. Biol. Chem.* **285**, 9594–9603

47. Tietz, D. R., Podust, L. M., Sherman, D. H., and Pochapsky, T. C. (2017) Solution conformations and dynamics of substrate-bound cytochrome P450 MycG. *Biochemistry* **56**, 2701–2714

SCIENTIFIC REP🞂RTS

OPEN

Proximal and distal control for ligand binding in neuroglobin: role of the CD loop and evidence for His64 gating

Cécile Exertier [1,2], Lisa Milazzo[3], Ida Freda [1,2], Linda Celeste Montemiglio[1,2,5], Antonella Scaglione[1,2,6], Gabriele Cerutti [1], Giacomo Parisi[1,2,7], Massimiliano Anselmi[4], Giulietta Smulevich[3], Carmelinda Savino[5] & Beatrice Vallone [1,2,5]

Neuroglobin (Ngb) is predominantly expressed in neurons of the central and peripheral nervous systems and it clearly seems to be involved in neuroprotection. Engineering Ngb to observe structural and dynamic alterations associated with perturbation in ligand binding might reveal important structural determinants, and could shed light on key features related to its mechanism of action. Our results highlight the relevance of the CD loop and of Phe106 as distal and proximal controls involved in ligand binding in murine neuroglobin. We observed the effects of individual and combined mutations of the CD loop and Phe106 that conferred to Ngb higher CO binding velocities, which we correlate with the following structural observations: the mutant F106A shows, upon CO binding, a reduced heme sliding hindrance, with the heme present in a peculiar double conformation, whereas in the CD loop mutant "Gly-loop", the original network of interactions between the loop and the heme was abolished, enhancing binding via facilitated gating out of the distal His64. Finally, the double mutant, combining both mutations, showed a synergistic effect on CO binding rates. Resonance Raman spectroscopy and MD simulations support our findings on structural dynamics and heme interactions in wild type and mutated Ngbs.

Phylogenetically ancient globins such as neuroglobin (Ngb), globin X and androglobin are endowed with heme iron hexacoordination, whereas pentacoordinated myoglobins (Mb) and hemoglobins (Hb), characterized by a complex respiratory role, appeared later in evolution. The current hypothesis is that hexacoordination could be the oldest coordination scheme. Furthermore, this iron binding mode triggers a protein conformational reorganization upon ligand binding that could play a role in gas sensing[1].

Regardless of the variety of functions that globins may perform (NO metabolism, lipid metabolism, anti-apoptosis, signal transduction, detoxification of ROS/RNS, respiration, O_2 diffusion), they share a common fold, with 8 helices named from A to H, the so-called 3-on-3 α-helical sandwich[2], and some very specific structural elements. The sequence conservation of globin structure involves two universally invariant positions: a histidine at the 8th position of the F-helix, that binds the heme iron proximally, and a phenylalanine in position CD1 that is packed against the heme[3]. Sequence alignment of a large set of globins also highlighted the strong conservation of other residues such as a histidine in position E7 on the heme distal side, which, in some cases, is involved in heme hexacoordination and/or in the stabilization of bound ligands. Sequence conservation in

[1]Dip. di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, P.le A. Moro 5, 00185, Rome, Italy. [2]Istituto Pasteur-Fondazione Cenci Bolognetti, Dip. di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, P.le A. Moro 5, 00185, Rome, Italy. [3]Dip. di Chimica "Ugo Schiff", Università di Firenze, Via delle Lastruccia 3–13, 50019, Sesto Fiorentino (FI), Italy. [4]Institute for Microbiology and Genetics, Georg-August University Göttingen, Justus-von-Liebig-Weg 11, 37077, Göttingen, Germany. [5]CNR Institute of Molecular Biology and Pathology, P.le A. Moro 5, 00185, Rome, Italy. [6]Present address: Neuroscience Initiative, CUNY Advanced Science Research Center, 85 St. Nicholas Terrace, 10031, New York, USA. [7]Present address: Department of Physiology and Cellular Biophysics, Russ Berrie Pavilion, Columbia University Medical Center, 1150 St Nicholas Ave, 10032, New York, USA. Correspondence and requests for materials should be addressed to B.V. (email: beatrice.vallone@uniroma1.it)
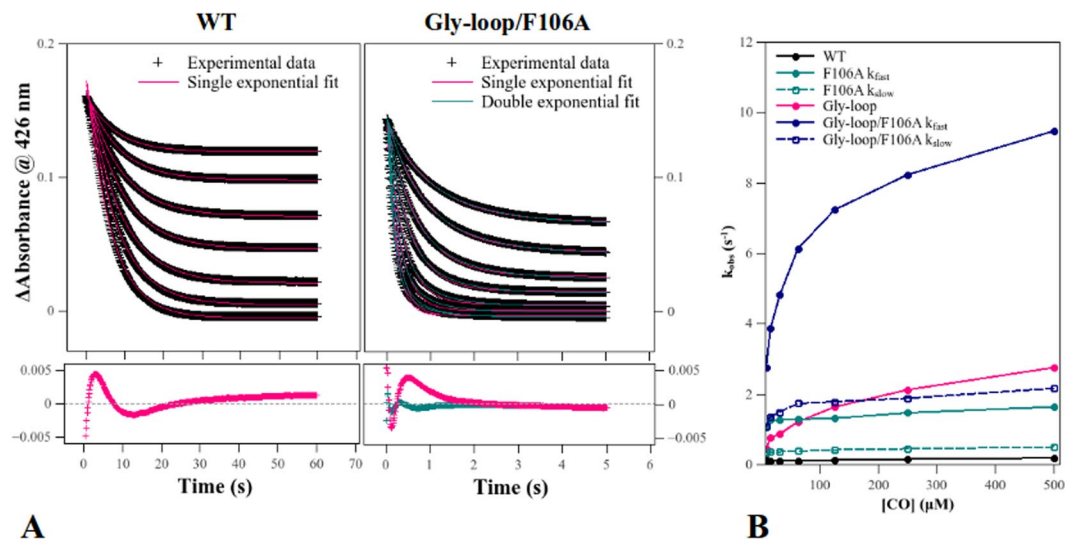
Figure 1. CO binding to ferrous neuroglobin mutants at 25 °C. CO binding rate constants k_{obs} were extracted from fitting experimental kinetic traces as single or double exponentials (**A**) and plotted as a function of CO concentration (**B**).

globins highlights residues involved in heme packing (His(F8), Phe(CD1), Leu(B10), Phe(B14), Phe(CD4)) or participating to the early stage of globin folding, with residues A8, G16, G12, A12, H5, H12 forming the A, G and H helices which act as a folding nucleus[4].

Ngb is over-expressed in hypoxic conditions and it may be involved in gas/redox sensing, radical scavenging, and/or signal transduction[5]. However, its precise mechanism of action remains to be clarified, despite Ngb function involves binding of diatomic gases, such as O_2, NO and CO. Notably, hexacoordination in Ngb accounts for a peculiar structural rearrangement and complex kinetics upon ligand binding[6–13], which takes place when the distal histidine (His(E7)64), engaged in internal iron hexacoordination, dissociates from the heme. The dissociation of His64 is the rate-limiting step for ligand binding and it triggers a repositioning of the heme which slides further inside a large internal cavity. This movement is associated with the motion of helices and loops[14–16], and probably confers to Ngb a more stable conformation. More recently, molecular dynamics (MD) simulations supported a His64 gating out movement coupled with a motion of the CD corner upon ligand binding, larger than the one described by crystallography[17,18]. Another structural feature of neuroglobin is the double insertion of the heme with 70:30 proportion[14].

Artificial point mutants of Ngb have been utilized to pinpoint major structural determinants for ligand binding regulation. Pentacoordinated Ngbs obtained by substituting the distal His64 with Leu, Val or Gln, display fast ligand binding and low heme iron auto-oxidation[6,11,19]. Proximal heme pocket mutants (M144W, and F106W) showed a hampered heme sliding, affecting both binding velocity and affinity[20]. Tyr44, in the CD loop, is an important heme-interacting residue involved in regulating binding kinetics[21] since its mutation into an aspartate yielded a 15-fold increase in O_2 binding velocity and a 1.5-fold increase in O_2 affinity. Finally, the CD loop was shown to be involved in the fine tuning of iron coordination by swapping CD corners of human Ngb and sperm whale Mb (swMb), which led to gain of hexa-coordination in swMb[22].

In this study, we spectroscopically and structurally characterized three Ngb mutants aiming at pinpointing key features: a CD loop mutant with enhanced flexibility (Gly-loop), the F106A mutant and the double Gly-loop/F106A mutant (Supplementary Fig. S1). We probed the role of heme sliding in the internal cavity, facilitated by reducing the bulk in position 106 and the role of the CD loop in governing ligand affinity and hexacoordination.

Results and Discussion

CO binding at 25 °C by rapid mixing. Wild type (WT) and mutant Ngbs were mixed with CO at 25 °C using a stopped flow apparatus. Traces were fitted as single (WT and Gly-loop, Fig. 1A) or double exponentials (F106A and Gly-loop/F106A, Fig. 1A). As observed previously[20], CO binding to WT Ngb exhibits a mono-phasic decay at 426 nm, and so does the Gly-loop mutant. However, for Phe106 mutants (F106A and Gly-loop/F106A), a double exponential trend was observed, suggesting the existence of two populations. Both Gly-loop and F106A mutations significantly increase CO binding velocities. The Gly-loop mutation has the largest effect on ligand association rates (k_{obs}(WT) = 0.19 s^{-1} vs k_{obs} (Gly-loop) = 2.8 s^{-1} at 500 μM CO) compared to the F106A mutation (k_{obs} slow(F106A) = 0.51 s^{-1} and k_{obs} fast(F106A) = 1.7 s^{-1}, Fig. 1B). However, these mutations have a synergistic effect rather than an additive one, as shown by the increase of rate constants (k_{obs} slow = 2.2 s^{-1} and k_{obs} fast = 9.5 s^{-1}) for the double Gly-loop/F106A mutant. Unlike WT and F106A Ngbs, both mutants containing the CD loop mutation (Gly-loop and Gly-loop/F106A) show a strong dependence on CO concentration in their binding rate constants k_{obs}.

Notably, CO binding to hexacoordinated Ngb is scarcely dependent on ligand concentration[20,21], since the main factor affecting binding is the spontaneous dissociation of the distal histidine (His64). This suggests that,
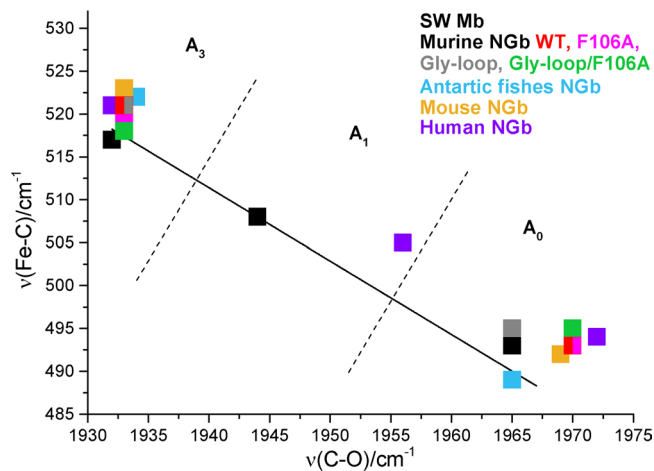
Figure 2. Back-bonding correlation line of the ν(Fe-C) and ν(C-O) stretching frequencies of various Ngbs[23–25,35]. The corresponding data of sperm whale Mb are also reported[36,37]. The dotted lines indicate the approximate delineation between the frequency zones of the A_0, A_1, and A_3 forms. The human Ngb and swMb show also a third weak H-bonded conformer (A_1) at 505/1956 and 508/1946 cm^{-1}, respectively, not reported in the Table S3.

in the Gly-loop mutant, the rupture of the His64–heme iron bond is still the rate limiting step, but an increase in its velocity (or decrease in the bond re-formation velocity) causes a marked dependence on CO concentration. For the F106A mutant the fraction of the amplitude of the two phases at 500 μM CO is 64% for the fast one and 36% for the slow one, for the Gly-loop/F106A, we observed that the slow fraction is of a lesser extent (13%, Supplementary Table S2).

For all the mutants, the increase in binding velocities is coupled to an overall augmented CO affinity with respect to WT, as calculated from the saturation at different CO concentrations. (Supplementary Fig. S2 and Table S2).

Structural Characterization of neuroglobin mutants by Resonance Raman (RR).

An increased velocity of the Gly-loop/F106A mutant in CO binding might be a consequence of the presence of some pentaco-ordinated species due to the modification of the CD loop[22]. However, resonance Raman spectroscopy did not reveal any pentacoordinated species in the ferrous form or in the photolyzed CO sample (Supplementary Fig. S3). This result is consistent with the H64L mutant behaviour (bimolecular CO rebinding rates or Ngb H64L = 230 μM.s^{-1} at 20 °C[19]), which is much faster than the fast phase detected for Gly-loop/F106A.

The RR spectra of the CO complexes of the WT, the F106A, and double Gly-Loop/F106A mutant (Supplementary Fig. S4) are very similar and show, consistently with other Ngbs[23–25], the presence of two protein conformers: an open A_0 form [ν(Fe-C) at 496 cm^{-1} and ν(CO) at 1970 cm^{-1}] and a more abundant closed one, A_3 [ν(Fe-C) at 521 (WT) and 518 (mutants) cm^{-1}, and ν(CO) at 1933 cm^{-1}], whose CO is H-bonded with His64. These data are reported in Fig. 2 which shows the back-bonding correlation line of the ν(Fe-C) and ν(CO) stretching frequencies for various Ngbs and swMb (the frequencies are reported in Supplementary Table S3).

It can be seen that, after CO binding, His64 can adopt two conformations, one that strongly interacts with the CO (closed form) and one with His64 far from the ligand (open cavity, consistent with a swung out His64). The Gly-loop mutant behaves differently, being the A_0 bands quite intense. We tentatively assigned this band to two overlapped A_0 open conformations.

CO binding kinetics of Ngb F106A is consistent with a double heme conformation.

Crystal structures of unliganded hexacoordinated F106A (1.8 Å) and carbomonoxy-F106A (F106A-CO, 1.75 Å), were determined to observe the effect of removal of the steric barrier to heme sliding posed by Phe106.

At this resolution, we clearly detect the double heme conformation (Supplementary Fig. S5A and S5B): the B conformer (canonical, 30% in ferric F106A, 27% in F106A-CO) and the A one (reversed, 70% in ferric F106A, 73% in F106A-CO). The assignment of the relative population of the two heme insertion modes is based on refinement and electron density residuals. The estimated error in this determination is therefore about 3%. As compared to WT Ngb, the structural difference between the two conformers is larger, likely due to the wider heme cavity, created by the Phe to Ala substitution that allows conformer B to adopt a deeper position within the pocket. We also observed double conformations for proximal and distal histidines.

The geometry of the His64-Fe-His96 bond, for both conformers, appears to be different with respect to WT Ngb (Supplementary Table S4) and it might contribute to reactivity with ligands. However, the most striking observation about heme conformations is the relative position of the A and B conformers, that are laterally and vertically shifted (Supplementary Fig. S5), with their iron atoms being about 1.0 Å and 0.90 Å apart, respectively in the unliganded and in the carbomonoxy form. In addition, the Fe-Fe distance between hexacoordinated F106A
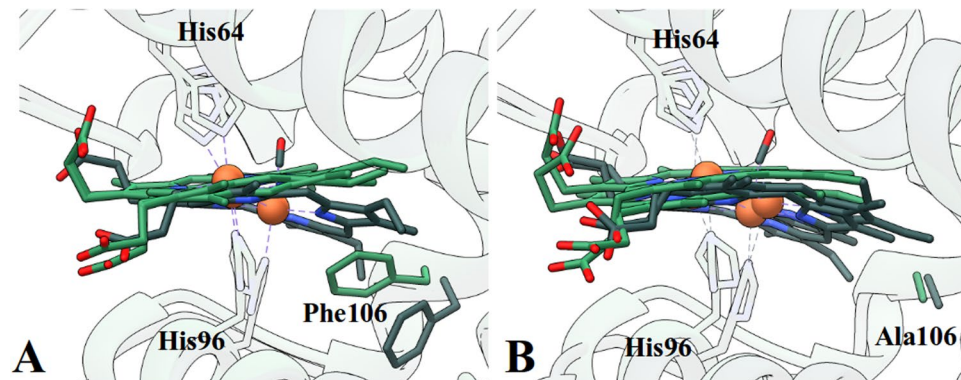
Figure 3. Comparison of wild type and F106A neuroglobin structures. (**A**) Unbound (green) and carbo-monoxy (grey) WT Ngb structures[14,15]. (**B**) Unbound (green) and carbo-monoxy (grey) F106A Ngb structures.
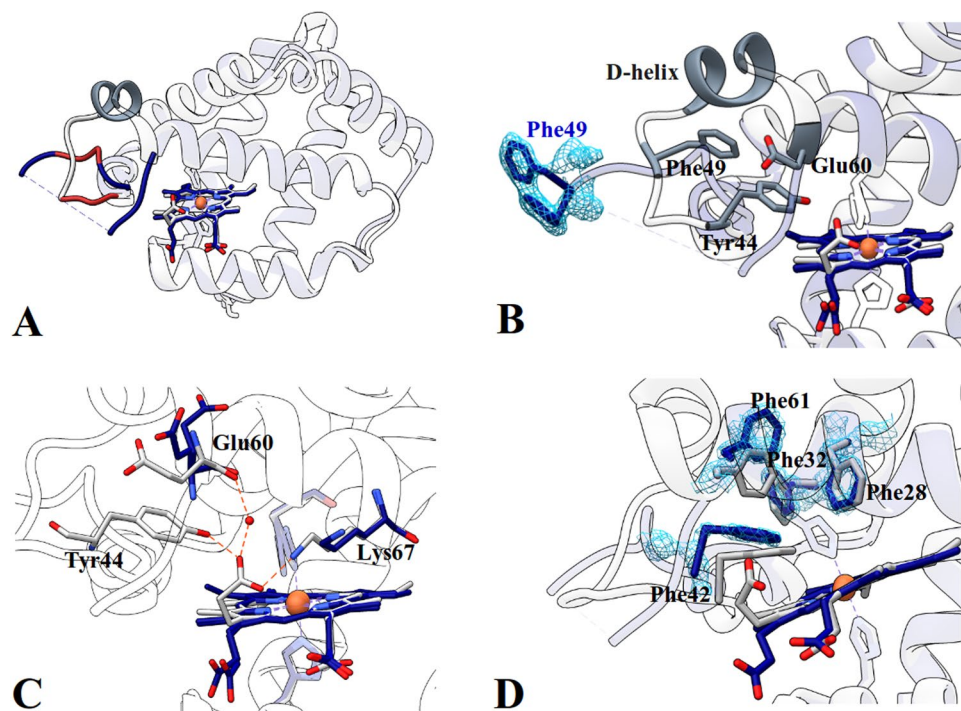

Figure 4. Superposition of Ngb Gly-loop (blue) and Ngb WT (grey) structures. (**A**) Comparison between the overal fold of the WT and the Gly-loop mutant. The segment missing from electron density maps (Phe49-Ser57) is indicated by a dashed line, the mutated sequence is in red, and the folded D-helix is indicated in dark grey. (**B**) In WT, the D-helix leans on Phe49, for the Gly-loop mutant, the 2Fo-Fc map is contoured at 1σ for Phe49. (**C**) Network of interactions between Tyr44, Glu60, Lys67, one of the heme propionates and a water molecule in WT Ngb. In the Gly-loop mutant, this network is abolished due to the absence of Tyr44. (**D**) The canonical hydrophobic patch in neuroglobin involved in heme support and maintenance of the globin fold is perturbated.

and F106A-CO shows, upon CO binding, a displacement of the heme A conformer of about 2.3 Å and of 2.9 Å for the heme B conformer, larger than the one observed for WT Ngb (2.0 Å[15]).

The enhanced displacement of B conformer is due to the absence of Phe106 ring onto which the heme leans in the WT protein. Consistently, Phe(CD1)42, a conserved residue involved in the heme packing through π-π stacking interactions, adopts a double conformation in a 70:30 proportion (lateral displacement of 1.0 Å at the level of Phe42 Cζ, not shown) in the structure of unliganded F106A.

In the F106A-CO structure, the absence of Phe in position 106, associated with the two heme insertion modes, induces residues such as Trp148 and the whole FG loop to explore multiple conformations upon heme sliding (Supplementary Fig. S6A–C). This structural feature could account for the biphasic behaviour detected in CO binding. The heme double conformation has already been proposed to be responsible for the slightly biphasic ligand binding in human Ngb, absent in murine WT Ngb, and which is greatly enhanced in murine F106A Ngb, due to the enhanced difference of A and B heme conformer positions[26,27].
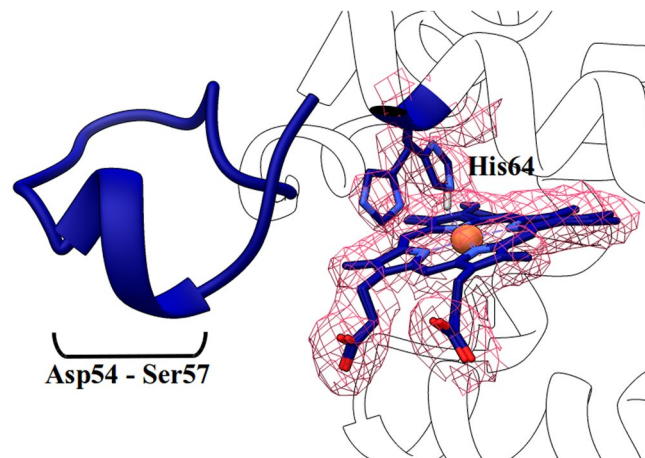
Figure 5. Evidence for His64 swing in Gly-loop structure bound to CO. The electron density indicates the presence of a fraction of swung out His64 (40%, the 2Fo-Fc map is contoured at 1σ). We could reconstruct the CD loop and we observed a re-shaping of the D-helix (from Asp54 to Ser57).

Finally, the absence of the aromatic side chain of Phe106 accounts for faster CO binding in both heme conformers, considering that in the WT, the phenylalanine ring has to flip to allow the heme to slide inside the cavity[15,20] (Fig. 3A), indeed, in the F106A mutant, the alanine side chain stays put upon heme repositioning (Fig. 3B). In mutant F106W[20], where steric hindrance is increased, CO binding velocity decreases of about five times and affinity is lowered by an order of magnitude (Supplementary Table S2).

The evaluation of the fast and slow binding phases in CO mixing experiment allows us to attribute to the B conformer the slow binding fraction and to the A conformer the fast binding one. An additional feature contributing to the increased CO binding velocities might reside in enhanced acceleration of the spontaneous rupture of the His64-Fe bond due to variations in its geometry reported in Supplementary Table S4.

The disruption of the heme-CD loop interaction network is responsible for increased CO binding velocities in the Gly-loop mutant.

The global Ngb 3D architecture is conserved in the CD corner "Gly-loop" mutant as shown in Fig. 4A. The double conformation of the heme in 70:30 proportion, along with corresponding histidine double conformations, is present and His64(Nε)-Fe-His96(Nε) angles are unchanged with respect to WT (Supplementary Table S4). Besides some alternate conformations of the N- and C-termini, of the AB and EF loops, due to crystal packing changes, only the CD loop was significantly altered. The lack of well defined electronic density did not allow to reconstruct the last three residues (aa 50–52) of the CD loop Fig. 4A, grey dashed line), which was completely reshaped. Moreover the D-helix was disordered (Fig. 4A, red) and no electronic density was observed up to the beginning of the E-helix at Pro59 (Fig. 4A,B).

Figure 4C shows the alteration of intra-molecular interactions in the Gly-loop Ngb. In the WT protein, the oxygen of one of the propionates interacts with Lys67 (Fig. 4B), this propionate and Tyr44 constitute a barrier to the swinging out of His64 (Supplementary Fig. S7) which in MD simulations, can take place only upon the upward movement of the CD loop and propionate flipping[17]. Tyr44 and Lys67 firmly interact at a 2.5 Å distance with a water molecule bridging the propionate to the main chain oxygen of Glu60. Owing to the mutation of Tyr44 into a glycine, the propionate flips toward the proximal side of the heme, and the amino group of Lys67 is rotated by 180°. Moreover in WT Ngb, the bond between Tyr44 and the heme propionate favors a conformation of the CD loop in which the D-helix leans on the Phe49 aromatic ring (Fig. 4B). The absence of Tyr44, that is not pinning any more the loop to the heme, allows the displacement of Phe49 which might explain the D-helix destabilization. In addition to pinning the CD loop to the heme propionate, Tyr44 appears to be part of the aromatic core around which the CD-loop-D helix segment is wound, with the significant stacking of the aliphatic portion of Glu60 onto its phenyl moiety. The CD-loop-D-helix unit seems therefore to be an intrinsically dynamic segment, whereas the C- and E-helices appear to constitute a rigid scaffolding in Ngb. The mobilization of the CD-loop-D-helix unit could facilitate the His64 dissociation from the heme iron, unlocking its swing-out displacement, thus leading to the increased CO binding velocities observed by rapid mixing. Indeed, faster O_2 association was also observed for Tyr44Asp Ngb, which has a behaviour remarkably similar to the Gly-loop mutant, including enhanced ligand concentration dependence and ligand affinity (P_{50}(Y44D) = 1.3 torr vs P_{50}(WT) = 2.2 torr)[21].

Quite surprisingly, the hydrophobic cluster involved in the support of the heme, and including the canonical Phe28(B10), Phe32(B14), Phe42(CD1) and Phe61(E4) is loosened in the Gly-loop mutant. As shown, in Fig. 4D, Phe42 and Phe61 are displaced with respect to the WT Ngb phenylalanine cluster: Phe42 is located about 0.6 Å further away from the heme (Cε(Phe42)-C3 methyl(heme) distance), whereas Phe61 is laterally shifted of approximately 2.4 Å (Cζ(Phe61 wt)-Cζ(Phe61 mutant) distance). These variations are structurally significant, considering that the universally conserved Phe42 is the residue in direct contact with the heme[4,28], and could constitute an additional factor affecting the heme mobility, possibly contributing to increased binding velocity observed by rapid mixing.
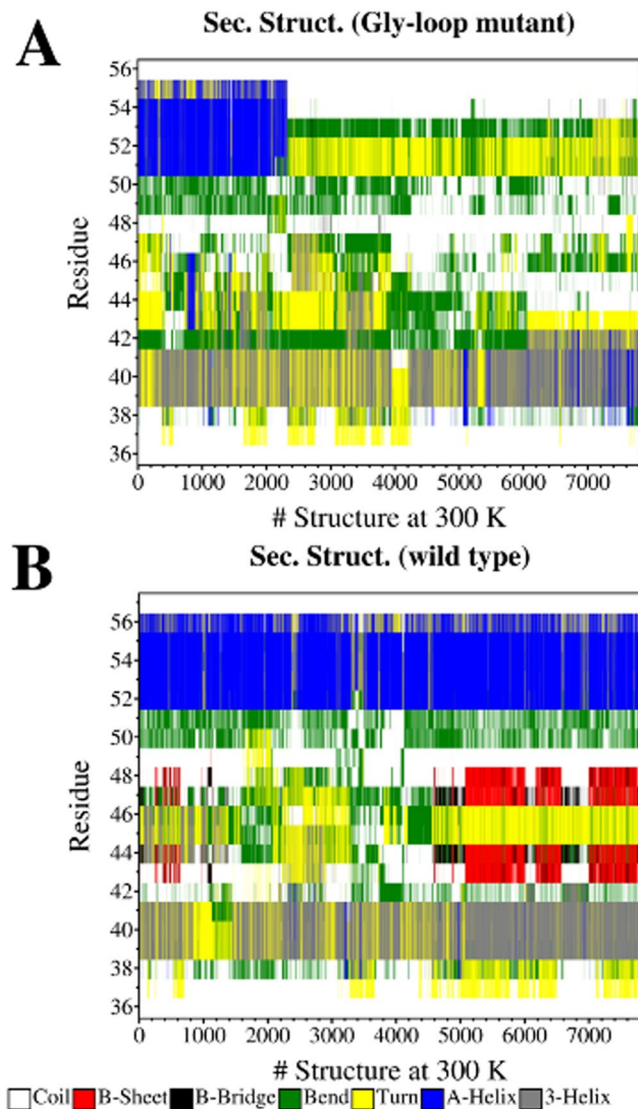
Figure 6. Secondary structure of CD corner at 300 K. The secondary structure was determined by DSSP (hydrogen bond estimation algorithm) on the ensemble of structures, sorted by time in ascending order, at 300 K, for Gly-loop Ngb (**A**) and WT Ngb (**B**).

Evidence of the distal histidine "gating out" movement and of the reshaping of the D-segment upon partial CO binding observed by X-ray crystallography. Crystals of the Gly-loop mutant tend to dissolve upon exposure to CO in the reduced Fe^{2+} state. However, by freezing crystals before complete disruption, we managed to determine a structure in which we observed about 40% CO ligation and a concomitant swung out histidine conformation (Fig. 5). Moreover electron density appeared for the 49–56 segment that could be successfully built. This indicates that, in the absence of the constraint imposed by the propionate-Tyr44 bond, a state with a swung out His64 can be trapped and structurally characterized, with the heme still in the unslid position. The transition of the CD-loop-D-helix segment is an indication that a remodeling of this region is associated to ligand binding, as also suggested by MD simulations, which is likely the cause for destabilization of the crystal lattice upon full ligation.

MD Simulations predict the CD loop-D-helix disordering and support the gating out movement of His64 promoting the partial refolding of D-helix. The complete 3D homology model of the Gly-loop mutant modeled with the same fold as WT Ngb (Supplementary Fig. S8) was used as a starting point for MD simulations in order to assess the rearrangement of the CD corner upon the triple glycine substitution and to compare the results with the WT MD. After equilibration using standard MD simulations, we performed simulated tempering MD (ST-MD) of the Gly-loop mutant and of the WT, initially in a range of 300–420 K.

In Fig. 6, we report the secondary structure of the CD corner at 300 K as a function of time, for the Gly-loop mutant and WT Ngb. The conformation of the C-helix fluctuates during the simulation between a 3-helix and a turn, both in WT and in the mutant, showing that, despite the peptide strand remains in a twisted conformation,

the optimal geometry of the hydrogen bond involving the main chain is continuously lost and immediately regained. On the other hand, while the D-helix remains folded during the whole simulation for the WT, in the mutant, the helix unfolds within the first 110 ns and remains unfolded. The reduced structural stability of the D-helix in the Gly-loop mutant is further confirmed by a 60 ns ST-MD simulation in a temperature range of 300–460 K. In that condition, a helix-coil transition in the D-helix was observed after 40 ns for the Gly-loop mutant, then steadily maintained at 300 K, without regaining the helical conformation. Due to high temperatures, a helix-coil transition was observed also for WT Ngb, but the D-helix folded again at the end of the simulation. In Supplementary Fig. S9A, we have reported the most representative configurations (80%), obtained by clustering the structures of the Gly-loop mutant at 300 K. The severe destabilization of the D-helix in the Gly-loop mutant was confirmed by the free energy landscapes that describe the helix unfolding (Supplementary Fig. S10) where the unfolded configurations are more stable of 5–8 kJ/mol than in WT Ngb.

To test the hypothesis that the prevailing effect of the glycine substitutions in the CD corner is to favor the gating out of His64 upon external ligation to the heme, resulting in the observed increased reactivity with CO, we studied the His64 swinging motion, both in the Gly-loop mutant and in WT, choosing the Fe-N distance between the heme and His64 as reaction coordinate. In Supplementary Fig. S11A, we report some snapshots describing the displacement of His64 and the opening of the His-gate. It is worth noting that in Gly-loop Ngb, a relevant structural rearrangement of the CD corner is shown throughout the swing in/out of His64, where the D-helix recovers its folding. Instead, such a rearrangement is not observed in WT, and the CD corner remains folded throughout the His64 displacement.

The free energy landscapes of the His64 displacement (Supplementary Fig. S11B) show that its removal from the sixth coordination position is thermodynamically and kinetically favored in Gly-loop Ngb. This effect is rather moderate up to the distance of 0.4 nm, where the removal of His64 is at least 2 kJ/mol more favorable, but at higher Fe-N distances, the free energy barrier decreases from 16.5 kJ/mol in WT to 6 kJ/mol in the mutant, and the swung-out conformations (corresponding to the open His-gate) are between 5 and 18 kJ/mol more stable in Gly-loop Ngb than in WT Ngb. The difference between the two profiles can be explained considering not only an increased mobility of the CD corner in the mutant but also the existence of a coupling between the gating out movement of the distal His64 and the structural rearrangement of the CD corner, that mainly involves the refolding of the D-helix. Thus, the swung out conformations are promoted by the return of the D-helix to a folded state, and therefore to more enthalpically favored conformations.

Concluding Remarks

By analyzing the kinetics, spectroscopic and structural features of three Ngb mutants, we have highlighted that the control of external ligand binding, mainly dependent on the spontaneous dissociation of the distal His64 is exerted (i) by the swinging out of distal His64 that depends on Tyr44 and on the dynamics of the CD-loop-D-helix segment and (ii) by the barriers to the heme sliding within its crevice consisting of residues on the proximal side, mostly Phe106[20].

The effect of the double heme insertion observed in Ngb is enhanced by the increased size of the large Ngb internal cavity created by the F106A mutation, which results in a set of alternate conformations of proximal side residues that account for the enhanced biphasic behaviour in ligand binding of mutants carrying this substitution.

We propose a sequence of events for ligand binding to neuroglobin, in which after spontaneous rupture of the 6th coordination bond of His64 to the heme iron, exogenous diatomic ligands such as CO can be coordinated by the heme only upon swinging out of His64, having isolated and determined the structure of this intermediate in the Gly-loop mutant where the Tyr44-propionate barrier is absent. Heme sliding within the heme crevice follows this intermediate step and leads to partial return of His64 in the distal cavity as indicated by resonance Raman spectroscopy. Removal of the barrier to His64 early swinging out increases the velocity of binding by one order of magnitude, reaching a much faster rate limiting step in the Gly-loop mutant with respect to WT Ngb. Consistently with our hypothesis, this indicates that, in the Gly-loop mutant (where the barrier to His64 gating out is removed and Phe106 still needs to rearrange to allow heme sliding), the fraction of the A_0 open conformation is predominant. Indeed, in Gly-loop-F106A, where the main barrier to sliding is removed, the A_0/A_3 ratio (open/closed) is restored to a value similar to the one observed for WT Ngb.

In the double mutant F106A/Gly-loop we found a more than additive effect of the loosening of both distal and proximal control as shown by the fifty times increase in CO binding rate, which might be due to a global destabilization of the hexacoordinated state (or stabilization of the ligand-bound one) over-and-above the local effect on the 6th coordination position. Raman spectroscopy and MD simulation support our interpretation of the structural and kinetics data.

Methods

Cloning, expression and purification.   We utilized as WT Ngb the C55S/C120S murine gene[14]. The CD loop "Gly-loop" mutation was introduced by means of the HiFi expand kit in combination with the Gibson Assembly kit (BioLabs). The F106A mutation was introduced using the QuickChange Lightning SDM kit (Stratagene). WT Ngb and mutants were expressed and purified as previously described[29]. Spectroscopy and rapid mixing experiments were performed in 0.1 M HEPES pH 7.4 at 25 °C.

Resonance Raman spectroscopy.   Final protein concentration was in the range of 40–150 μM. Ferrous samples were prepared by adding Na-dithionite (20 mg/mL). The CO derivative was obtained by CO equilibration (^{12}CO or ^{13}CO (Rivoira, Milan, Italy)), followed by reduction of the heme with Na-dithionite. RR spectra were obtained by excitation with the 413.1 nm line of Kr$^+$ laser (Innova 300C, Coherent, Santa Clara, CA, USA) and the 441.6 nm line of a He–Cd laser (Kimmon IK4121R-G). The RR spectra were calibrated with indene, carbon tetrachloride and acetonitrile as standards to an accuracy of 1 cm^{-1} for intense isolated bands. All RR measurements

were repeated several times to ensure reproducibility. To improve the signal-to-noise ratio, a number of spectra were accumulated and summed only if no spectral differences were noted. All spectra were baseline-corrected. Absorption spectra were measured both prior to and after RR measurements to ensure that no degradation had occurred (Supplementary Materials and Methods).

Rapid mixing experiments.   Stopped-flow experiments were performed with a 1 cm light path Applied Photophysics apparatus. Ngb solutions and buffers were degassed by N_2 equilibration. Heme iron reduction from Fe^{3+} to Fe^{2+} (ferric to ferrous) was achieved by anaerobically adding 200 μM Na-dithionite. Solutions of CO concentrated from 1 mM to 15 μM were prepared by consecutive dilutions of the 1 mM CO stock solution with N_2-equilibrated buffer. The binding reaction between the 10 μM ferrous Ngbs (WT and Gly-loop/F106A) or 5 μM ferrous Ngbs (F106A and Gly-loop) and CO was followed at 426 nm. Rate constants and amplitudes were determined using the Qtiplot software.

Crystallization, data collection and analysis.   Crystals were obtained using the vapor diffusion method in hanging drops. Data collection was carried out at ESRF (France) and Elettra (Italy), and structures were solved by molecular replacement. (Supplementary Table S1).

Molecular Dynamics (MD) simulations.   The initial coordinates for WT Ngb were from PDB entry 1Q1F[14], for the Gly-loop variant the missing residues Phe49-Ser57 were modeled by homologs. Proteins were solvated with explicit TIP3P waters. All MD simulations were performed with the GROMACS package[30] version 2016.4 using CHARMM36m force field[31]. The detailed protocol for MD simulations is given in Supplementary Materials and Methods. Simulated tempering (ST)-MD were performed for both systems for 300 ns. The free energies profiles were obtained using umbrella sampling and weighted histogram analysis method (WHAM)[32]. The initial configurations, one for each umbrella window, were generated by steered molecular dynamics (SMD) simulations[33] coupled with simulated tempering[34]. The configuration for each window was obtained collecting and grouping the configurations at 300 K in sub-ensemble according their projection along the reaction coordinate, and clustering them using the GROMOS method. Starting from the structures reported in Supplementary Fig. S10, an equilibration run of 100 ns was performed. Then the protein was switched to the pentacoordinated state. After an energy minimization, and equilibration of other 100 ns, the Fe-N distance between the heme and the distal histidine was constrained at 2.2 Å during a ST-MD of 10 ns. Then a representative structure at 300 K was chosen and the His64 was pulled away from the iron atom. From that simulation, the configurations for the following umbrella sampling were extracted.

References

1. Burmester, T. & Hankeln, T. Function and evolution of vertebrate globins. *Acta Physiologica* **211**, 501–514, https://doi.org/10.1111/apha.12312 (2014).
2. Holm, L. & Sander, C. Structural alignment of globins, phycocyanins and colicin A. *FEBS Letters* **315**, 301–306, https://doi.org/10.1016/0014-5793(93)81183-Z (1993).
3. Bashford, D., Chothia, C. & Lesk, A. M. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **196**, 199–216, https://doi.org/10.1016/0022-2836(87)90521-3 (1987).
4. Ptitsyn, O. B. & Ting, K. L. H. Non-functional conserved residues in globins and their possible role as a folding nucleus. *J. Mol. Biol.* **291**, 671–682, https://doi.org/10.1006/jmbi.1999.2920 (1999).
5. Baez, E. *et al.* Protection by neuroglobin expression in brain pathologies. *Front. Neurol.* **7**, 146, https://doi.org/10.3389/fneur.2016.00146 (2016).
6. Dewilde, S. *et al.* Biochemical Characterization and Ligand Binding Properties of Neuroglobin, a Novel Member of the Globin Family. *J. Biol. Chem.* **276**, 38949–38955, https://doi.org/10.1074/jbc.M106438200 (2001).
7. Kiger, L. *et al.* Neuroglobin ligand binding kinetics. *IUBMB life* **56**, 709–19, https://doi.org/10.1080/15216540500037711 (2004).
8. Kriegl, J. M. *et al.* Ligand binding and protein dynamics in neuroglobin. *Proc. Natl. Acad. Sci. USA* **99**, 7992–7997, https://doi.org/10.1073/pnas.082244399 (2002).
9. Smagghe, B. J., Sarath, G., Ross, E., Hilbert, J. L. & Hargrove, M. S. Slow ligand binding kinetics dominate ferrous hexacoordinate hemoglobin reactivities and reveal differences between plants and other species. *Biochem.* **45**, 561–570, https://doi.org/10.1021/bi051902l (2006).
10. Fago, A. *et al.* Allosteric regulation and temperature dependence of oxygen binding in human neuroglobin and cytoglobin: Molecular mechanisms and physiological significance. *J. Biol. Chem.* **279**, 44417–44426, https://doi.org/10.1074/jbc.M407126200 (2004).
11. Hamdane, D. *et al.* The Redox State of the Cell Regulates the Ligand Binding Affinity of Human Neuroglobin and Cytoglobin. *J. Biol. Chem.* **278**, 51713–51721, https://doi.org/10.1074/jbc.M309396200 (2003).
12. Nienhaus, K. & Nienhaus, G. U. A spectroscopic study of structural heterogeneity and carbon monoxide binding in neuroglobin. *J. Biol. Phys.* **31**, 417–432, https://doi.org/10.1007/s10867-005-0173-0 (2005).
13. Trent, J. T. & Hargrove, M. S. A ubiquitously expressed human hexacoordinate hemoglobin. *J. Biol. Chem.* **277**, 19538–19545, https://doi.org/10.1074/jbc.M201934200 (2002).
14. Vallone, B., Nienhaus, K., Brunori, M. & Nienhaus, G. U. The structure of murine neuroglobin: Novel pathways for ligand migration and binding. *Proteins: Struct. Funct. Genet.* **56**, 85–92, https://doi.org/10.1002/prot.20113 (2004).
15. Vallone, B., Nienhaus, K., Matthes, A., Brunori, M. & Nienhaus, G. U. The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc. Natl. Acad. Sci. USA* **101**, 17351–6, https://doi.org/10.1073/pnas.0407633101 (2004).
16. Moschetti, T., Mueller, U., Schulze, J., Brunori, M. & Vallone, B. The structure of neuroglobin at high Xe and Kr pressure reveals partial conservation of globin internal cavities. *Biophys. J.* **97**, 1700–1708, https://doi.org/10.1016/j.bpj.2009.05.059 (2009).
17. Anselmi, M., Brunori, M., Vallone, B. & Di Nola, A. Molecular dynamics simulation of deoxy and carboxy murine neuroglobin in water. *Biophys. J.* **93**, 434–41, https://doi.org/10.1529/biophysj.106.099648 (2007).
18. Anselmi, M., Brunori, M., Vallone, B. & Di Nola, A. Molecular dynamics simulation of the neuroglobin crystal: comparison with the simulation in solution. *Biophys. J.* **95**, 4157–4162, https://doi.org/10.1529/biophysj.108.135855 (2008).
19. Nienhaust, K., Kriegl, J. M. & Nienhaust, G. U. Structural dynamics in the active site of murine neuroglobin and its effects on ligand binding. *J. Biol. Chem.* **279**, 22944–22952, https://doi.org/10.1074/jbc.M401561200 (2004).

20. Avella, G. *et al.* Engineering the internal cavity of neuroglobin demonstrates the role of the haem-sliding mechanism. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **70**, 1640–1648, https://doi.org/10.1107/S1399004714007032 (2014).
21. Giuffrè, A., Moschetti, T., Vallone, B. & Brunori, M. Neuroglobin: Enzymatic reduction and oxygen affinity. *Biochem. Biophys. Res. Commun.* **367**, 893–898, https://doi.org/10.1016/j.bbrc.2008.01.021 (2008).
22. Boron, I. *et al.* Engineered chimeras reveal the structural basis of hexacoordination in globins: A case study of neuroglobin and myoglobin. *Biochimica et Biophys. Acta - Gen. Subj.* **1850**, 169–177, https://doi.org/10.1016/j.bbagen.2014.10.006 (2015).
23. Giordano, D. *et al.* Biophysical Characterisation of Neuroglobin of the Icefish, a Natural Knockout for Hemoglobin and Myoglobin. Comparison with Human Neuroglobin. *PLoS One* **7**, https://doi.org/10.1371/journal.pone.0044508 (2012).
24. Ishikawa, H. *et al.* Neuroglobin dynamics observed with ultrafast 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. United States Am.* **104**, 16116–16121, https://doi.org/10.1073/pnas.0707718104 (2007).
25. Sawai, H. *et al.* Structural characterization of the proximal and distal histidine environment of cytoglobin and neuroglobin. *Biochem.* **44**, 13257–13265, https://doi.org/10.1021/bi050997o (2005).
26. Bocahut, A. *et al.* Heme orientation modulates histidine dissociation and ligand binding kinetics in the hexacoordinated human neuroglobin. *J. Biol. Inorg. Chem.* **18**, 111–122, https://doi.org/10.1007/s00775-012-0956-2 (2013).
27. Fago, A., Mathews, A. J., Dewilde, S., Moens, L. & Brittain, T. The reactions of neuroglobin with CO: Evidence for two forms of the ferrous protein. *J. Inorg. Biochem* **100**, 1339–1343, https://doi.org/10.1016/j.jinorgbio.2006.03.009 (2006).
28. Ota, M., Isogai, Y. & Nishikawa, K. Structural requirement of highly-conserved residues in globins. *FEBS Lett.* **415**, 129–133, https://doi.org/10.1016/S0014-5793(97)01110-1 (1997).
29. Arcovito, A. *et al.* An X-ray diffraction and X-ray absorption spectroscopy joint study of neuroglobin. *Arch. Biochem. Biophys.* **475**, 7–13, https://doi.org/10.1016/j.abb.2008.03.026 (2008).
30. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free, https://doi.org/10.1002/jcc.20291 (2005).
31. Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73, https://doi.org/10.1038/nmeth.4067 (2016).
32. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021, https://doi.org/10.1002/jcc.540130812 (1992).
33. Izrailev, S. *et al.* Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*, 39–65, https://doi.org/10.1007/978-3-642-58360-5{_}2 (1999).
34. Marinari, E. & Parisi, G. Simulated tempering: A New Monte Carlo Scheme, https://doi.org/10.1209/0295-5075/19/6/002 (1992).
35. Couture, M., Burmester, T., Hankeln, T. & Rousseau, D. L. The heme environment of mouse neuroglobin. Evidence for the presence of two conformations of the heme pocket. *J. Biol. Chem.* **276**, 36377–36382, https://doi.org/10.1074/jbc.M103907200 (2001).
36. Howes, B. D., Helbo, S., Fago, A. & Smulevich, G. Insights into the anomalous heme pocket of rainbow trout myoglobin. *J. Inorg. Biochem.* **109**, 1–8, https://doi.org/10.1016/j.jinorgbio.2012.01.007 (2012).
37. Morikis, D., Champion, P. M., Springer, B. A. & Sligar, S. G. Resonance Raman Investigations of Site-Directed Mutants of Myoglobin: Effects of Distal Histidine Replacement. *Bioch.* **28**, 4791–4800, https://doi.org/10.1021/bi00437a041 (1989).

Acknowledgements

Author Contributions

B.V. conceived the study. B.V., C.E., M.A. and G.S. planned the experiments, interpreted data and wrote the manuscript. C.E., L.C.M. and I.F. carried out the kinetics characterization. C.E., C.S., I.F. and G.C. carried out the structural determination. C.E., A.S. and G.P. cloned and purified proteins. G.S. and L.M. carried out resonance Raman spectroscopy and its interpretation. M.A. carried out molecular dynamics and its interpretation.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41780-3.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.