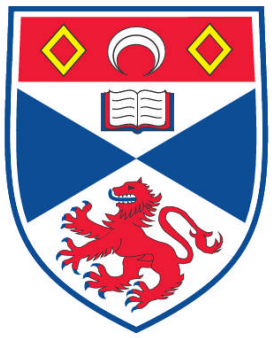
# PYRH AND PRNB CRYSTAL STRUCTURES

**Walter De Laurentis**

**A Thesis Submitted for the Degree of PhD**
**at the**
**University of St. Andrews**



**2006**

# “PyrH and PrnB Crystal Structures”

A thesis submitted for the degree of

Doctor of Philosophy

School of Chemistry and
Centre for Biomolecular Sciences
University of St Andrews

Walter De Laurentis
December 2006

**Supervisor – Prof. J H Naismith**

# ABSTRACT

Determination of the three-dimensional structure of enzymes at atomic resolution is a key prerequisite for elucidation of molecular mechanisms of catalysis and catalysis mechanism prediction. X-ray protein crystallography is the most widely used method today for determining protein structures.

In this thesis we describe the expression, purification, crystallization and structure solution of two new enzymes: PyrH and PrnB.

PyrH is a member of the new emerging family of FADH dependent tryptophan halogenases. It catalyzes the regioselective halogenation of tryptophan at the C-5 position of the indole ring. Elucidation of its structure (Chapter 2) and comparison with PrnA, a regioselective $7^{th}$ tryptophan halogenase whose structure has already been solved confirmed the proposed mechanism of action for this class of enzymes.

PrnB is the only enzyme known to perform exquisite and peculiar ring rearrangement chemistry: it converts 7-Cl-tryptophan and tryptophan into respectively monodechloroaminopyrrolnitrin and aminophenylpyrrole. We developed a method for expression and purification of milligrams of pure and homogeneous recombinant PrnB (Chapter 3). We identified suitable crystallization conditions and determined PrnB structure (Chapter 4). Analysis of the PrnB structure helped us to propose a reaction mechanism for this unique enzyme.

# DECLARATION

I, Walter De Laurentis, hereby certify that this dissertation, which is approximately 26000 words in length, has been written by me, that is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: Signature of candidate:

…../…../….. ……

I was admitted as a research student in January, 2004 and as a candidate for the degree of Doctor of Philosophy in August, 2006; the higher study for which this is a record was carried out in the Biomolecular Sciences (BMS), School of Chemistry department at the University of St. Andrews between 2004 and 2006.

Date: Signature of candidate:

…../…../….. ……

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit the dissertation in application for that degree.

Date: Signature of supervisor:

…../…../….. ……

(Prof. J. H. Naismith)

# COPYRIGHT DECLARATION

Unrestricted access

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date: Signature of candidate:

…../…../….. …….

# LIST OF ABBREVIATIONS

A: Absorbance

Å: Angstrom

aa: Amino Acid

APP: Aminophenylpyrrole

APRN: Aminopyrrolnitrin

bp: Base Pair

BSA: Bovine Serum Albumin

CV: Column Volume

conc: Concentration

Da : Dalton

dNTP: Deoxynucleotide Triphosphate

DTT: Dithiothreitol

EDTA: Ethylene Diamine Tetra-Acetic Acid

EK: Enterokinase

FAD/FADH: Flavin-Adenine Dinucleotide

FMN/FMNH: Flavin-Mononucleotide

K: Kilo

Kbp: Kilo Base Pair

KDa: Kilo Dalton

MAD: Multiwavelenght Anomalous Diffraction

MBP: Maltose Binding Protein

MDA: monodechloroaminopyrrolnitrin

MES: 2-Morpholinoethanesulfonic acid

$M_w$: Molecular Weight

NAD/NADH: Nicotinamide-Adenine-Dinucleotide

NADP/NADPH: Nicotinamide-Adenine-Dinucleotide Phosphate

O/N: Over Night

ORF: Open Reading Frame

PAGE: Polyacrylamide Gel Electrophoresis

PCR: Polymerase Chain Reaction

PEG: Polyethylene Glycol

PMM: *Pseudomonas* Minimal Medium

PRN: Pyrrolnitrin

SDS: Sodium Dodecyl Sulphate

TAE: Tris Acetate EDTA Buffer

TEV: Tobacco Etch Virus

TRP: Tryptophan

Amino Acid Single Letter Code

A: alanine

C: cysteine

D: aspartic acid

E: glutamic acid

F: phenylalanine

G: glycine

H: histidine

I: isoleucine

K: lysine

L: leucine

M: methionine

N: asparagine

P: proline

Q: glutamine

R: arginine

S: serine

T: threonine

V: valine

W: tryptophan

Y: tyrosine

**Chapter 1**
**Introduction** **6**
1.1 Halogenated Natural Compounds 7
1.2 Biological Halogenation 9
1.3 Haloperoxidase 11
1.4 Perhydrolases 16
1.5 Flavin Dependent Halogenase 18
1.6 PrnA a FADH dependent tryptophan 7-halogenase 22
1.7 Non-heme FeII $\alpha$-ketoglutarate and $O_2$-dependent halogenases 29
1.8 Nucleophilic chlorination: Methyl-Tranferase 32
1.9 Nucleophilic chlorination: Fluorinase 33
1.10 Pyrrolnitrin 35

**Chapter 2**
**Expression, Purification and Structure of PyrH** **43**
2.1 Summary 44
2.2 FADH dependent tryptophan halogenases 44
2.3 PyrH a FADH dependent tryptophan 5 halogenase 46
2.4 Aim of our study 47
2.5 Materials and Methods 48
2.6 Results 59
2.7 Discussion 71
2.8 Future Work 85

**Chapter 3**
**PrnB Cloning, Expression and Purification** **86**
3.1 Summary 87
3.2 PrnB sequence analysis 87
3.3 Aim of our study 88
3.4 Materials and Methods 89
3.5 Results 97
3.6 Discussion 113

**Chapter 4**
**PrnB Structure Determination and Analysis** **114**
4.1 Summary 115
4.2 Introduction 115
4.3 Materials and Method/Results 116
4.4 Discussion 124
4.5 Future Work 146
**REFERENCES** **147**

**ACKNOWLEDGMENTS** **157**

# Chapter 1

# Introduction

## 1.1 Halogenated Natural Compounds

With seven outer electrons in their highest occupied energy level the halogens (or “salt-formers”) are among the most reactive of the elements and their chemistry is dominated by their high electronegativity. In their pure forms, the Group VII elements exist as diatomic molecules. With metals they form ionic compounds, with non-metal they form covalent compounds. Of the five elements in the group fluorine, chlorine, bromine and iodine are found in biological molecules. The least aboundant and the largest halogen, astatine, is radioactively unstable and not found in nature. The first report of a halogen-containing natural product (halometabolite) was that of the iodinated amino acid diiodotyrosine from the coral *Gorgonia cavolii* in 1896 [1]. Thirty years ago only 200 organohalogen compounds were documented, today more than 4000 organohalogen compounds are described in literature, approximately 2200 organochlorine, 1950 organobromine, 95 organoiodine and 100 organofluorine [2]. Halogenated compounds are formed during natural abiogenic processes, such as volcanoes, forest fires, and other geothermal processes. Natural halogenated compounds are defined as those produced by living organisms. Oceans are the largest source of organohalogens, biosynthesized by seaweeds, sponges, corals, tunicates, bacteria. Terrestrial plants, fungi, lichen, bacteria, insects, some higher animals, and even humans [3] produce halometabolites. The number of organohalogens metabolites isolated from living organisms is steadily increasing [4], mainly as consequence of the general revitalization of interest in natural products as a potential source of new medicinal drugs [3]. The ratio of discovery of new organohalogens mirrors the ratio of known ones with chlorinated compounds dominating. The best known halogenated natural compound is vancomycin, the last resort antibiotic against MRSA, which depends on chlorine for its full bactericidal activity [5; 6]. Chloramphenicol is a dichlorinated broad-spectrum antibacterial antibiotic produced by *Streptomyces venezuelae* and other actinomycetes [7]. Rebeccamycin is a potent DNA topoisomerase inhibitor, in which the removal of its chlorine atoms results in weaker antiproliferative action [8]. Cryptophycin A is a newly discovered antimitotic agents isolated from *cyanobacteria* with *in vivo* potency 100-1000 times more potent than microtubule binding drugs paclitaxel and vinblastine [9], where the presence of chlorine is needed to obtain optimal cytotoxicity [10]. (See Fig. 1.1)

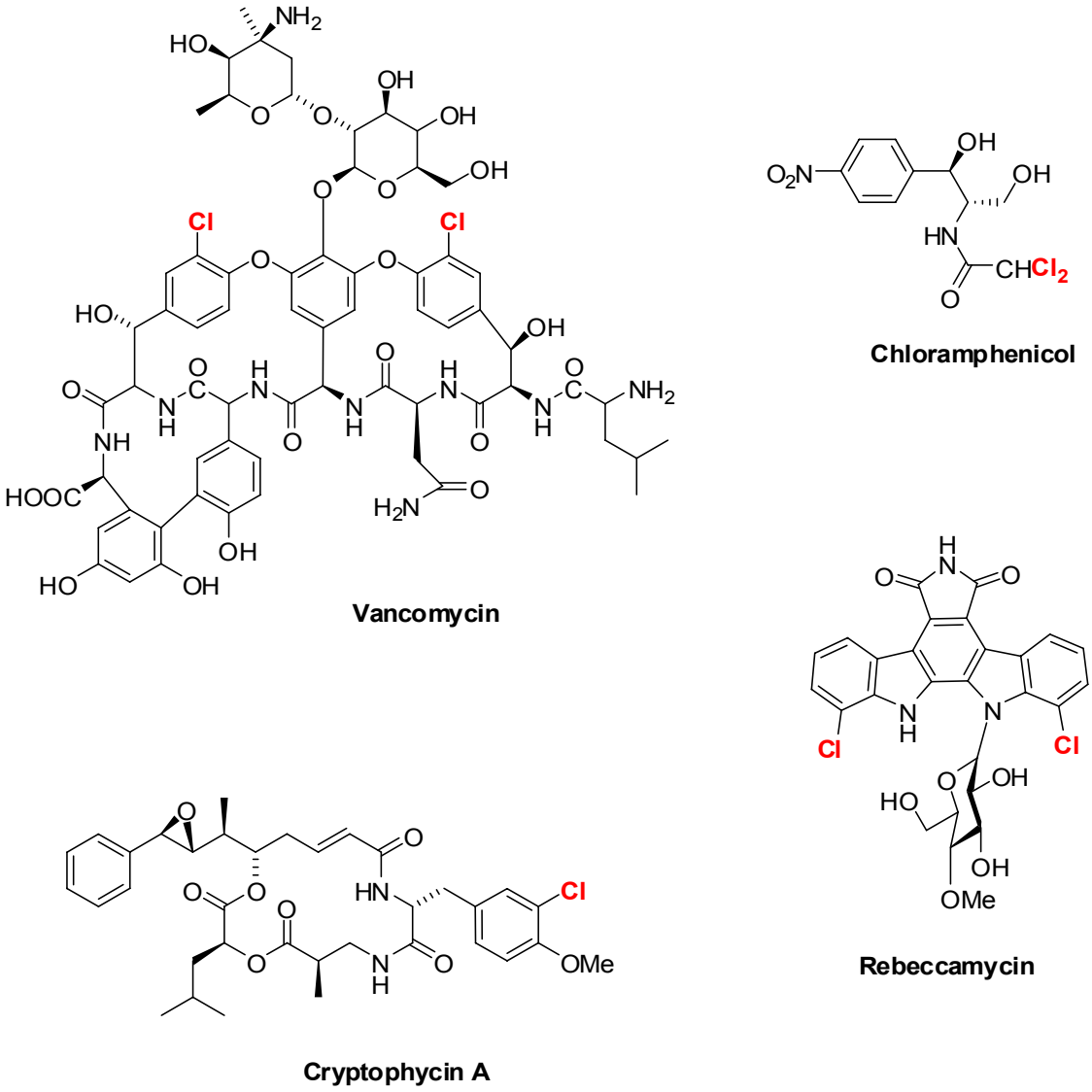



**Figure 1.1**

The full biological activity of these compounds depends on how halogenated molecules interact with cellular components. Halogen bonding described as oxygen–halogen interactions shorter than their respective Van der Waals Radii ($R_{VdW}$) has been known in organic chemistry since the 1950s. A halogen bond in biomolecules can be defined as a short C—X···O—Y interaction (C—X is a carbon-bonded chlorine, bromine, or iodine, and O—Y is a carbonyl, hydroxyl, charged carboxylate, or phosphate group), see Fig. 1.2.

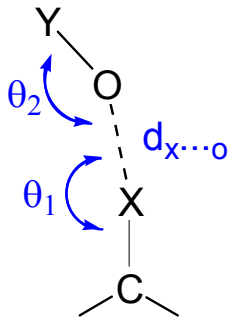



**Figure 1.2**
Schematic representation of halogen (X) interaction to various oxygen-containing functional groups (where O—Y can be a carbonyl, hydroxyl, or carboxylate when Y is a carbon; a phosphate when Y is a phosphorus; or a sulfate when Y is a sulfur). The geometry of the interaction is defined by the normalized $R_{X\cdots O}$ distance [$R_{X\cdots O} = d_{X\cdots O}/R_{vdW(X\cdots O)}$], the $\theta_1$ angle of the oxygen relative to the C—*X* bond, and the $\theta_2$ angle of the halogen relative to the O—Y bond.

A recent survey of the Protein Data Bank (PDB) database (Release July 2004) by Auffinger *et al*. [11] identified 66 different protein and six different nucleic acid structures (with resolution between 0.66 and 3.0 Å; average, ~2.1 Å) presenting halogen–oxygen distances ($R_{X\cdots O}$) shorter than their respective van der Waals Radii ($R_{vdW}$). These short halogen–oxygen (C—X···O) interactions in proteins and nucleic acids confirms that the halogen bond geometries in biological systems conform generally to those seen in small molecules. The interaction is defined by distances that are as short as 80% of $R_{vdW}$ of the interacting atoms as short as 80% of $R_{vdW}$ and are directional relative to the C—X bond, despite non perfect linear angles ($\theta_1 \neq 180°$). A characteristic perpendicular approach of the halogen toward peptide bond oxygen is attributed to the involvement of peptide bonds $\pi$ electrons. The high number of $\pi$-system donors identified in the survey shows perpendicular interactions to the peptide backbone to be very important in biomolecular halogen interactions. Halogen bonding together with the ability of halogens to enhance membrane binding and permeation [12] is responsible for the growing interest in halogenated drug like compounds.

## 1.2 Biological Halogenation

Half of the drugs currently in clinical use are of natural product origin [13]. Chemical synthesis of complex natural products such as antibiotics is often too expensive or too difficult to be practical so that production of those metabolites relies usually on

fermentation. Modification of natural products can be achieved in several ways. The first and simplest is to chemically modify the natural product itself by simple functional-group transformations. This semisynthetic approach has the benefit of providing analogs rapidly, but it is limited in terms of variety. Another possibility is to use genetic engineering to reconstruct biosynthetic pathways, the so called combinatorial biosynthetic approach [14]. Introduction of halogen atom in drug like compounds can positively alter their activity or pharmacokinetic parameters. Amplification of chemical diversity by introduction of halogen atom in natural products using either chemoenzymatic approaches or combinatorial biosynthesis would be possible with a deeper understanding of the mechanisms that control regioselective biological halogenation. Only then it will be possible to rationalize the introduction of halogens into new and altered structures using nature's biosynthetic machinery [15].

Chloroperoxidase, discovered in 1959, was the first enzymes shown to catalyze incorporation of chlorine into natural products [16]. Since then chloroperoxidases have been recognized as the catalyst involved in biological halogenation reactions. The ease by which this enzyme activity could be detected, using a convenient spectrophotometric assay developed by the researchers involved in the initial discovery of this class of enzyme, led to the discovery of many haloperoxidases from a variety of sources without the need to identify their natural substrates [17; 18]. However, the broad substrate acceptance, lack of regiospecificity and the difficulties in determining Michaelis constants for organic substrates were inconsistent with a role for these enzymes in the biosynthesis of complex halogenated metabolites [4]. The discovery that the halogenating reagent formed during the chloroperoxidase-catalysed reaction was hypochlorous acid, which upon release from the active site, reacts spontaneously with suitable organic compounds in the surrounding medium without any selectivity definitely established that biological chlorination cannot be catalyzed by this class of enzyme [4]. In the late 1990s, 45 years after the detection of haloperoxidases, Dairi *et al.* [19] were the first to report the discovery of a gene that did not code for a chloroperoxidase, but was required for the halogenation step of chlortetracycline biosynthesis. Similarly the identification and functional analysis of the gene cluster for pyrrolnitrin biosynthesis in *Pseudomonas fluorescens* revealed two genes responsible for chlorination, neither of which coded for a chloroperoxidase [20; 21]. This new class of

halogenases is now known as FADH dependent halogenase and is found to catalyze regioselective halogenation of aromatic substrates [4]. Very recently Walsh and colleagues demonstrated the novel halogenating activity of a class of nonheme $Fe^{II}$, $\alpha$-ketoglutarate-dependent enzymes that accounts for the regiospecific incorporation of chlorine into saturated aliphatic substrates [22].

## 1.3 Haloperoxidases

Studies on the biosynthesis of the chlorine containing fungal metabolite caldariomycin (See Fig. 1.3) led to the discovery of the first halogenating enzyme from the mold *Caldariomyces fumago* in 1959. For the catalysis of halogenation reactions, this enzyme requires hydrogen peroxide and halide ions (only chloride, bromide and iodide [17]) and was thus named chloroperoxidase.

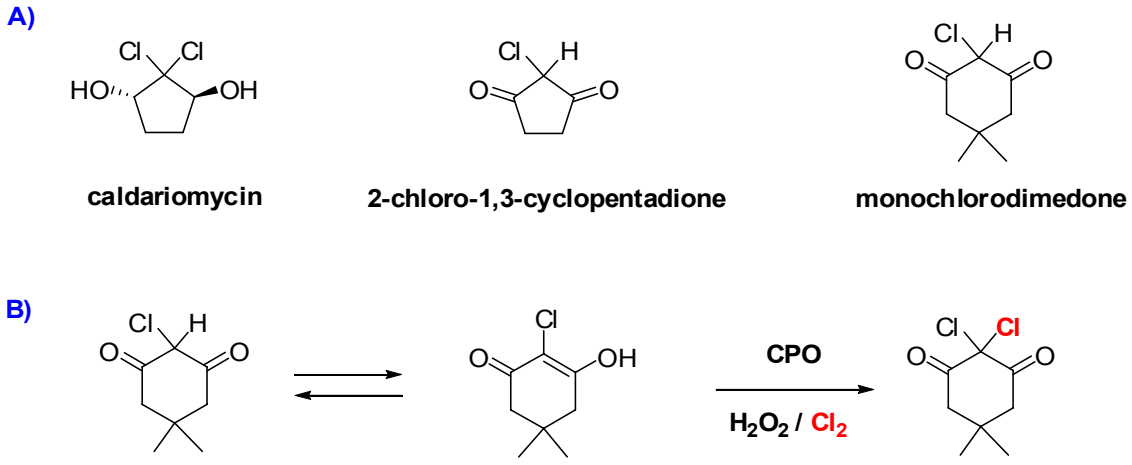



**Figure 1.3**
A): chemical structures of caldariomycin, 2-chloro-1,3-cyclopentanedione, an intermediate in caldariomycin biosynthesis, and the synthetic compound monochlorodimedone. Second row B): halogenation of monochlorodimedone has been used for years as the standard assay in the search for halogenating enzymes. CPO: chloroperoxidase

The spectroscopic assay for haloperoxidase detection is based on the chlorination or bromination of monochlorodimedone, a synthetic compound with structural similarity to

2-chloro-1,3-cyclopentanedione, an intermediate in caldariomycin biosynthesis (See Fig. 1.3). The assay relies on the loss of $A_{278nm}$ that accompanies the conversion of monochlorodimedone to dichlorodimedone [23]. Using this assay many other chloro-, bromo- and iodoperoxidases have been detected and isolated from bacteria, fungi, marine algae, marine invertebrates, and mammals [24; 25]. Chloroperoxidase (EC 1.11.1.10) is a secreted heavily glycosylated monomeric hemoprotein [26]. Biochemical characterization showed that *Caldariomyces fumago* chloroperoxidase contains a cysteine bounded heme cofactor, exhibits catalase activity and can catalyzes P450-type reactions [25]. The enzyme three-dimensional structure [27] and biomimetic studies [28] led to the reaction mechanism (See Fig. 1.4). Heme-type haloperoxidases produce free hypohalous acid as the halogenating agent which is produced by the reaction of $H_2O_2$ in the presence of the heme cofactor and halide ions. Halogenation by chloroperoxidase requires a low pH (<3) and leads rapidly to the inactivation of the enzyme due to reaction with hypohalous acid. Substrates susceptible to halogenation include a variety of organic compounds and structures, for example, alicyclic ketones, phenols, flavonoids, aromatic acids, polycyclic aromatic hydrocarbons, biphenyls, lignin and lignin model steroids acetic and other aliphatic short-chain carboxylic acids as well as alkenes and alkynes [24].

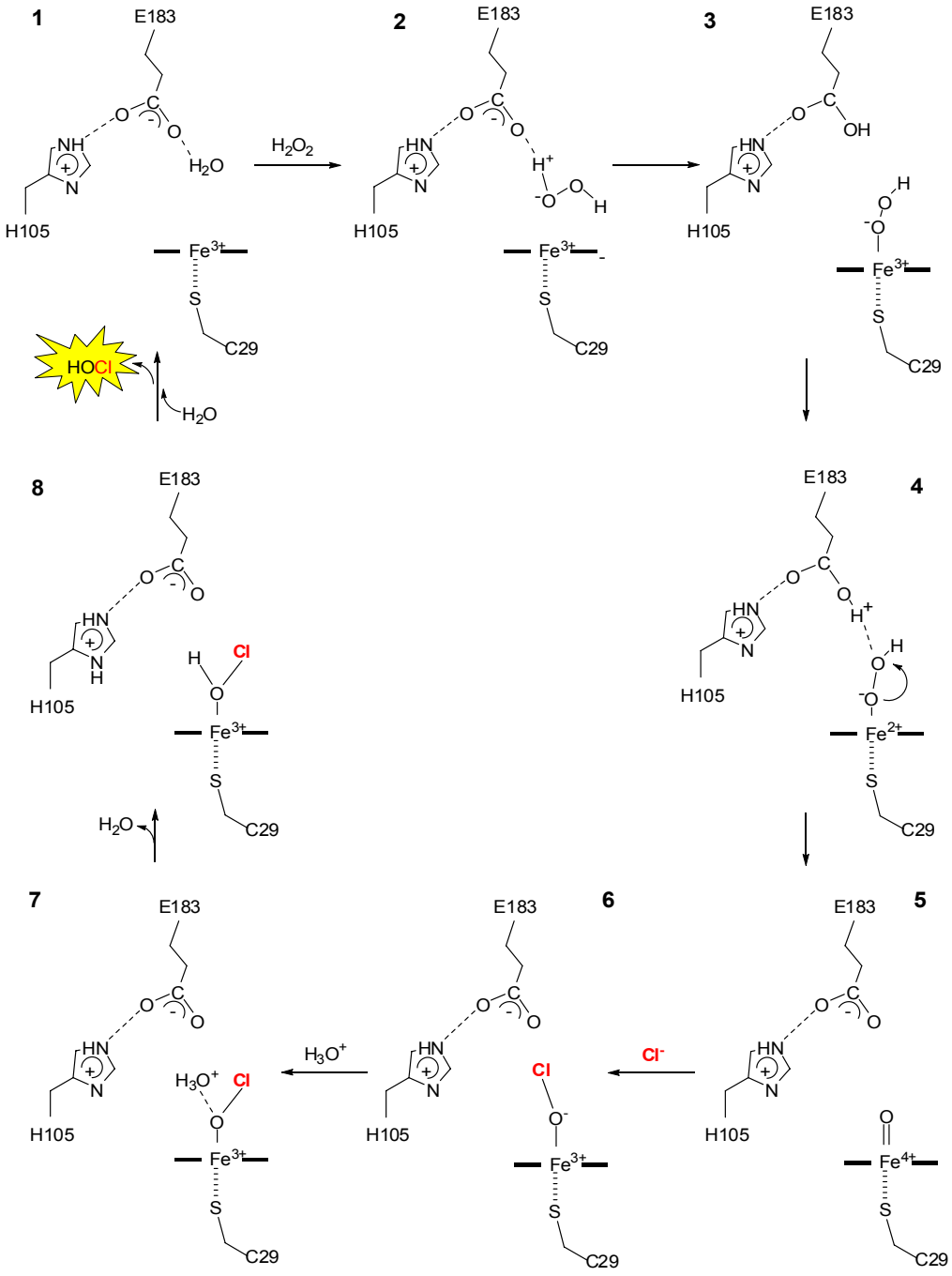



**Figure 1.4**

Reaction mechanism of heme-containing haloperoxidases showing the formation of hypohalous acid. Only the heme ion without the porphyrin ring system is shown. (**1**) enzyme active site in resting state. (**2**) hydrogen peroxide enters the active site. The reaction proceeds by heterolytic cleavage of the O–O peroxide bond facilitated by the glutamate residue 183 (E183) acting as an acid–base catalyst. (**3**) first deprotonating hydrogen peroxide. (**4**) then reprotonating the heme-bound ionized peroxide, releasing water and (**5**) leaving an oxyferryl centre (compound I). (**6**) Compound I reacts with Cl, (**7**) generating a Fe–OCl adduct, and as the chlorination reaction proceeds optimally at pH 3, (**8**) this is most likely protonated to Fe–HOCl. The release of HOCl returns the enzyme to its native state and the hypohalous acid reacts with the organic substrate outside the active site [1].

Following studies on enzymatic halogenation led to the identification of a second class of haloperoxidase comprising novel non iron enzymes which had a similar reactivity to Chloroperoxidases. These enzymes isolated from marine algae, lichen, and fungi contain Vanadium as cofactor [29]. Investigation of the reaction mechanism and elucidation of the X-ray structure of the vanadium-enzyme from the fungus *Curvularia inaequalis* showed that this enzyme also produces hypohalous acid as the halogenating agent [30] (See Fig 1.5). This class of enzyme is known as vanadium dependent chloroperoxidase.

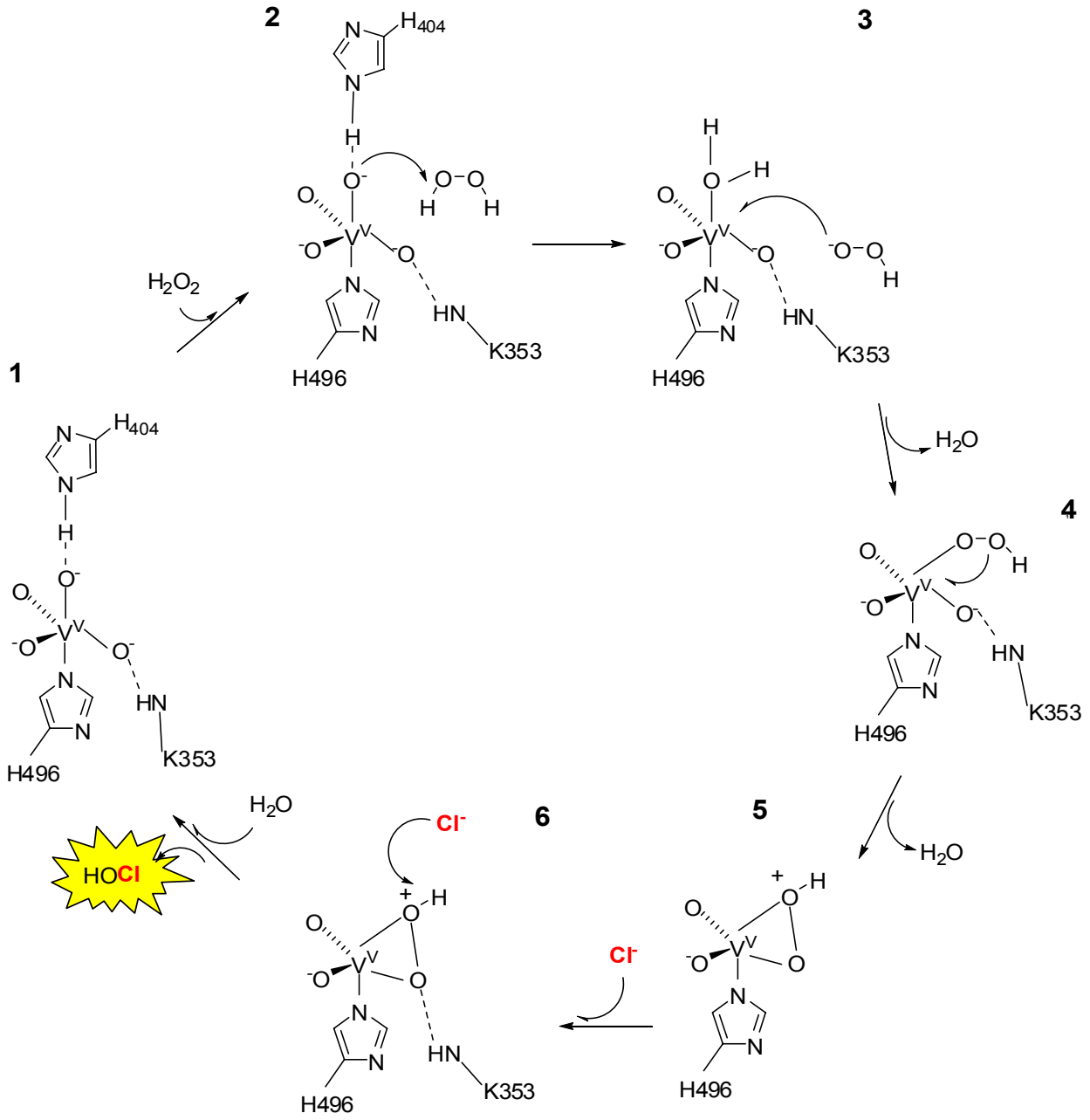



**Figure 1.5**
Proposed reaction mechanism of vanadium-containing haloperoxidases from the fungus *Curvularia inaequalis* [30; 31]: (**1**) in resting state Vanadium is coordinated by four oxygen atoms and Histidine 496 (H496). The apical oxygen is hydrogen bonded to His404. (**2,3**) peroxide molecule approaches and gets singly deprotonated. The water molecule leaves the vanadium-coordinated sphere. (**4**) the hydroperoxide coordinates to the vanadium at this empty coordination site. (**5**) the OH ligand is displaced by peroxide oxygen. (**6**) nucleophilic attack by chloride occurs with consequent formation of hypohalous acid as actual halogenating agent [31; 32; 33].

## 1.4 Perhydrolases

Using the monochlorodimedone assay, further halogenating enzymes requiring hydrogen peroxide for their activity, were identified from halogenated metabolites-producing *Pseudomonas* and *Streptomyces*. These halogenating enzymes were shown to contain neither a heme group nor any other metal ion. Although they required hydrogen peroxide for their halogenating activity, they appeared different to the peroxidases class. The three-dimensional structure of these halogenating enzymes shows a catalytic triad consisting of a serine, an aspartate and a histidine residue revealing that they belong to the α/β hydrolase fold enzyme family [34; 35]. These enzymes catalyze the peroxidation of organic acid to peroxoacid in the presence of $H_2O_2$ through a reaction mechanism that closely resembles transesterification. Subsequently, the halide ion is oxidized by the peroxoacid to hypohalous acid, and then non-enzymatic halogenation of organic compounds such as monochlorodimedone by the resulting hypohalous acid takes place [36] (See Fig. 1.6).

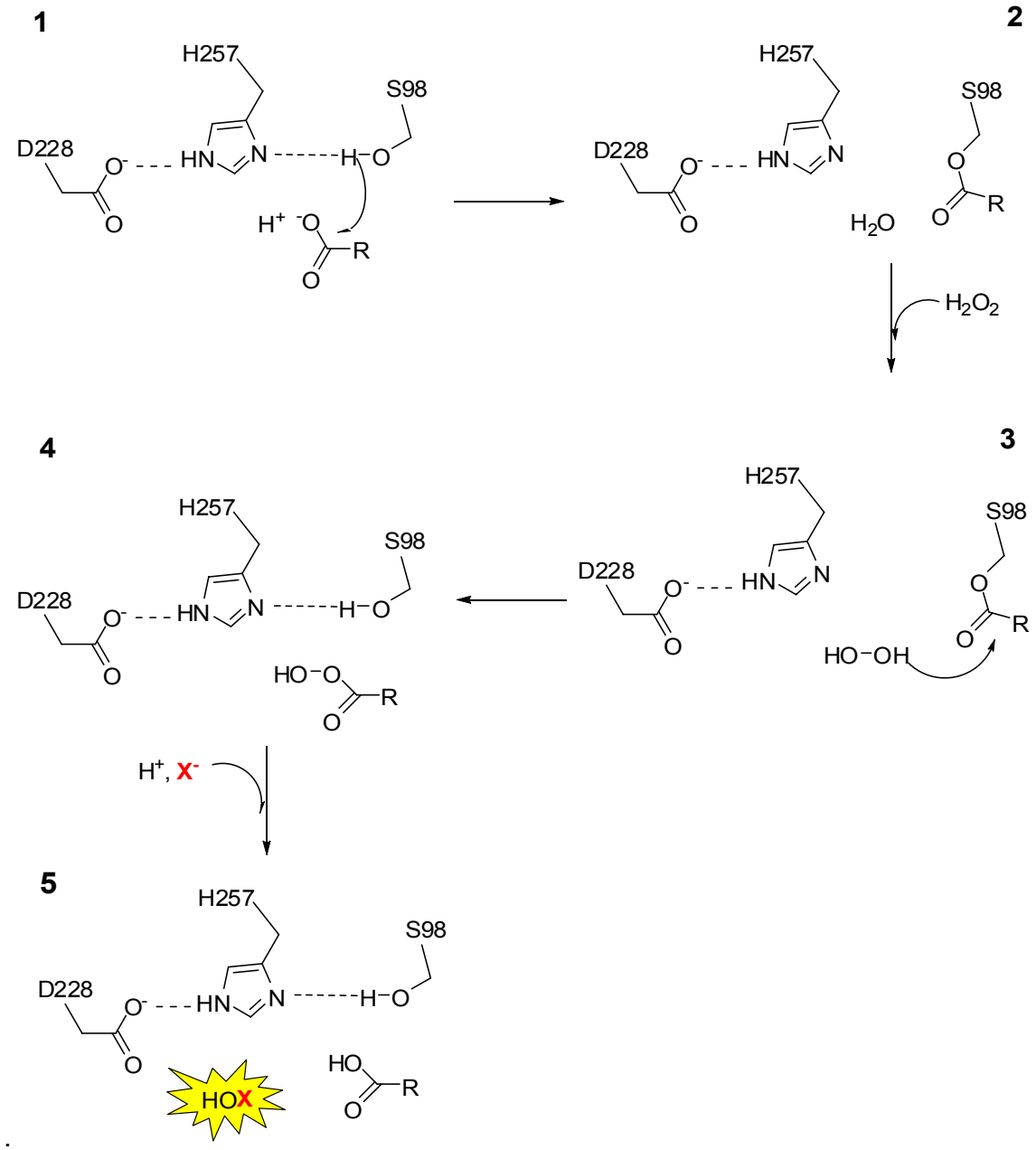



**Figure 1.6**
Proposed reaction mechanism of perhydrolase *Streptomyces aureofaciens* [35]. Three residues that make up the catalytic triad, Serine 98 (S98), Histidine 257 (H297) and Aspartate 228 (D228), catalyze the overall reaction. At the optimum pH 6.0, the active H257 is hydrogen-bonded to S98 and shares the proton at Nσ1 with the carboxylate group of D228. (**1-2**) the first part of the reaction is the nucleophilic attack on the carboxyl carbon atom of the organic acid by S98 and the formation of the acyl-enzyme by elimination of water. (**3-4**) the second part shows the hydrolysis of the acyl-enzyme by nucleophilic attack of hydrogen peroxide and the formation of the peroxoacid. The mechanism is formulated in analogy to the well-known mechanism of the serine proteases and esterases. (**5**) the final halogenation reaction is nonenzymatic and to proceed by the formation of hypohalous acid where the hydrophobic environment of the active site pocket protects the peroxoacid against hydrolysis, while halide is transported to the active site of the enzyme [32].

Perhydrolases had been, as a result of this, thought to be involved in the synthesis of halogenated metabolites *in vivo*, and had been called 'non-heme haloperoxidases', to distinguish them from the 'true' haloperoxidases (heme- or vanadium-dependent enzyme). There are now evidences that perhydrolases do not participate in the biosynthesis of halogenated compounds. A perhydrolase-deficient mutant of *Pseudomonas fluorescens* yielded a chlorinated metabolite, pyrrolnitrin, like the parent strain [37], and the *Streptomyces* enzyme, which chlorinates tetracycline, is not related to the perhydrolases family [19].

Since, heme/vanadium-containing haloperoxidases and perhydrolases are halogenating enzymes without any substrate specificity and regioselectivity; it is very unlikely that haloperoxidases and perhydrolases could be involved in the biosyntheses of complex halogenated natural products that present exquisite and specific modification. Notably the involvement of the haloperoxidases myeloperoxidase, in defense mechanisms, where it halogenates molecules in a non specific manner is well established [38]. From these observations it had been concluded that other halogenating enzymes than haloperoxidases must exist [25].

## 1.5 Flavin Dependent Halogenase

A convincing example of the non-involvement of haloperoxidases in biological halogenations is in the biosynthesis of the antifungal compound pyrrolnitrin by a number of *Pseudomonas* species, in which a new type of halogenating enzyme has been discovered. Four genes encode the proteins involved in the biosynthetic pathway to pyrrolnitrin; *prnABCD* (See Paragraph 1.10) and the functions of the genes have been determined by identifying the intermediates that accumulated in cultures of *prn* deletion mutants [20; 21]. The gene products of *prnA* was proved to catalyze regiospecific chlorination of tryptophan, while *prnC* catalyses the halogenation of monodechloroaminopyrrolnitrin to aminopyrrolnitrin [39]. Both enzymes required $FADH_2$ for activity and NADH was necessary for the reduction of the flavin by a non-specific reductase [18]. After the discovery of these two enzymes the number of FADH-dependent halogenases involved in the biosynthesis of natural halogenated compound has been increasing faster [4], in Table 1.1 are reported some of the better characterized enzymes [4; 39]; in Figure 1.7 their substrate and products .

| *Producing strains* | *Enzyme* | *Substrates* | *Halometabolite* | **Identity to PrnA / PrnC** |
|---|---|---|---|---|
| *Pseudomonas fluorescens* BL915 | Tryptophan 7-halogenase (PrnA) | L- and D-tryptophan and some tryptophan and indole derivatives (1) | Pyrrolnitrin (5) | - / 14% |
| *Lechevalieria aerocolonigenes* | Tryptophan 7-halogenase (RebH) | L- and D-tryptophan (1) | Rebeccamycin (6) | 52% / 12% |
| *Streptomyces rugosporus* | Tryptophan 5-halogenase (PyrH) | L- and D-tryptophan (1) | Pyrrindomycin (7) | 37% / 13% |
| *Streptomyces albogriseolus* | Tryptophan 6-halogenase (Thal) | L- and D-tryptophan (1) | Thienodolin (8) | 53% / 16% |
| *Pseudomonas fluorescens* BL915 | Monodechloroamino-pyrrolnitrin 3-halogenase (PrnC-Hal) | Monodechloroamino-pyrrolnitrin (2) | Pyrrolnitrin (9) | 15% / - |
| *Actinoplanes* sp. ATCC 33002 | HalB | 2-(3,5-Dibromophenyl)pyrrole NOT THE PHYSIOLOGICAL SUBSTRATE (3) | Pentachloropseudilin (10) | 17 % / 40% |
| *Pseudomonas fluorescens* Pf-5 | PltA | Pyrrolyl-*S*-carrier protein (4) | Pyoluteorin (11) | 13% / 16% |

**TABLE 1.1**

Through comparisons of the amino acid sequences of PrnA and PrnC with the other halogenases it is clear that two classes of $FADH_2$-dependent halogenases are emerging: one that halogenates tryptophan and indole (PrnA-type) and the other that halogenates phenyl and the more reactive pyrrole (PrnC-type) [18]. The majority of the flavin-dependent halogenases have only been identified on the basis of sequence homology and not by experimental evidence. All flavin-dependent halogenases consist of about 500 amino acids and have two absolutely conserved regions. One is the GxGxxG motif located near the amino terminal end, which is involved in the binding of the flavin co-substrate, also found in flavin dependent monooxygenase enzymes.

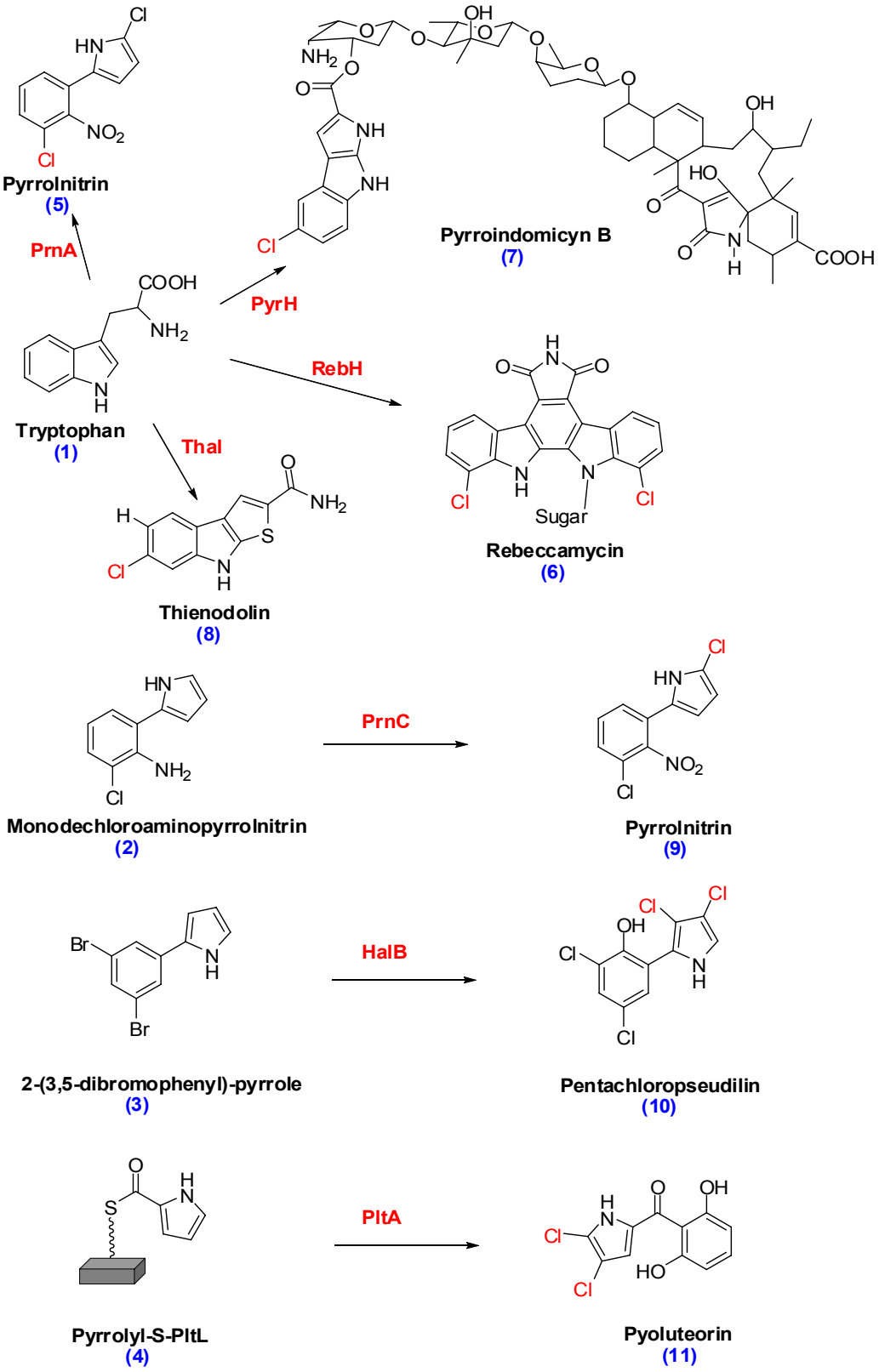

Pyrrolnitrin
(5)
PrnA
Tryptophan
(1)
PyrH
RebH
Thal
Pyrroindomicyn B
(7)
Rebeccamycin
(6)
Sugar
Thienodolin
(8)
PrnC
Monodechloroaminopyrrolnitrin
(2)
Pyrrolnitrin
(9)
HalB
2-(3,5-dibromophenyl)-pyrrole
(3)
Pentachloropseudilin
(10)
PltA
Pyrrolyl-S-PltL
(4)
Pyoluteorin
(11)

**Figure 1.7**

Recently the crystal structure of PrnA has been resolved by Changjiang Dong in our group, shedding new light on the mechanism of tryptophan halogenation [15]. Paradoxically, the proposed mechanism involves the formation of hypochlorous acid by nucleophilic attack by $Cl^-$ on flavin hydroperoxide (See Fig. 1.8). The HOCl is not released into the solvent but rather travels along to the substrate where aminoacids, conserved in all the know tryptophan halogenase, help the hypochlorite mediated halogenation. This type of halogenating chemistry is limited to electron rich substrates such as aromatic ring containing molecules [4; 18].

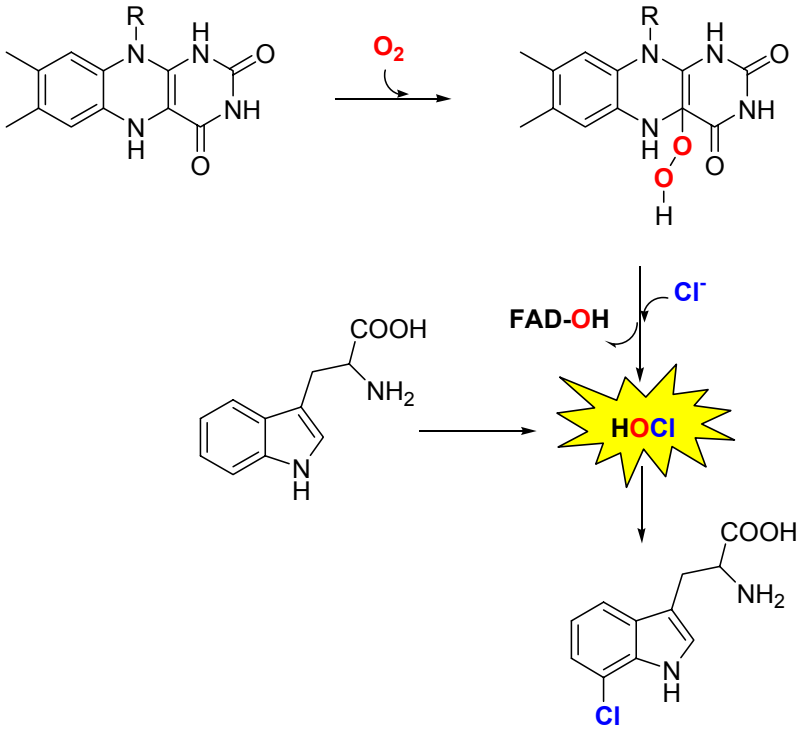



**Figure 1.8**
Proposed reaction mechanism for flavin-dependent halogenases. Tryptophan is shown as an example for a substrate. The mechanism was suggested by Dong *et al.* [15] based on the three-dimensional structure of PrnA and biochemical data. HOCl depicts a hypochlorite molecule that is not released from the active site of the enzyme.

## 1.6 PrnA a FADH dependent tryptophan 7-halogenase

PrnA is a regioselective tryptophan halogenase, involved in the biosynthesis of pyrrolnitrin, that catalyses incorporation of chlorine into the 7 position of the tryptophan indole ring. Discovered in 1997 [20; 40] by a genetic approach, it was soon realized that the protein required FADH for the catalysis of the halogenation reaction [40]. The isolation of pure PrnA was achieved in 2000 [41]. In the first step of the purification protocol, a protein component absolutely necessary for the enzyme activity was partially removed. This protein component is a flavin reductase that reduces FAD to FADH using NADH. The gene coding for this reductase is not part of the pyrrolnitrin biosynthetic cluster. Recently the entire genome sequence of *Pseudomonas fluorescens* Pf-5 strain has been released [42]. Analysis of the DNA sequence next to the pyrrolnitrin operon revealed the presence of a putative protein containing a FAD reductase domain [4]. The *Pseudomonas* flavin reductase activity can be substituted by flavin reductase from other bacteria such as *Fre*, the *E.coli* NAD(P)H-flavin oxidoreductase; *Ssue*, the flavin reductase component of the alkanesulfonate monooxygenase from *E.coli* or by NADH oxidase from *Thermus Thermophilus* [41]. To investigate whether a direct contact between the flavin reductase and the PrnA is necessary for the halogenating activity, the two protein components were separated by a dialysis membrane. Even under these conditions, chlorination of tryptophan occurred, although only with about 60% of efficiency seen normally [43]. The characterization of RebH, a FADH dependent tryptophan 7-halogenase, involved in the biosynthesis of rebeccamycin RebH has recently been reported by Yeh E., *et al*. [44]. RebH, shares 55% identity with PrnA[45]. Its partner reductase, RebF, has been identified within the rebeccamycin gene cluster [44; 45]. *In vitro* assays revealed a very low $K_{cat}$ for both PrnA 0.1$min^{-1}$ [15] and RebH 1.4$min^{-1}$ [44]. For PrnA the lower turn over number is ascribed to the missing proper reductase partner. The fact that PrnA can utilize chemically reduced diffusible FADH [43] does not exclude a possible interaction between the two components that could give better kinetic parameters. Such protein-protein interaction is thought to occur in the case of the FADH-dependent 4-hydroxyphenylacetate monooxygenase (HpaB) and its flavin reductase (HpaA) from *E.coli* [46] or the styrene monooxygenase (StyA) and its flavin reductase (StyB) from *Pseudomonas* sp. VLB120 [47] but no structural evidences has been reported so far. Only in the luciferase/flavin reductase system from *Vibrio harveyi* a FMNH tunnel system in

the donor-acceptor enzyme has been described [48]. Another intriguing hypothesis is that a third partner protein would function as shuttle between reductase and halogenase [43]. Speculative reaction mechanisms for the FAD-dependent halogenase family have been proposed since its initial purification[4] Only when PrnA structure was solved with bonded substrate, product and cofactor [15; 49] a sensible reaction mechanism could be proposed. Yeh E. [44] demonstrated that when the RebH/RebF halogenation reaction was prepared anaerobically and carried out introducing $O_2$ slowly product formation was increased by 10 fold. Thus a low $pO_2$ environment could better reflect the bacteria cytoplasmatic environment and be more favorable for observing the two component system reaction.

### 1.6.1 PrnA Overall Structure

The PrnA monomer (residues 2 to 518) is a single domain protein, formed by a parallelepiped "box" with a triangular "pyramid" leaning on one side of it (See Fig. 1.9). The "box" (residues 1-102, 159-401) was identify as the FAD binding module (See Fig. 1.10) and contains the glycine-rich phosphate binding consensus sequence [GXGXXG] conserved at the N-terminal of all the known FAD dependent halogenases. FAD is bound in a solvent-exposed groove adjacent to one or the two β sheet present in the domain (See Fig. 1.9). A second absolutely conserved motif, [WxWxIP] lies in the middle of the enzymes. Located near the flavin the two tryptophan residues are suggested to block the binding of a substrate close to the flavin preventing the enzyme from catalyzing monooxygenase type reactions [4]. The "pyramid" (residues 103-158, 402-518) is formed only by helices and represents the substrate binding module residues.

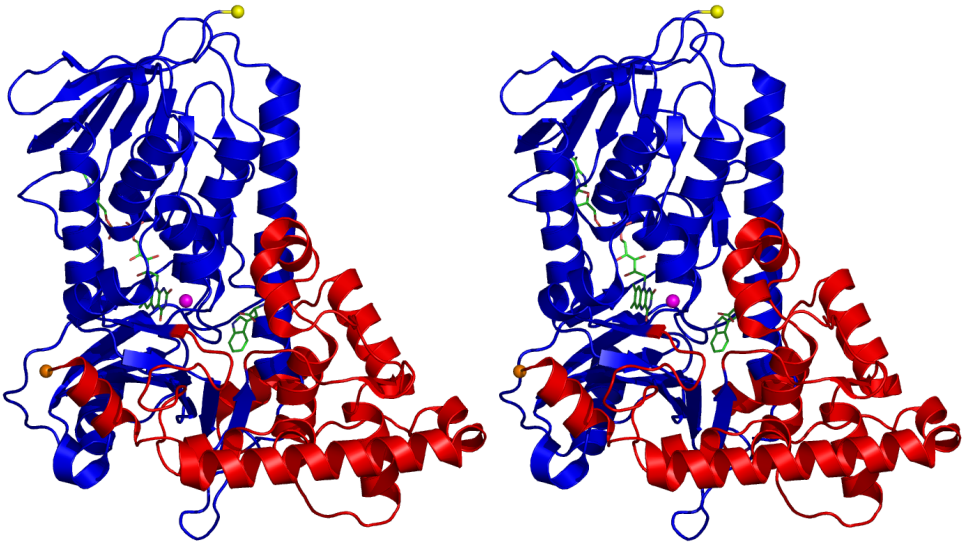


**Figure 1.9**
PrnA monomer shown in cartoon format. The flavin binding module is colored in blue and the substrate module colored in red. The N-terminal is denoted by a yellow sphere and C-terminal by a orange one. FAD and tryptophan are shown as sticks, carbon green, oxygen red, nitrogen blue and phosphorous orange . The $Cl^-$ ion is shown as a purple sphere.

Sequence alignment shows that only the flavin binding module is conserved in the halogenase superfamily [15]. The flavin binding module of PrnA shows significant similarity to the structure of *p*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens* [15] (See Fig 1.10 and 1.11), A $Cl^-$ anion is bound next to the isoalloxazine ring. No other $Cl^-$ was identified in the structure despite the presence of 50 mM NaCl in the crystallization solution. Complexes with tryptophan and 7-chlorotryptophan show that both the reaction substrate and the reaction products lie 10Å apart from the isoalloxazine ring in the same position. The indole ring keeps the tryptophan in place being stacked between aromatic residues W455 and H101 on one face and F103 on the opposite one. E346 forms a hydrogen bond with 7 tryptophan position. (See Fig. 1.13).

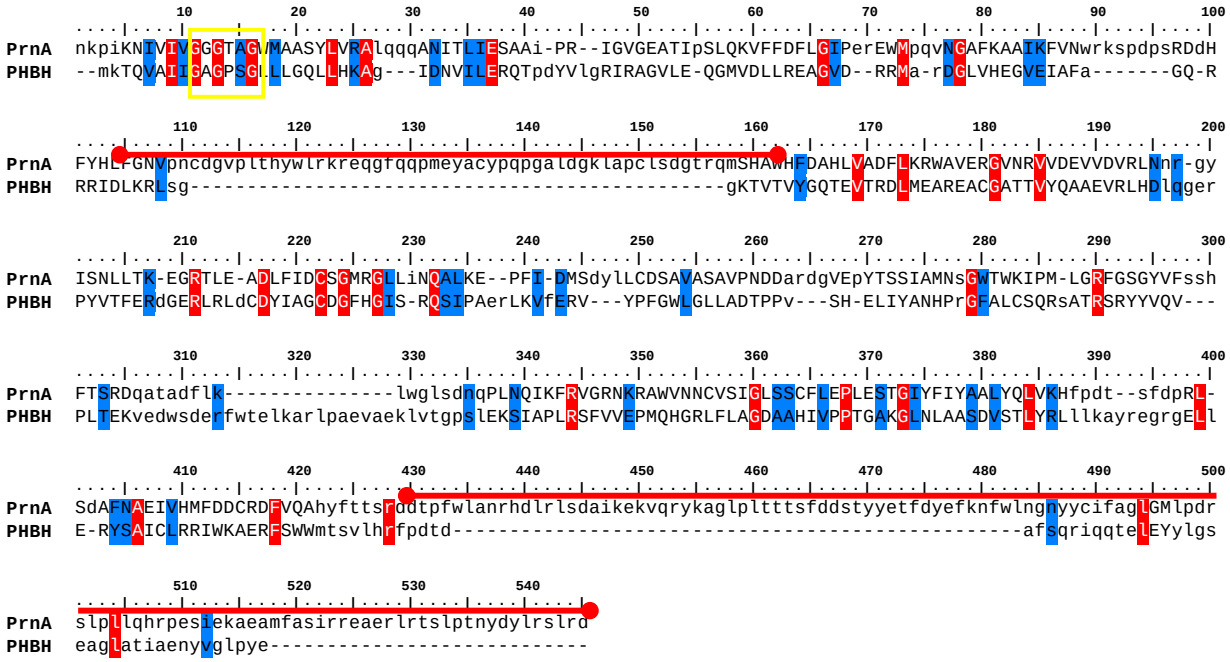



**Figure 1.10**

Sequence alignment based on structural superposition of PrnA on the *p*-hydroxybenzoate hydroxylase (PHBH) using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50] Shaded in red identical residues, in blue similar ones. The yellow box indicate the GxGxxG nucleotide binding motif. The red thick lines indicate the substrate binding domain regions present only in Prna.

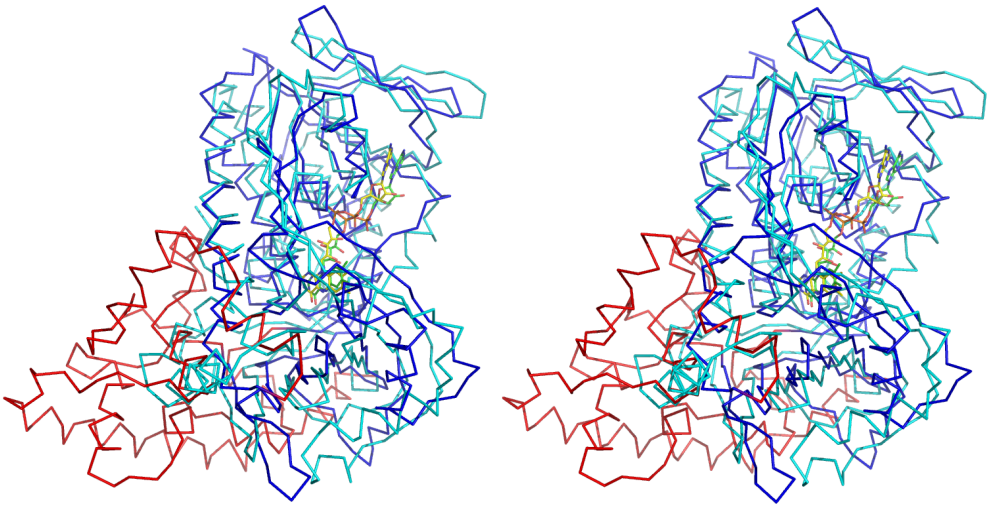


**Figure 1.11**

Superposition of PrnA and PHBH using the SSM algorithm of WinCOOT [51]. PHBH in cyan and PrnA, in blue the flavin binding module and in red the substrate binding module. FAD molecules are shown as a stick representation, yellow in PHBH green in PrnA.

### 1.6.2 PrnA Proposed Reaction Mechanism

Structural similarity to flavin-dependent monooxygenases helped to formulate a reaction mechanism for PrnA. The mechanism of these monooxygenases is well established [52; 53]. The enzyme is first reduced to give $FADH_2$, which then binds molecular oxygen to form a spectroscopically characterized highly reactive peroxide-linked flavin [52; 53]. This reactive intermediate is decomposed by the nucleophilic attack of an adjacent aromatic substrate, resulting in the transfer of an oxygen atom. Consistent with this established mechanism in PrnA $FADH_2$ should bind molecular oxygen forming the same peroxide-linked isoalloxazine ring (See Fig. 1.12). In PrnA, unlike PHBH, there is no room for an organic molecule to bind adjacent to the isoalloxazine ring, due to the presence of the conserved tryptophan residues W272 and W274. $Cl^-$ is ideally positioned to make a nucleophilic attack on the flavin peroxide, resulting in the formation of hydroxylated FAD and HOCl (See Fig. 1.12, 1.13).

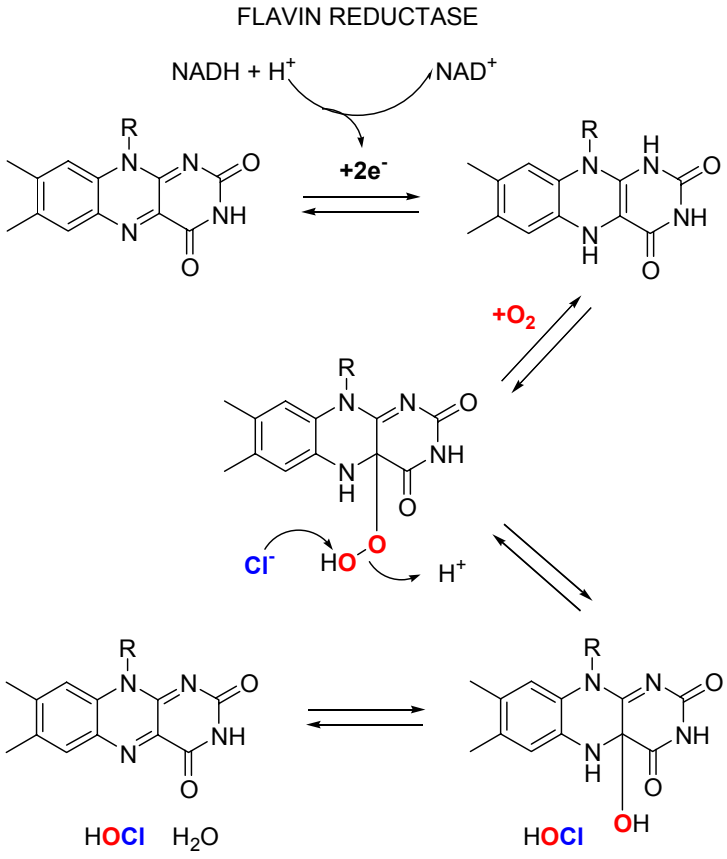



**Figure 1.12**
Generation of HOCl. Reduced flavin is generated by flavin reductase. The formation of the peroxy-flavin molecule is known to occur in related monooxygenase enzymes. The generation of HOCl by PrnA requires a nucleophilic attack of $Cl^-$ on the peroxy-flavin.

Very recently Walsh *et al*. were able to detect the peroxyflavin species generated during RebH (7 tryptophan halogenase) catalytic cycle [54] in accordance with the enzyme proposed mechanism. The substrate is 10Å distant at the end of a tunnel (See Fig 1.13). Within the tunnel there are no residues which would be predicted to be oxidized or chlorinated by HOCl (See Fig. 1.13). HOCl, after its formation, is prevented from diffusing into solvent by the protein structure and instead enters the tunnel, moving towards tryptophan. In the native PrnA structure, K79 makes a hydrogen bond with a water molecule located at the end of the tunnel adjacent to the Cl atom of 7-chlorotrypotphan (See Fig. 1.13 and 1.14).

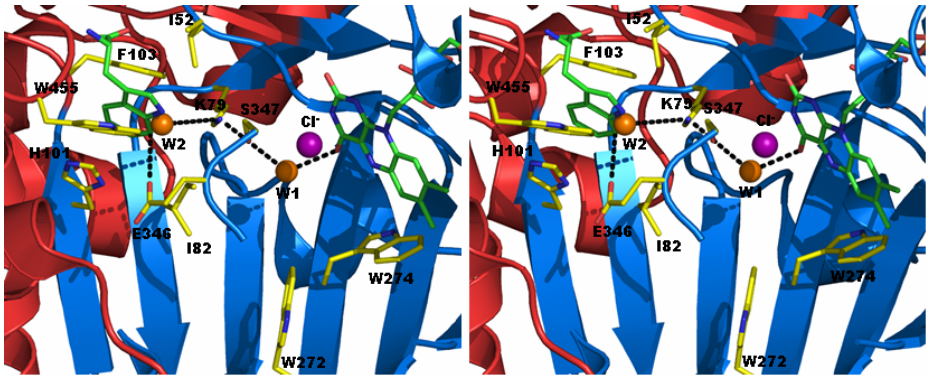



**Figure 1.13**
A tunnel connects the FAD and tryptophan binding sites. WaterN#1(W1), shown as an orange sphere, is found in all structures. WaterN#2 (W2) is a second water molecule found only in the native structure, where it is hydrogen bonded to K79. It has been placed in this figure to illustrate the path followed by HOCl through the protein tunnel from $Cl^-$ (Purple sphere) to tryptophan substrate. W2 is absent in both co-complexes because it would sterically clash with the ligands. The interaction of E346 with 7 position of tryptophan is shown as a black dotted line.

K79 may hydrogen bond to HOCl and thus position it to react with tryptophan; while E346 would help deprotonation of indole ring 7 position (See Fig. 1.13 and 1.14) The controlled spatial presentation of HOCl to substrate has been proposed as the base for regioselective halogenation [4; 15]. Consistent with this proposed mechanism chlorinating activity could not

be detected in a PrnA K79→A79 mutant enzyme. While the $K_{cat}$ of the PrnA E346→Q376 mutant decrease by two order of magnitude without affecting the $K_M$ [15].

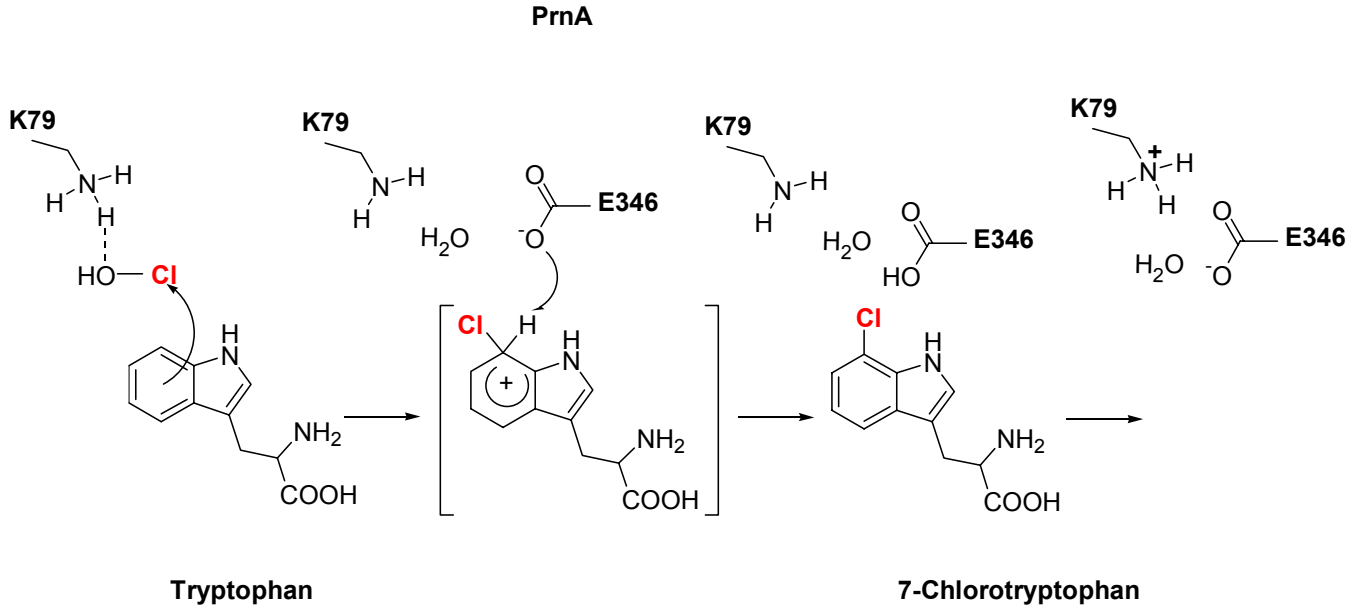



**Figure 1.14**
Proposed halogenation of tryptophan proceeds by electrophilic aromatic substitution at the 7 position. The complex shown in [] would be stabilized by interaction with E346. The indole ring stacks with W455 and H101 on one face and F103 on the other which may further stabilize the intermediate.

In Chapter 2 I describe the purification to crystal trial standard, the crystallization and structure of PyrH a tryptophan 5 halogenase member of the family of flavin dependent halogenases. I started this work to confirm the way this new class of enzymes achieve substrate halogenation and to shed light into the way tryptophan regioselective halogenation is accomplished.

## 1.7 Non-heme FeII α-ketoglutarate and $O_2$-dependent halogenases

In only very few case the substrate of the halogenated natural products are known or available. Due to this lack of knowledge *in vitro* halogenase activity could only be demonstrated for a few halogenases. This explains why progress on the investigation of halogenases involved in halometabolite biosynthesis has been slow and has only been accomplished with tryptophan halogenases so far [4]. There are only two examples where genes of flavin-dependent halogenases have been detected in gene clusters for the biosynthesis of chlorinated aliphatic compounds. One is *cmlS* from chloramphenicol biosynthesis and the other is ORF3 from the neocarzilin biosynthetic gene cluster. While chloramphenicol contains a dichloracetyl moiety of yet unknown origin, neocarzilin has a trichloromethyl group. A number of halometabolites from various *cyanobacteria* contain trichloromethyl groups, but no genes for flavin-dependent halogenases have been found in their biosynthetic gene clusters [4]. The mechanism of enzymatic chlorination of aliphatic compounds has been examined in the *Pseudomonas syringae* pv. *Syringae* produced phytotoxic lipodepsipeptide syringomycin E 7 [18; 22], and in the cyanobacterium, *Lyngbya majuscula*, produced molluscicidal compound barbamide [18; 55]. Biosynthetic investigations using isotope labelled compounds indicated that the biosynthetic origin of the 4-chlorothreonine residue in syringomycin E 7 is threonine and the trichloromethyl group of barbamide originates from the *pro-S* methyl group of leucine (See Fig. 1.15).

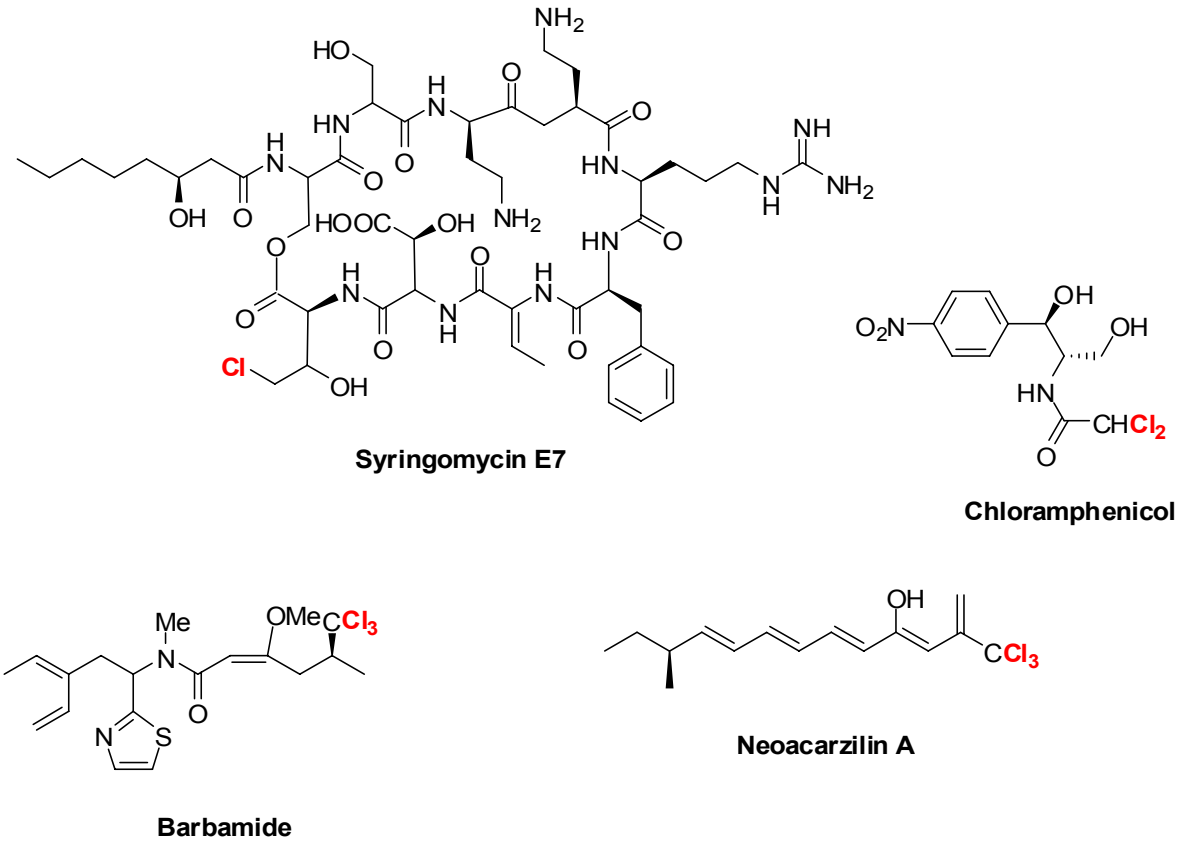



**Figure 1.15**

Such precursors are not obvious substrates for $FADH_2$-dependent halogenases. Feeding experiments with deuterated leucine demonstrated that C3 and C4 of leucine remain saturated during its incorporation into barbamide, thus the *pro-S* methyl group is not activated prior to chlorination, suggesting a halogenating mechanism involving radicals [18]. Similarities have been identified between the deduced amino acid sequences of BarB1/BarB2 in the barbamide gene cluster and SyrB2 in the syringomycin gene cluster. These proteins belong to a class of nonheme $Fe^{II}$, α-ketoglutarate-dependent enzymes and recently Walsh and colleagues demonstrated the novel halogenating activity of SyrB2, which chlorinates L-Thr linked to the peptidyl carrier protein SyrB1 but does not chlorinate free threonine [56]. Elucidation of the crystal structure of SyrB2 with both a chloride ion and α-ketoglutarate coordinated to the iron at 1.6Å resolution helped Walsh and colleagues to propose a rection mechanism for this new class of non heme

$Fe^{II}$, α-ketoglutarate-dependent halogenases [22] (See Fig. 1.16). Presumably the BarB1 and BarB2 enzymes, that have been shown to catalyze *in vitro* the triple chlorination of their substrate [55], catalyze the chlorination of leucine *via* a similar radical mechanism. This radical halogenating mechanism is able to chlorinate aliphatic unreactive substrates.

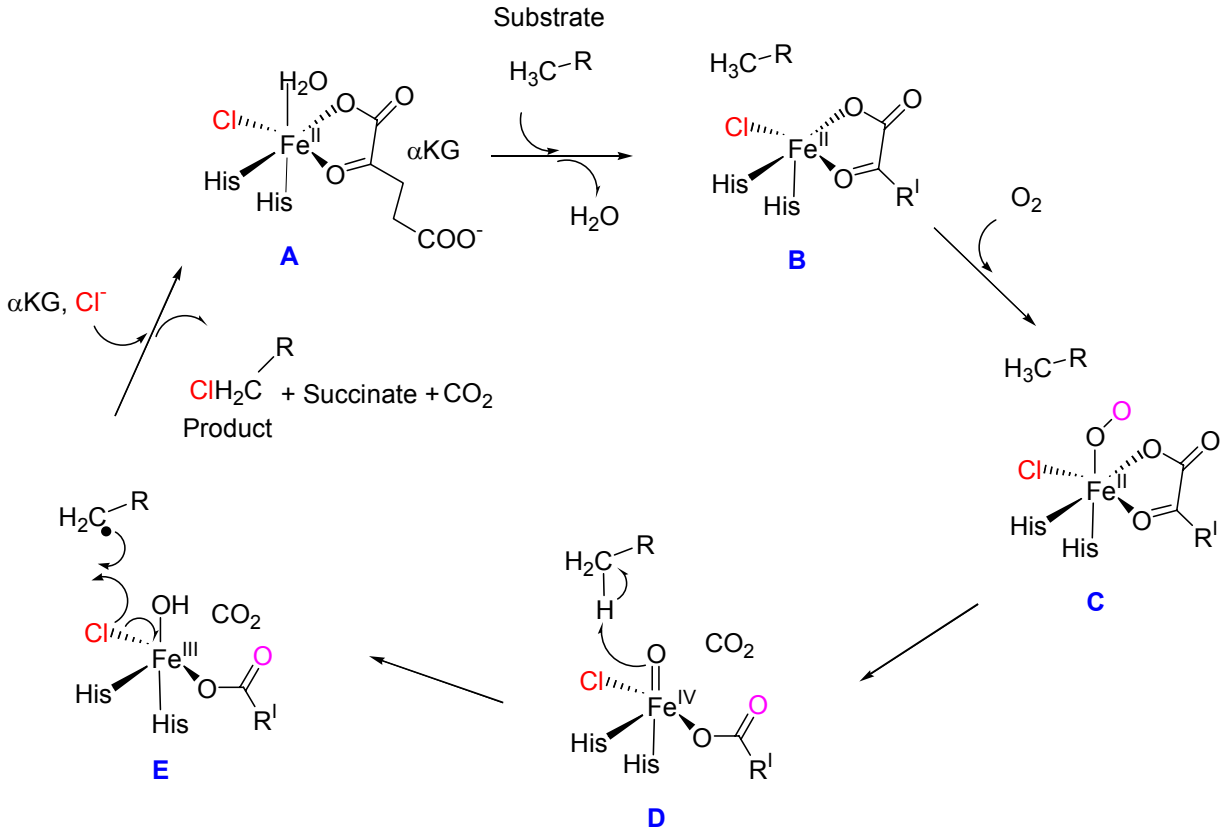
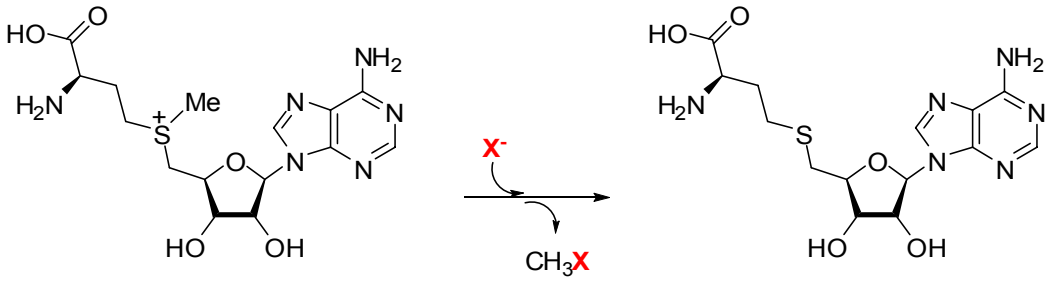



**Figure 1.16**

Walsh proposed mechanism for halogenation by SyrB2. As observed crystallographically, αKG, chloride and water coordinate the iron in the resting Fe(ii) state (A). L-Thr-*S*-SyrB1 binding would then exclude water from the active site and allow dioxygen to bind, (B and C). Decarboxylation of αKG would lead to the formation of a high-energy ferryl-oxo intermediate (D), which would then abstract a hydrogen atom from the substrate. The substrate radical would abstract the chlorine atom (E), producing chlorinated L-Thr-*S*-SyrB1 and regenerating the reduced Fe(ii) centre. After formation of the substrate carbon radical, however, some competition between transfer of $Cl^{\bullet}$ and $OH^{\bullet}$ would be expected. None threonine hydroxylation side reaction has been detected, indicating that $Cl^{\bullet}$ transfer is greatly favoured. There are several possible explanations for this selectivity, including the positioning of the substrate in the active site during the reaction, and the lower potential of $Cl^{\bullet}$ versus $OH^{\bullet}$[22].

## 1.8 Nucleophilic chlorination: Methyl-Tranferase

Halomethanes, particularly chloromethane, are produced by fungi, algae and higher plants. Investigations of cell-free extracts have led to the identification of methyl transferase enzymes, which transfer a methyl group from *S*-adenosylmethionine (SAM) to a chloride, bromide or iodide ion (See Fig. 1.17). Most of these enzymes are labile, making purification and characterization difficult. Kinetic measurements indicate that the preference of halides is I > Br > Cl, although the concentration of halide ions in the environment probably determines the proportions of the halomethanes produced by the organism. The biosynthesis of halomethanes may regulate the concentrations of halide ions in algae or contribute to halotolerance adaptations in plants [1].




***S*-adenosylmethionine** ***S*-adenosylhomocysteine**

$X^-$ = $Cl^-$, $Br^-$, $I^-$

**Figure 1.17**
Reaction catalyzed by *S*-adenosylmethionine: halide ion methyl transferase

As the enzyme can only utilize *S*-adenosylmethionine as methyl donor, it is limited to a single product.

## 1.9 Nucleophilic chlorination: Fluorinase

Fluorine, the most electronegative element, has a van der Waals radius similar to hydrogen Incorporating fluorine into an organic compound can alter its electronic properties without substantial steric effects. Fluorinated compounds have different biological activities than their non-fluorinated analogues. Thus many pharmaceutical compounds contain fluorine, such as the anticancer drug fluorouracil, the serotonin uptake inhibitor fluoxetine (Prozac®, Eli Lilly) and fluoroquinolone antibiotics such as ciprofloxacin [1]. An enzyme that catalyses the formation of C–F bonds under relatively mild conditions would be a very useful biocatalyst. Enzymatic formation of C–F bonds is possible with active-site mutants of glycosidase enzymes, where a nucleophilic glutamate residue is replaced with glycine, alanine or serine. The requirement for a nucleophile in the active site is satisfied by fluoride at high concentrations (2M), but the glycosyl fluorides formed are transitory [57]. The bacterium *Streptomyces cattleya* produces fluoroacetate and 4-fluorothreonine as secondary metabolites from fluoride ion [1]. During the last few years the enzyme responsible the formation of the C–F bond in *S. cattleya* has been identified [58] and its crystal structure solved [59]. The enzyme, also known as "fluorinase" (E.C. 2.5.1.63), catalyses the synthesis of 5′-fluoro-5′-deoxyadeonsine from *S*-adenosylmethionine and $F^-$ via an $S_N2$ reaction mechanism (See Fig 1.18). More recently it has been shown that can also utilize $Cl^-$ as substrate generating 5′-chloro-5′-deoxyadeonsine (See Fig 1.18). The reactions with both $F^-$ and $Cl^-$ are reversible. The equilibrium of the chlorination reaction lies substantially in favour of the substrates and was only possible to detect the chlorinated product using a coupled assay [60].

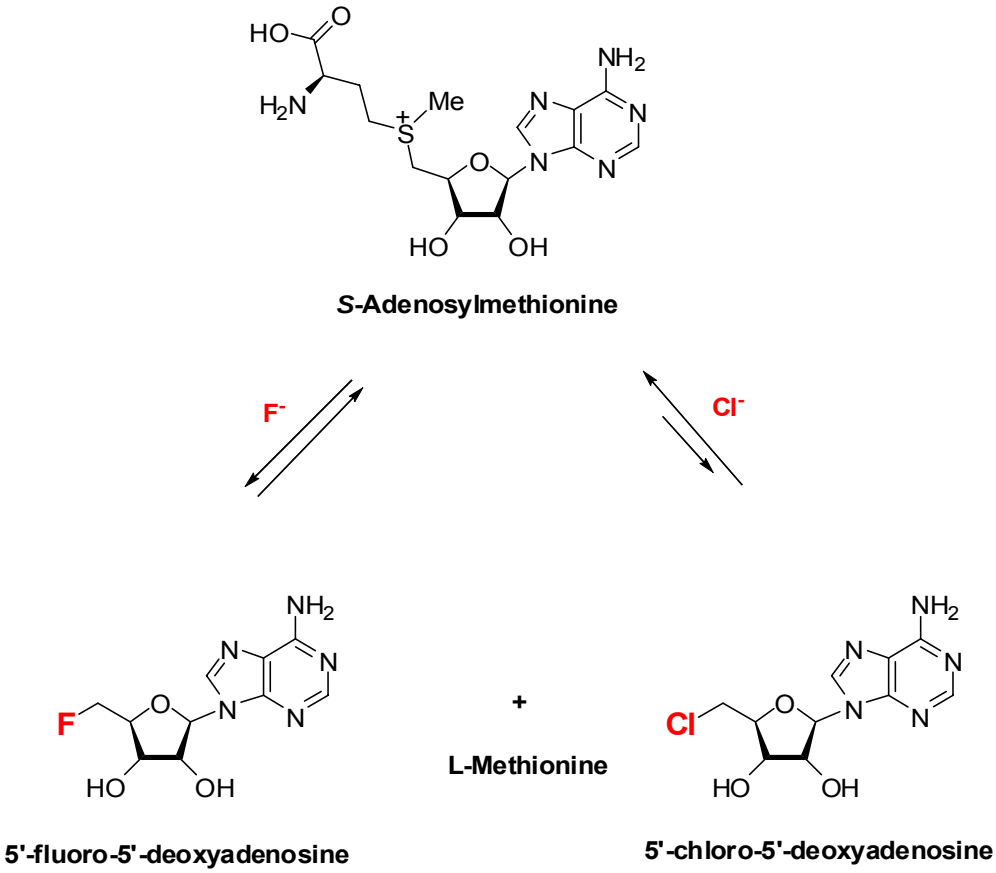



**Figure 1.18**
Nucleophilic chlorination and fluorination of SAM catalyzed by the fluorinase from *S. cattleya*

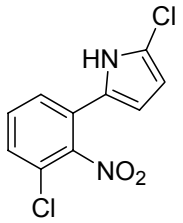
Interestingly, the discoveries of bacteria that biosynthesize fluorinated compounds were fortuitous and was a consequence of fluoride impurities present in the culture medium. It is very likely that other microorganisms have similar biosynthetic capabilities and a screening programme for such compounds would therefore be worthwhile [1].

## 1.10 Pyrrolnitrin

Pyrrolnitrin is one one the best known and most studied halogenated natural products. It constitutes the active ingredient of drugs for the treatment of superficial dermatophytic fungal infections and was used as lead structure for the development of the world marketed phenylpyrrol antifungal Fludioxonil [61]. Research on pyrrolnitrin antifungal and antibacterial activity, mode of action and biosynthesis has been conducted during the last 40 years. Despite that the enzymatic mechanisms that lead to its synthesis have not been completely elucidated, yet [61].

### 1.10.1 Discovery of Pyrrolnitrin

Pyrrolnitrin (3-chloro-4-(2'-nitro-3'-chlorophenyl)pyrrole; (See Fig. 1.19) is a tryptophan derived secondary metabolite with broad-spectrum anti-fungal activity which was first isolated in 1964 from *Pseudomonas pyrrocinia* [62]. Subsequently, pyrrolnitrin has been identified in several isolates of *Pseudomonas* and *Burkholderia* (previously *Pseudomonas*) and has been implicated as an important mechanism of biological control of soil-borne fungal plant pathogens by these strains [63; 64; 65]. Pyrrolnitrin production has also been documented for strains of *Enterobacter agglomerans* [66], *Myxococcus fulvus* [67], *Corallococcus exiguous* [67], *Cystobacter ferrugineus* [67] and *Serratia* spp [68].




**Pyrrolnitrin**

**Figure 1.19**

Tryptophan was identified as pyrrolnitrin precursor, feeding cultures with isotopically labelled compounds. Both D- and L-tryptophan isomers were efficiently incorporated into

pyrrolnitrin. Interestingly only the D-isomer was shown to be able to increase the antibiotic production [61; 69]. Early attempts to identify the pyrrolnitrin biosynthetic gene cluster using a genetic approach resulted in identification of *Pseudomonas* secondary metabolism global regulators [20; 70]. Only in 1997 van Peé and co-workers were able to identify the pyrrolnitrin biosynthetic gene cluster from *Pseudomonas fluorescens* BL915 [20; 21; 61] elucidating the definitive pyrrolnitrin biosynthetic pathway.

### 1.10.2 Pyrrolnitrin Biosynthetic Gene Cluster and Pathway

*P. fluorescens* BL915 is a biological control strain described by Hill *et al.* [70]. It has been reported to produce pyrrolnitrin and to be an effective biocontrol agent for plant diseases caused by fungal pathogens including the damping-off pathogen *Rhizoctonia solani*. Van Peé and co-workers [20; 21] isolated a 6.2-kb genomic DNA fragment from this strain containing a cluster of four genes (*prnABCD*) required for the biosynthesis of pyrrolnitrin (See Fig 1.20). *P. fluorescens* BL915 deletion mutants in any of the four genes resulted in a pyrrolnitrin -nonproducing phenotype. Reintroduction of the entire *prnABCD* gene cluster in a *prnABCD*$^-$ mutant was shown to complement the pyrrolnitrin-nonproducing phenotype. One by one introduction of the four *prn* genes in a *prnABCD*$^-$ mutant was shown to complement the enzymatic activity proposed for each of the four genes. Furthermore, transfer of the entire *prnABCD* cluster to *E. coli* resulted in the production of pyrrolnitrin, thereby demonstrating that the four genes were the only ones required to encode the pathway for pyrrolnitrin biosynthesis [20; 21; 70].

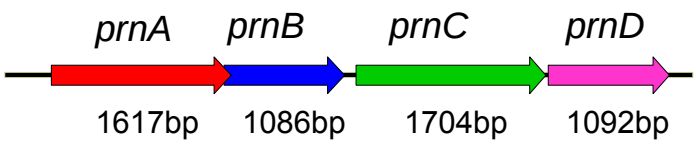



**Figure 1.20**

The succession of the *prn* genes in the operon is identical to the order of the reactions in the biosynthetic pathway proposed by van Peé (See Fig. 1.21): the *prnA* gene product catalyzes the chlorination of L-TRP to form 7-chloro-L-TRP.

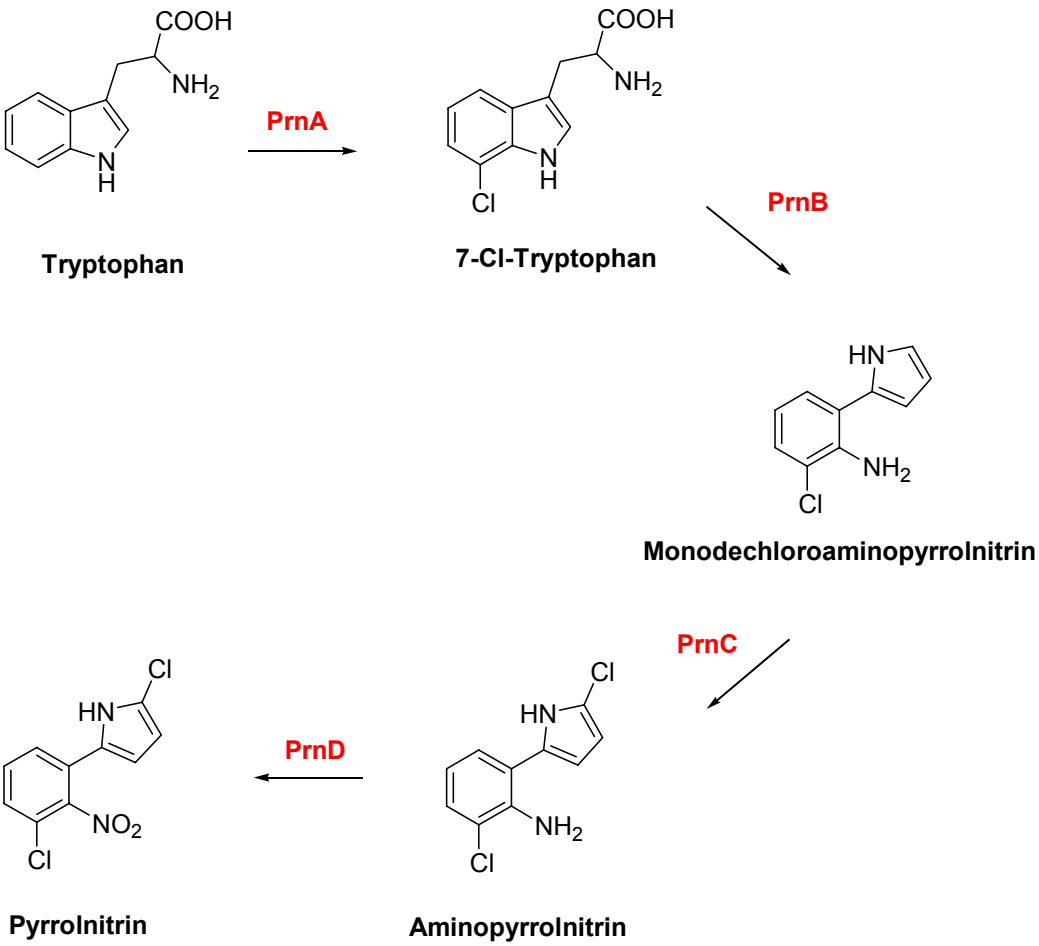



**Figure 1.21**

The *prnB* gene product catalyzes a ring rearrangement and a decarboxylation converting 7-chloro-L-TRP to monodechloroaminopyrrolnitrin (MDA). The *prnC* gene product chlorinates MDA at position 3 of the pyrrole ring to form aminopyrrolnitrin (APRN). The *prnD* gene product catalyzes the oxidation of the amino group of APRN to a nitro group to form pyrrolnitrin.

### 1.10.3 The Pyrrolnitrin Biosynthetic Gene Cluster is Conserved

More recently, the *prnABCD* cluster from *P. fluorescens* BL915 was used to identify homologous gene clusters from the pyrrolnitrin producing bacteria *Pseudomonas pyrrocinia*, *Burkholderia cepacia* LT4-12-W and *Myxococcus fulvus* Mx f147 [71]. The *prnA* gene fragment from strain BL915 was used to clone *prnA* homologues from *Pseudomonas fluorescens* CHA0 and *Pseudomonas aureofaciens* ACN [71]. Both orientation and length of the *prnABCD* genes appear to be conserved within the pyrrolnitrin biosynthetic gene clusters. The only exception is the position of *M. fulvus prnA*, which is located on the operon antisense strand downstream of *prnD* (See Fig. 1.22).

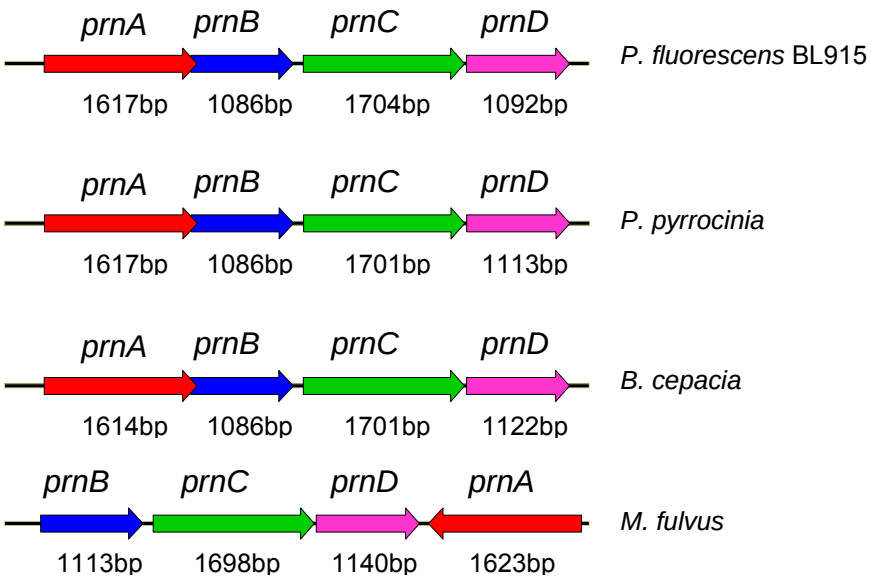



**Figure 1.22**

With the exception of *M. fulvus prnA*, the deduced amino acid sequences are >59% similar among the strains. This indicates that the biochemical pathway for pyrrolnitrin biosynthesis is highly conserved (See Table 1.2).

**Table 1.2**

| | *P. fluorescens* BL915 | *P. aureofaciens* ACN | *P. fluorescens* CHA0 | *P. pyrrocinia* | *B. cepacia* |
|---|---|---|---|---|---|
| *prnA* | | | | | |
| *M. fulvus* | 44.7 | 45.5 | 44.1 | 44.5 | 45.2 |
| *B. cepacia* | 89.4 | 90.1 | 88.5 | 92.2 | |
| *P. pyrrocinia* | 94.4 | 93.1 | 92.9 | | |
| *P. fluorescens* CHA0 | 94.8 | 93.1 | | | |
| *P. aureofaciens* ACN | 95.0 | | | | |
| *prnB* | | | | | |
| *M. fulvus* | 61.6 | | | 61.9 | 59.4 |
| *B. cepacia* | 80.9 | | | 85.6 | |
| *P. pyrrocinia* | 86.2 | | | | |
| *prnC* | | | | | |
| *M. fulvus* | 79.3 | | | 79.5 | 79.2 |
| *B. cepacia* | 93.7 | | | 94.2 | |
| *P. pyrrocinia* | 95.2 | | | | |
| *prnD* | | | | | |
| *M. fulvus* | 62.1 | | | 62.0 | 61.2 |
| *B. cepacia* | 87.4 | | | 87.9 | |
| *P. pyrrocinia* | 91.2 | | | | |

Predicted aminoacid sequences were compared among the strain using the CLUSTAL alignment algorithm. The values shown are the percentage similarity. Adapted from Hammer *et al* 1999 [71].

Further pyrrolnitrin biosynthetic operons have been identified in recently sequenced *Pseudomonas* related strain such as *Pseudomonas fluorescens* Pf-5 [42], *Burkholderia ambifaria* AMMD, *Burkholderia pseudomallei* strains 1710a,1710b, 688,1655, S13 (Source DOE Joint Genome Insitute Integrated Microbial Genomes webpages).

#### *1.10.3.1 prnA*

The *prnA* gene product PrnA, a member of the family of $FADH_2$ dependent halogenases, catalyzes the chlorination of L-tryptophan to form 7-chloro-L-tryptophan and has been discussed earlier in Paragraph 1.6.

#### *1.10.3.2 prnB*

The *prnB* gene product catalyzes the conversion of 7-chloro-L-tryptophan to monodechloroaminopyrrolnitrin coupling the rearrangement of the indole ring to a phenylpyrrole and a decarboxylation. Van Peé *et al*. have reported that this enzyme is active on tryptophan too [21]. The predicted *prnB* product codes for a 361aa protein with an $M_w$ of about 39KDa. This predicted protein does not contain any known signature sequence, cofactor binding site or conserved motif. Furthermore, it has not any statistically significant similarity profile match with any known protein included in any public protein database [20; 21]. Interestingly *prnB* GTG initiation codon overlaps with *prnA* gene stop codon, suggesting translational coupling of the two genes reflecting a close metabolic regulation between 7-Cl-Tryptophan synthesis and its conversion to monodechloroaminopyrrolnitrin [20].

PrnB is a new unique enzymatic protein completely uncharacterized so far. In Chapter 3 we describe PrnB cloning, over-expression and purification to crystal trial standard and in Chapter 4 its crystallization and structure solution. This work is a step foward in the understanding of the mechanism of action of the enzyme and the complete elucidation of the pyrrolnitrin biosynthetic pathway.

#### *1.10.3.3 prnC*

The *prnC* gene product is a 65KDa protein (567aa), which chlorinates monodechloroaminopyrrolnitrin at position 3 of the pyrrole ring to form aminopyrrolnitrin. The aminoacid sequences of the two chlorinating enzymes PrnC and PrnA show an identity of 18% and a similarity of 26%. The low sequence similarity could explain the different substrate specificity. Despite that, both PrnC and PrnA contain two FAD-dependent halogenase conserved motifs. The nucleotide binding site with the consensus [GxGxxG] motif and the double tryptophan motif [WxWxIP], suggesting that the two enzymes work in a similar way [20; 21].

### *1.10.3.4 prnD*

The *prnD* gene product encodes a 363aa protein with a $M_w$ of 40KDa. It catalyzes the final oxidation step of the amino group to a nitro group in APRN to form pyrrolnitrin. The N-terminal sequence of the enzyme contains two conserved motifs characteristic of the family of Rieske type non-heme oxygenases. The first motif, a Rieske type iron sulphur centre, consist of the consensus [CxH 15-17aa CxxH] and is involved in the binding of a [2Fe-2S] cluster. The second one, 90aa apart from the first motif, is the highly conserved [DxxHxxxxH] sequence which binds the mononuclear nonheme Fe(II) that binds molecular oxygen (See Fig. 1.23).

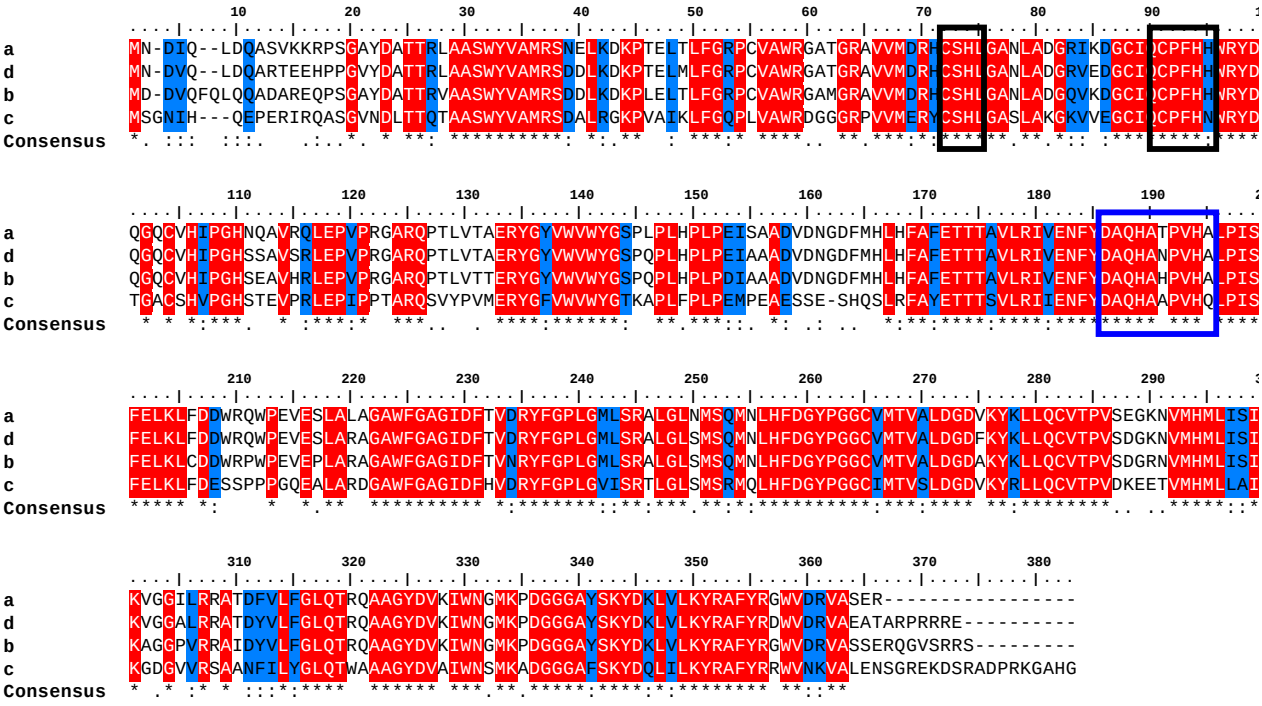



**Figure 1.23**
Sequence alignment of PrnD from *P. fluorescens* (a) *B. cepacia* (b) *M. fulvus* (c) *B. pyrrocina* (d); Rieske type iron sulphur binding motif highlited in black, iron binding motif highlighted in blue. Shaded in red identical residues, in blue similar ones (threshold 100%).

PrnD is a member of the dioxygenase superfamily. Despite that, it shows less than 20% homology to any known oxygenase structure and performs completely different chemistry [20; 21]. First attempts to obtain PrnD over-expression in *E.coli* led to production of insoluble protein [21]. More recently soluble MBP tagged prnD over-expression in *E.coli* has been

reported together with the cleavage of the MBP tag and reconstitution of the iron sulphur cluster [72]. Availability of reconstituted enzymatically active prnD helped to propose the first enzyme mechanism for the conversion of arylamines into arylnitro compounds in which the enzyme catalyzes at least three consecutive reactions [73] (See Fig. 1.24).

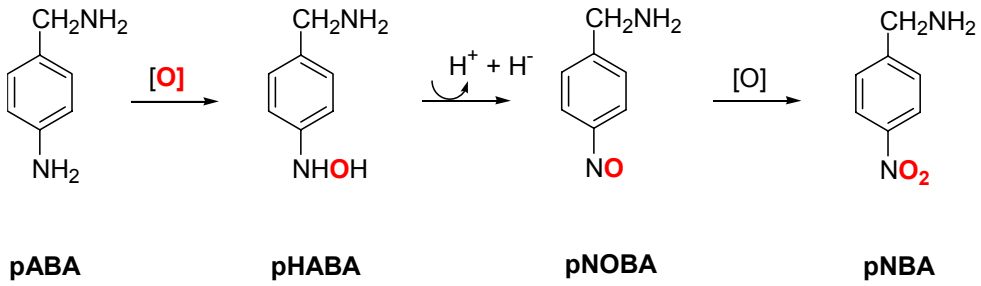


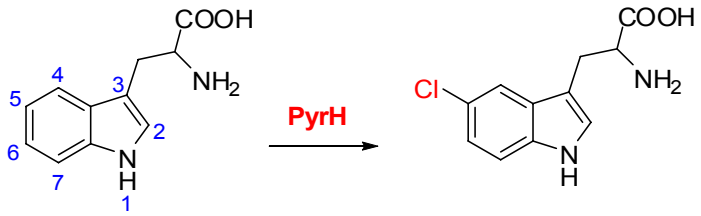

**Figure 1.24**
Proposed PrnD reaction mechanism [72; 73]. Due to the difficulties in obtain APRN 4-aminobenzylamine (pABA) was used as substrate for the enzyme. Since Rieske-oxygenase are typically comprised of two protein components: a terminal oxygenase and a flavin reductase *E.coli* Ssue reductase was added to reconstituted PrnD together with NADPH, FMN and pABA. 4-hydroxylaminobenzylamine (pHABA) and 4-nitrosobenzylamine (pNOBA) and 4-nitrobenzylamine (pNBA) were detected into the reaction mixture. Appropriate negative and positive controls together with kinetic analysis lead to the definition of PrnD reaction mechanism were conversion of arylamine into arylnitro compounds proceeds with two monooxygenation steps and one dehydrogenation step via hydroxylamine and nitroso compound as intermediates.

# Chapter 2

# Expression, Purification and Structure of PyrH

## 2.1 Summary

PyrH, a member of the flavin dependent halogenase family, regioselectively chlorinates tryptophan at the C-5 position of the indole ring (See Fig. 2.1). The protein is involved in the biosynthesis of the pyrroindomycin B antibiotic in *Streptomyces rugosporus* [74]. The PyrH gene was overexpressed in *Pseudomonas* and PyrH purified to homogeneity. The pure protein has been crystallized and a dataset at 2.4Å collected. The enzyme structure reveals the same FAD binding module as PrnA but differences in the substrate binding modules. The comparison of PyrH and PrnA structure gives insight into the mechanism that controls regioselectivity inside the tryptophan halogenase family.




**Tryptophan** **5-Cl-Tryptophan**

**Figure 2.1**

## 2.2 FADH dependent tryptophan halogenases

Among the numerous FADH dependent tryptophan halogenases identified so far [4], *in vitro* activity has been confirmed only for the tryptophan 7-halogenase from the pyrrolnitrin biosynthetic pathway PrnA [40]; the 7-halogenase from rebeccamycin biosynthetic pathway RebH [44]; the 5-halogenase from pyrroindomycin B biosynthetic pathway and the 6-halogenase ThaL from the thienodolin biosynthetic pathway [4]. The tryptophan halogenases similarity and homology values compared to PrnA are shown in Table 2.1.

**Table 2.1**

| To PrnA | Similarity (%) | Identity (%) |
|---|---|---|
| RebH | 66 | 52 |
| PyrH | 50 | 36 |
| ThaL | 67 | 53 |

Similarity and homology values from the sequence aligment in Fig. 2.2

The sequence alignment of these proteins is shown in Figure 2.2.

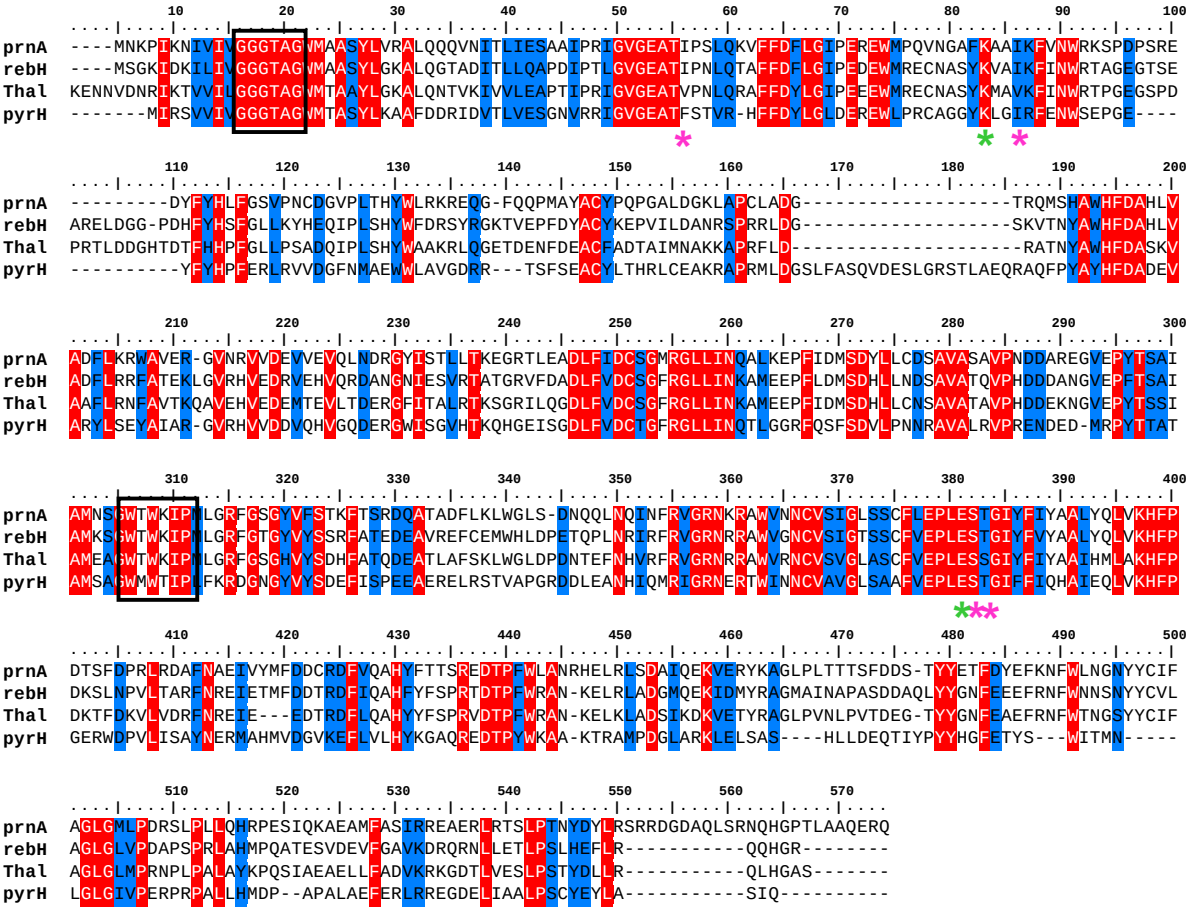



**Figure 2.2**
Sequence alignment of PrnA from *P. fluorescens* BL915[40]; RebH from *Lechevalieria aerocolonigenes*[45]; ThaL from *Streptomyces albogriseolus*[4] and PyrH from *Streptomyces rugosporus*. Highlighted in black the two halogenase conserved motifs. Highlighted by a star are aa essential for PrnA reaction mechanism: in purple aa constituting the HOCl channeling tunnel; in green E347 and K79. Shaded in red identical residues, in blue similar ones (threshold 100%).

Due to the high homology between the proteins and the proposed reaction mechanism it has been suggested that the regioselectivity of the tryptophan halogenating family is regulated by the way the substrate (tryptophan) binds into the active site such that the indole ring position to be halogenated will face the hypochlorite [4; 15].

## 2.3 PyrH a FADH dependent tryptophan 5 halogenase

PyrH is a regioselective tryptophan 5 halogenase involved in the chlorination of the pyrroloindole moiety of the pyrroindomycin B (See Fig. 2.4) antibiotic in *Streptomyces rugosporus* [74].

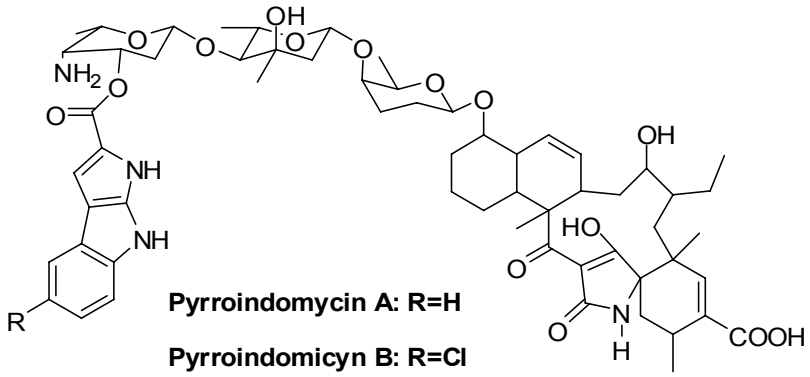



**Figure 2.4**
The pyrroindomycin are composed of a pyrroindole entity linked via an ester bond to an unbranched deoxytrisaccharide. A polyketide macro-ring system containing a is connected to the other end of the trisaccharide through a glycosidic linkage. Whereas pyrroindomycin B contain a cholrine ato, pyrroindomycin is the nonhalogenated derivative.

The pyrroindomycins (See Fig 2.4) were first isolated in 1994 from fermentations of culture LL-42D005, a strain of *Streptomyces rugosporus*. Pyrroindomycins posses potent antimicrobial activities against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci*, and are the first natural products that contain the highly unsaturated pyrroloindole moiety [75]. Tryptophan is presumed to be the precursor for the pyrroloindole moiety [76], and by analogy to pyrrolnitrin and rebeccamycin its halogenation should take place as the first step of the pyrroloindole synthesis. Starting with this assumption Zhener *et al.* were able to identify the gene of this tryptophan 5-halogenase by PCR using primers derived from conserved regions of the tryptophan halogenase genes already identified [74]. The gene named *pyrH*, was cloned and expressed. It showed 5 chloro and 5 bromo halogenase activity both *in vivo* and *in vitro*. PyrH involvement in pyrroindomycin B biosynthesis was proved by gene disruption. A *Streptomyces rugosporus* ∆*pyrH* mutant was shown to produce

pyrroindomycin A the non chlorinated pyrroindomycin, but failed to produce the chlorinated one, pyrroindomycin B [74].

## 2.4 Aim of our study

In order to test the reaction mechanism proposed for the tryptophan FAD dependent halogenase family we decide to pursue the crystal structure of PyrH together with its substrate and/or product. The structure of PyrH is expected to be highly similar to that of PrnA. If the proposed PrnA reaction mechanism is correct we would expect the localization of important residues involved in the catalytic mechanism such as those that define the hypohalous acid tunnel, base E346 and the lysine K79 to be conserved. While differences in the tryptophan binding region should account for the different substrate orientation responsible for the enzyme regioselectivity. To achieve this goal crystallography quality pure PyrH protein had to be produced and crystallization condition obtained.

# 2.5 Materials and Methods

## 2.5.1 Strains used in this study

**Table 2.2**

| N# | Strain | Relevant genotype | Organism / Use | Supplier/ References |
|---|---|---|---|---|
| **1** | BL915 ΔORF1 | PRN$^-$ (ΔprnA) *km*$^R$ | *Pseudomonas fluorescens* / Expression | 20; 70 |
| **2** | TOP10 | *F- mcrA D(mrr-hsdRMS-mcrBC) f80lacZDM15 DlacX74 deoR recA1 araD139 D(ara-leu)7697 galU galK rpsL (Str$^R$) endA1 nupG* | *Escherichia coli* / Cloning | Invitrogen |
| **3** | S17.1 | *Tp$^R$ Sm$^R$ recA, thi, pro, hsdR$^-$ M$^+$RP4: 2-Tc:Mu: Km Tn7 λ pir* | *Escherichia coli* / Conjugation | Biomedal |
| **4** | BL915 ΔORF1-4 | PRN$^-$ (ΔprnA, ΔprnB, ΔprnC, ΔprnD) *km*$^R$ | *Pseudomonas fluorescens* / Expression | 20; 70 |

## 2.5.2 Plasmids

**Table 2.3**

| N# | Plasmid | Antibiotic Resistance / Host | Type of Vector / Characteristic | Supplier/ References |
|---|---|---|---|---|
| **1** | pCIB-HIS-*pyrH* | Tetracycline / *E.coli*; *Pseudomonas sp.* | HIS tag expression, *Ptac* promoter<br>Broad host range gram -/+ vector (~22Kbp) | 74 |
| **2** | pCIB-HIS (~21Kbp) | Tetracycline / *E.coli*; *Pseudomonas sp.* | HIS tag expression, *Ptac* promoter<br>Broad host range gram -/+ vector (~21Kbp) | 20; 21; 74; 77; 78 |
| **3** | pFastBac HT A | Ampicillin, / *E.coli* | cloning for Bacmid generation; high copy number ideal for subcloning<br>(4856bp) | Invitrogen |
| **4** | pLAH31 | Tetracycline / *E.coli*; *Pseudomonas sp.* | HIS tag expression, *T5* promoter<br>Broad host range gram -/+ vector<br>(~21Kbp) | 79; 80; 81 |

### 2.5.3 Bacterial Growth and Treatments

#### *2.5.3.1 Media*

Luria-Bertani-Medium (LB) for *E. coli*:

LB-Medium: 10g/l Tryptone
5g/l Yeast Extract
5g/l NaCl

LB-Agar: LB-medium
Agar 15g/l

dissolve in $H_2O$, pH 7,0 with 0,1M NaOH, autoclave

*E. coli* cells were grown at 37˚C on plates. Liquid cultures were incubated on an orbital shaker at about 200rpm at temperatures ranging from 37 to 18˚C. Growth was monitored following the $A_{600nm}$ (Spectrophotometer: Pharmacia Biotech Ultrospect 1000).

HNB medium for *Pseudomonas*:

HNB-Medium: 3g/l Meat extract
3g/l Yeast extract
5g/l Peptone from meat
5g/l NaCl

HNB-Agar: HNB-Medium
Agar 15g/l

dissolve in $H_2O$, pH 7,0 with 0,1M NaOH, autoclave

PMM medium for *Pseudomonas*:

PMM-Agar: 8,0 g/l $K_2HPO_4$ X $3H_2O$
3,0 g/l $KH_2PO_4$
1,0 g/l $(NH_4)_2SO_4$
6.7g/l Disodium Succinate
18 g/l Agar

dissolve in $H_2O$, pH 7,0 with 0,1M NaOH, after autoclaved add 1,5ml/l of 1M $MgSO_4$

*Pseudomonas* cells were grown at 30˚C on plates. Liquid cultures were incubated on an orbital shaker at approximately 180 to 210rpm at 30˚C.

#### *2.5.3.2 Antibiotics*

The following antibiotics were used when appropriate:

| Antibiotic | Stock at | Used at |
|---|---|---|
| Ampicillin Sodium | 50 mg/ml in $H_2O$ at -20°C | 100 µg/ml |
| Kanamycin Sodium sulphate | 25 mg/ml in $H_2O$ at -20°C | 50 mg/ml |
| Spectinomycin | 50 mg/ml in $H_2O$ at -20°C | 100 µg/ml |
| Streptomycin | 50 mg/ml in $H_2O$ at -20°C | 30 µg/ml |
| Tetracycline | 5 mg/ml in EtOH at -20°C | 15µ g/ml *E. coli*<br>30 µg/ml *Pseudomonas sp*. |

#### *2.5.3.3 Glycerol Stocks*

Fresh liquid bacterial cultures were mixed with sterile 60%(v/v) glycerol to a final concentration of 30%(v/v). After 5 min incubation on ice the samples were frozen in liquid nitrogen and stored at –80 ˚C until required.

### 2.5.4 Genetic Techniques

#### *2.5.4.1 Purification of Plasmid-DNA from Pseudomonas*

Plasmid DNA purification was carried out by using both the *QIAprep Spin Miniprep Kit* by using the microcentrifuge method or the *QIAGEN Maxi Plasmid Kit*. Because of the big plasmid size that reduce the standard yield, miniprep extraction of plasmid pLAH31 and pCIBHIS from *E.coli* and *Pseudomonas* were carried out starting from 15ml of cultures [82; 83].

#### *2.5.4.2 Agarose Gel Electrophoresis*

DNA-fragments were analyzed by using 0.6 to 1.4% (w/v) agarose gels, depending on the DNA fragment size. Polymerized agarose gels contained 1x buffer TAE and 0.01% ethidium bromide. The DNA samples were mixed with 6x loading buffer, loaded onto the gel and separated at 5 V/cm in 1x buffer TAE. Afterwards gels were visualized and photographed by using a *Biorad Geldoc* apparatus.

TAE 50X: 242 g Tris Base
57.1 mL Glacial Acetic Acid
100 mL 500 mM EDTA, pH 8.0
600 mL ddH2O
Mix. Bring volume to 1 L. Autoclave

6x Loading buffer: Tris-HCl 60 mM pH 7.6
EDTA 6 mM
Bromphenole Blue 0.03 % (w/v)
Xylene Cyanole 0.03 % (w/v)
Orange G 0.03 % (w/v)
Glycerol 30 % (v/v)

### *2.5.4.3 DNA Purification from Agarose Gel and PCR Purification*

After electrophoresis DNA was visualized under UV light. Slices of agarose gel containing the DNA fragments of interest were excised with a sterile scalpel. For the gel extraction procedure the *QIAquick Gel Extraction Kit* was used. PCR products were purified using the *QIAquick PCR Purification Kit* [84].

### *2.5.4.4 Polymerase Chain Reaction – PCR*

PCR has been used to amplify the *pyrH* gene from plasmid-DNA. The reaction was performed using the *GeneAmp® PCR System 2400* thermal cycler *(Applied Biosystem)*. A standard reaction contained the following reagents:

| **Reagent** | **Amount - Stock [C]** | **Final [C]** |
|---|---|---|
| dNTPs | 1µl -10mM | 0.2mM |
| Oligos (forward and reverse) | 1µl - 100µM | 0.5µM |
| Polymerase buffer | 5µl - 10X | 1X |
| Polymerase* | 1µl - 2U/µl | 2U |
| Template DNA | 1µl – 50ngr/µl | 50ngr |
| $H_20$ | Up to 50µl | - |

| **PROGRAM** | |
|---|---|
| 98°C X 3'00'' | |
| 98°C X 45'' | 30 cycles |
| 47°C X 45'' | |
| 72°C X 2'10'' | |
| 72°C X 10'00'' | |

***OLIGOS:***

| | | |
|---|---|---|
| 5'-pyrH-TEV-BamHI | *pyrH* amplification; red BamHI site, blue HindIII site, yellow TEV recognition sequence, in bold ATG and STOP codons | ATTCGTAGCGGATCCAGAAAACCTGTATTTTCAGGGC**ATG**ATCCGATCTGTGGTGATCGTGGGTGGTGGC |
| 3'-pyrH-TEV-HindIII | | GTATGCTACC**AAGCTT**CTATCATTGGATGCTGGCGAGGTACTCGTAG |

Reagent concentrations and procedures were chosen as recommended by the manufacturers of *Vent Polymerase (NEW ENGLANDS BIOLAB Cat. No.M0254L)* chosen because of the high fidelity due to its 3'-5' exonuclease activity (Proof reading).

#### *2.5.4.5 DNA Sequencing*

The pCIB-HIS-TEV-*pyrH* plasmid DNA was sequenced at the University of Dundee Sequencing Service and analyzed with the *Contig Express feature of The Vector sequence analysis software (InforMax).* Plasmid DNA and DNA primers were sent to the Sequencing Service according to the Unit Standard Procedures.

| Name | Description | Sequence |
|---|---|---|
| 5'-pyrH-seq-Forward (1) | pyrH sequencing; forward from 571 to 591bp from 2nd ATG, reverse from 872 to 852 from 2nd ATG | TGGTCGACGACGTGCAGCAC |
| 3'-pyrH-seq-Reverse (1) | | TTGCCGTCGCGCTTGAACAG |
| 5'-pFastBac HT A-For | Sequencing of DNA cloned into pFastBac HT A polylinker (For 3995-4015; Rev 4319-4301) | TATTCCGGATTATTCATACC |
| 3'-pFastBac HT A-Rev | | GTTTCAGGTTCAGGGGGAG |

#### *2.5.4.6 DNA Restriction / DNA Ligation*

Restriction endonucleases were purchased from *NEW ENGLANDS BIOLAB*. The digestion mix contained 1x restriction buffer (as recommended by the manufacturer) 10-20 U/µg restriction enzyme(s) and the DNA to be digested**.** Plasmid DNA was incubated at 37 ˚C for 1 hour, PCR products from 2 to 4 hours. For double digestions the best buffer was chosen according to manufacturer's catalog. DNA ligations were performed using T4 DNA ligase *(ROCHE Cat. No. 0799009).*

#### *2.5.4.7 Transformation of Pseudomonas by Conjugation*

pCIB-HIS-TEV-*pyrH* and pLAH31-*pyrH* plasmids were transformed into *E. coli* S17.1 (Table 2.2 N#3). This strain has chromosomally integrated conjugal transfer functions (RP4 transfer functions) and is able to recognize the origin of transfer (*oriT* of RK2 plasmid) contained in the pLAH31 and pCIB-HIS plasmids. *E. coli* S17.1 was then used to transfer the plasmid DNA to *Pseudomonas fluorescens* BL915 strain ΔORF1-4 (Table 2.2 N# 4), (Conjugation). For conjugation 5-10ml of overnight liquid cultures of *Pseudomonas* and *E. coli* S17.1 carrying the appropriate plasmid were grown.

The cultures were harvested at room temperature by centrifugation and the cell pellets washed twice with saline solution. Both pellets were resuspended in 50μl sterile HNB, the two solutions mixed, plated on a PMM-agar plate and incubated O/N at 30°C for mating. The next day all the cells were scraped from the plate with a sterile loop and resuspended in 500μl of saline solution. Serial dilutions of the suspension (1:10 to 1:10000) were spread on PMM-agar plates containing the required antibiotics for the *Pseudomonas* strain and the plasmid. The plates were incubated at 30°C 2 to 4 days before the exconjugant bacteria were visible (1mm diameter colonies). Single colonies were picked and spread onto HNB plates with the required antibiotics for the *Pseudomonas* strain and the plasmid of interest plus 100μg/ml of Ampicillin. The plates were incubated one night at 30°C. Because *Pseudomonas* is resistant to Ampicillin; Ampicillin can be used to remove *E. coli* cells still present on the plates. This last step was repeated at least once to be sure to isolate only *Pseudomonas* bacteria. *Pseudomonas* exconjugant clones were screened by plasmid DNA extraction and restriction digestion.

### 2.5.5 Protein Biochemistry

#### *2.5.5.1 Recombinant Protein Expression in Pseudomonas*

Overnight cultures of *P. fluorescens* carrying the expression plasmids were diluted 1:20 in fresh medium containing the appropriate antibiotics. Cells were incubated for 2 days at 30˚C on an orbital shaker at 30˚C/200rpm. After centrifugation at 13000xg (25min, 4 ˚C; *Beckman Coulter™ Avanti™ J-20 XP*), the cell pellets were stored at –20 ˚C until required. For large preparation a 10 liters fermentor was utilized using the same growth condition.

#### *2.5.5.2 Purification of His-Tagged Fusion Proteins from Pseudomonas Using $Ni^{2+}$ Chelating Matrix*

Bacterial cell pellets were completely resuspended in 2ml/g binding buffer containing *Complete EDTA free Protease Inhibitor Cocktail Tablets (ROCHE Cat. No.1873580)*. After adding DNAse (final conc. 5μg/ml) and Lysozyme (final conc. 1mg/ml) the cell suspensions were incubated 20min at room temperature before being sonicated or passed twice through a Cell Distruptor, *The Basic Z Constant System Ltd.*, to achieve

complete cell disruption. The lysate was centrifuged at 48000xg (40min, 4˚C; *Beckman Coulter™ Avanti™ J-20 XP*) to pellet insoluble fractions and debris. Before loading onto the $Ni^{2+}$-column, the soluble fractions were syringe filtered (w/0.22μm). The column *HisTrap HP™ 5ml (Amersham Biotech Pharmacia)* connected to a *P1 peristaltic pump* (*Amersham Biotech Pharmacia)* was regenerated before use as follows: washed with 20ml water, stripped with 20ml 50mM EDTA and after a further wash step with 20ml water, loaded with 10ml 400mM $NiSO_4$ and washed again. After equilibration with 15ml of binding buffer, the protein solution was applied to the column and the flow through retained for analysis. The column was washed with washing buffer until all the unbound proteins were removed from the column. The protein content of the flow through was monitored with Bradford assay. Bound proteins were eluted applying the elution buffer containing 250mM imidazole. The protein content of eluted fractions, collected by a *FRAC100 fraction collector (Amersham Biotech Pharmacia)*, was checked by Bradford assay and analyzed by SDS PAGE. The fractions, which contained the purified protein were concentrated *(VIVASCIENCE VivaSpin Concentrators)* by centrifugation (3000 rpm, 4 ˚C) and afterwards dialyzed against an appropriate buffer to remove imidazole.
The buffers used are detailed in Table 2.4.

**Table 2.4**

| **Binding Buffer:** | **Washing Buffer:** | **Elution Buffer:** |
|---|---|---|
| $KH_2PO_4$ 20mM pH7.4 | $KH_2PO_4$ 20mM pH7.4 | $KH_2PO_4$ 20mM pH7.4 |
| NaCl 0.1M | NaCl 0.1M | NaCl 0.1M |
| Imidazole 20mM | Imidazole 35mM | Imidazole 250mM |

#### *2.5.5.3 Anion Exchange Chromatography (HiPrep™ 16/10 Q FF)*

*BIOCAD 700E Perfusion Chromatography Workstation (Applied Biosystem)* was used at room temperature connected to an *Advantec SF 2120 (Advantec)* automatic fraction collector. After loading the protein on a *HiPrep™ 16/10 Q FF (Amersham Biotech Pharmacia)* it was eluted against an increasing linear salt gradient. The protein concentration was monitored by measuring $A_{280}$ and protein containing fractions were analyzed by SDS PAGE.

#### *2.5.5.4 Gel Filtration Chromatography (HiLoad™16/60 Superdex™200, Superose 12 HR10/30)*

*BIOCAD 700E Perfusion Chromatography Workstation (Applied Biosystem)* was utilized at room temperature connected to an *Advantec SF 2120 (Advantec)* automatic fraction collector. A *HiLoad™ 16/60 Superdex™ 200 (Amersham Biotech Pharmacia)* gel filtration column was used for preparative purpose while a *Superose 12 HR 10/30 (Amersham Biotech Pharmacia)* gel filtration column was used to determine the oligomeric state of the protein. The columns were equilibrated with approximately 2CV of buffer and then the protein solution was applied to the column. After 0,2CV 1ml fractions were collected for 1CV. The protein conc. was monitored by measuring $A_{280}$. Fractions corresponding to protein peak(s) were analyzed by SDS-PAGE, pooled and concentrated.

#### *2.5.5.5 Determination of Protein Concentration*

The protein concentration was calculated by the measurement of the $A_{280}$ *(Spectrophotometer: Pharmacia Biotech Ultrospect 1000).* Depending on the concentration of the protein solution a dilution was made, the $A_{280}$ was measured three times and the average was used in the following calculation:

$[(FV/P) \times A_{280}] / EC$ = concentration in mg/ml

[concentration in mg/ml] / $M_W$ = concentration in mM

FV = final volume in µl

P = volume of protein solution in µl

EC = theoretical extinction coefficient of the protein [*]

$M_W$ = molecular weight of the protein

[*] obtained from http://us.expasy.org/cgi-bin/protparam by entering the amino acid sequence of the protein

Alternatively the Bradford method was used**.** This protein assay is based on complexing of proteins with Brilliant Blue G-250. The protein sample was mixed with the reagent and then read at 595nm after a short incubation at room temperature. The protein concentration was calculated by determination of the $A_{595}$ *(Spectrophotometer: Pharmacia Biotech Ultrospect 1000).* Depending on the concentration of the protein solution a dilution was made and the $A_{595}$ was measured three times. The average result was used to extrapolate the protein concentration on a calibration curve obtained by measuring the $A_{595}$ of different dilution of a BSA standard.

*Bradford Reagent (Cat. No. B6916),SIGMA*

#### *2.5.5.6 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS PAGE)*

For separation of proteins under denaturating condition Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis was used. Precast gels were obtained from *INVITROGEN NuPage™ 4-12 % Bis-Tris Gels, 1mm x 10, 12 or 17 wells* and run in 1x *MES* buffer *Invitrogen NuPage™ MES SDS Running Buffer (20x)* or *NuPage™ 7 % Tris-Acetate Gels 1mm x 10* and run in 1X *TRIS-ACETATE* buffer *Invitrogen NuPage™ TRIS-ACETATE SDS Running Buffer (20x)*. The samples were mixed with sample buffer *(Invitrogen NuPage™ LDS 4x sample buffer)*, denaturated for 5-10 min at 100˚C and then applied to the gel. Protein samples loaded onto MES gels were run at 200 V for 35 min, protein samples loaded onto TRIS-ACETATE gels were run at 150 V for 60 min [85].

#### *2.5.5.7 Coomassie Staining*

SDS-PAGE gels were stained ca. 10min in Coomassie blue staining solution with slight shaking. Following staining, the gels were incubated in destaining solution (See Table 2.5). After the background was destained the gels were washed in distilled water for 10 min and dried with the *DryEASE Kit (INVITROGEN)*.

*DryEASE Mini Gel Drying System (Version E), INVITROGEN*

**Table 2.5**

| Coomassie blue staining solution: | Destaining solution: |
|---|---|
| 50 % Methanol (v/v) | 45 % Methanol (v/v) |
| 10 % Acetic Acid (v/v) | 10 % Acetic Acid (v/v) |
| 0.5 % Brilliant Blue R (w/v) | |

### *2.5.5.8 TEV purification and TEV protease digestion*

The recombinant histidine tagged 27Kda catalytic domain of the Nuclear Inclusion a (NIa) protein encoded by the tobacco etch virus (commonly called TEV protease) was purified from *E. coli* BL21 codon plus (STRATAGENE) cells containing plasmid pRK793-HIS-TEV[S219V]Arg5 (David S. Waugh, Macromolecular Crystallography Laboratory NCI at Frederick). O/N *E. coli* culture in LB containing 100µg/ml Ampicillin and 30µg/ml Chloramphenicol were diluted 1 to 30 in fresh medium. Bacteria were grown till mid log phase ($OD_{600nm}$ ~ 0.7) at 37°C 200rpm on orbital shaker incubator. Protein expression was induced by adding IPTG to a final concentration of 1mM and the temperature was reduced to 30°C. After 4hrs of induction, cells were harvested by centrifugation. TEV was purified on Nickel chelating column as described in Paragraph 2.5.5.2. Immediately after the elution step all the protein fraction were diluted four times in 50mM $NaHPO_4$, 150mM NaCl 1mM EDTA pH 8.0 buffer and kept on ice. Fractions containing TEV protease were pooled; protein was concentrated at 2mg/ml; 10% glycerol added to the buffer; protein was snap frozen in liquid $N_2$ and stored at -80°C

TEV protease digestion was carried out in 50mM $NaH_2PO_4$, 50mM Tris-HCl pH 7,5, 400mM NaCl, 0,5mM EDTA with or without 1mM DTT. Usually 1µg of TEV was used for every 1mg of protein to be digested.

#### *2.5.5.9 PyrH assay*

The PyrH assay was performed as described by Zehner *et al.* [74]; the reaction mix contained:

| Reagent | Amount - Stock [C] | Final [C] |
|---|---|---|
| ***E.coli* Fre** | **0.84 Ul/ml*** | **0.42mU** |
| **PyrH** | **10µl - 175µM** | **1.75µM** |
| **NADH (Sigma)** | **20mM - 125µl** | **2.5mM** |
| **FAD (Sigma)** | **2mM - 5µl** | **10µM** |
| **$MgCl_2$** | **50µM – 250mM** | **12.5mM** |
| **D/L-TRP** | **100µl – 6mM** | **0.6mM** |
| **Buffer (20mM $KH_2PO_4$ pH 7.2)** | **Up to 1ml** | |

The reaction was carried out at 30˚C. To stop the reaction samples were boiled (100°C) for 5min. Precipitated protein were separated by centrifugation (13000rpm RT) on a microcentrifuge. Soluble fraction samples (20µl) were analyzed by isocratic HPLC, Methanol:$H_2O$ 60:40, using a Varian OmniSpher C-18 250 * 4.6 mm column. Flow rate was 1ml/min run time 30min and detected by $A_{220nm}$. The retention time of the product 5-Cl-TRP was comparable with previously described data by Zehner [74]. *E. coli* Fre (ferredoxin reductase) was kindly provided by Khim Leang of our group. The column was calibrated using pure 5-DL-Cl-TRP (AT N# 32606C from Apin Chemicals Ltd., UK).

## 2.6 Results

### 2.6.1 PyrH Expression Constructs

*Pseudomonas fluorescens* BL915 strain ΔORF1 [20] carrying the pCIB-HIS-*pyrH* [74] plasmid was provided by our collaborator Professor K.H. van Peé from the Institut für Biochemie at the Technische Universität Dresden, Germany. This DNA codes for N-terminal tagged PyrH protein (See Fig. 2.4). The protein tag consisted of a classical hexahistidine signature, to allow easy purification by $Ni^{2+}$ chelating resin, followed by a linker region and a protease recognition site, Enterokinase (EK) light chain recognition site that can be recognized by the commercially available antibody, Anti-Xpress Antibody, (Invitrogen), (See Fig. 2.4).

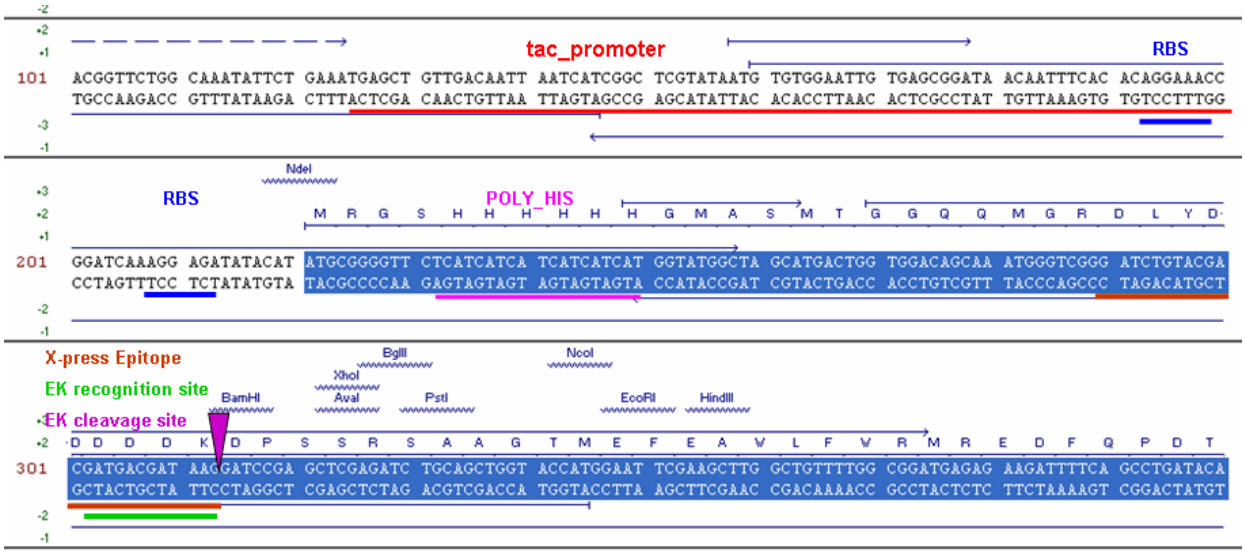



**Figure 2.4**
pCIB-HIS promoter and polylinker region. Underlined: in red the tac promoter sequence; in blue the two RBS present in the promoter region; light purple the hexahistidine sequence; in brown the anti X-press epitope; in green the EK recognition site and dark purple the EK cleavage site.

Given preliminary unsuccessful crystal trial results with PyrH purified from this strain (See Paragraph 2.6.5) concerns about possible detrimental effects exerted by the 32aa long tag, and cleaving problems with the EK protease that prevent its use in an effective way (Naismith lab unpublished results) we decided to reclone the *pyrH* gene. Two different strategies were chosen: firstly recloning *pyrH* into pCIB-HIS with a TEV protease recognition site immediately before the first Met; secondly using a different plasmid with a

shorter N-terminal hexahistidine tag. To accomplish the first objective *pyrH* was re-amplified by PCR from pCIB-HIS-*pyrH* plasmid using a set of primers that carry the TEV protease recognition site. The PCR product was then ligated into plasmid pFastBac-HT-A through BamHI and HindIII restriction site. *E. coli* TOP10, cloning strain, cells were transformed with 5μl of the reaction mix. Plasmid DNA was extracted from Ampicillin resistant clones. Once the TEV-*pyrH* gene was proven to be correct by sequencing the TEV-*pyrH* gene was cloned into pCIB-HIS vector for *Pseudomonas* overexpression. The second approach consisted in choosing a different expression vector with a shorter N-terminal hexahistidine tag. Iris Bertani and Vittorio Venturi from the ICGEB Bacteriology Group, Italy sent us their pLAH31 vector [79] (See Fig. 2.5).

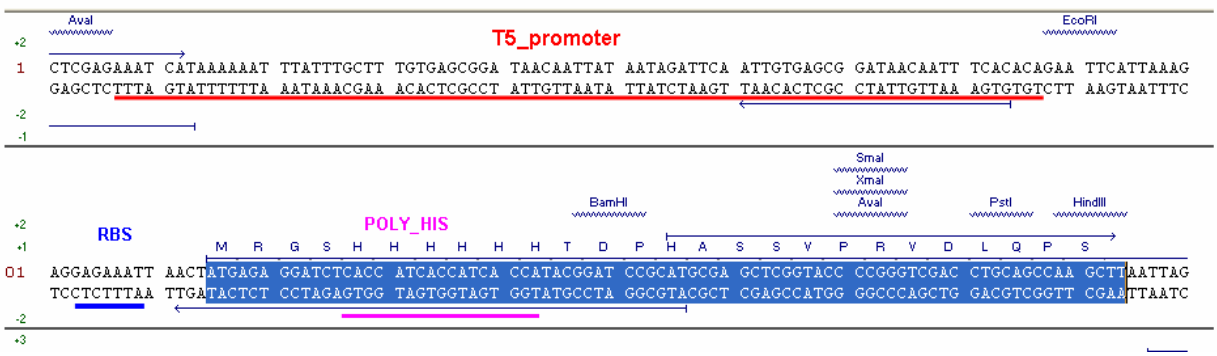



**Figure 2.5**
pLAH31 promoter and polylinker region. Underlined: in red the T5 promoter sequence; in blue the RBS; light purple the hexahistidine sequence.

The *pyrH* gene was excised from pCIB-HIS-*pyrH* plasmid with a BamHI HindIII double digestion and reintroduced into the pLAH31 plasmid using the same restriction sites preserving the reading frame. *E. coli* TOP10 (cloning strain) cells were transformed with 5μl of the ligation mix. Plasmid DNA was extracted from Tetracycline resistant clones. The insert presence was checked by restriction digestion and DNA gel electrophoresis before introducing the plasmid DNA into *Pseudomonas* by conjugative mating (Fig. 2.6 shows DNA gel examples).

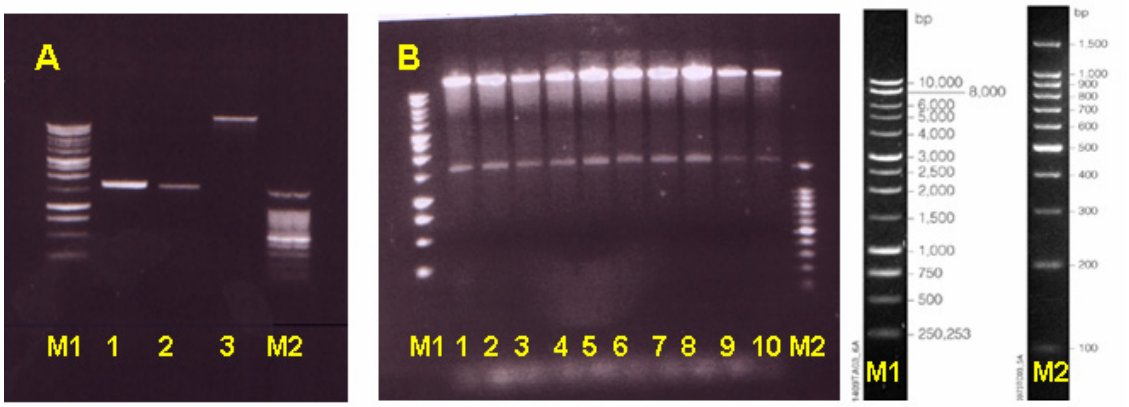



**Figure 2.6**
DNA gel examples. A: M1, DNA ladder (*1kbp DNA Ladder, Promega, Cat. No. G5711)*; 1, TEV-*pyrH* PCR amplified fragment; 2, *pyrH* extracted from pCIB-HIS-*pyrH*; 3, pLAH31 linearized BamHI HindIII; M2, DNA ladder *100bp DNA Ladder (Promega, Cat. No. G2101)*. B: M1, DNA ladder (*1kbp DNA Ladder, Promega, Cat. No. G5711)*; 1,2,3,4,5,6,7,8,9,10, pLAH31-*pyrH* plasmid digested BamHI HindIII, M2, DNA ladder (*100bp DNA Ladder, Promega, Cat. No. G2101)*.

pCIB-HIS-TEV-*pyrH* and pLAH31-*pyrH* plasmids were transformed into *E. coli* S17.1 and then by conjugation to *Pseudomonas fluorescens* BL915 strain ΔORF1-4.

### 2.6.2 PyrH Expression and Purification

PCIB-HIS-*pyrH*, pCIB-HIS-TEV-*pyrH* and pLAH31-*pyrH* plasmids contain constitutively active promoters and all gave over-expression of soluble HIS-tagged PyrH. The overexpressed proteins run on SDS PAGE between the 66.3KDa and the 55.4KDa markers consistent with the expected proteins $M_w$: pLAH-31-PyrH 59852Da; pCIB-HIS-PyrH: 62020Da; pCIB-HIS-TEV-PyrH: 62734Da. Furthermore protein identity was confirmed with peptide fingerprinting by the *St Andrews University BMS MASS SPECTROMETRY AND PROTEOMICS FACILITY* PyrH was purified by a three step procedure from fresh or frozen bacteria cell pellet, consisting of a first Nickel chelating affinity chromatography followed by an anion exchange column and a gel filtration column. The last step was used to polish the protein preparation and exchange the protein buffer to Tris-HCl, PyrH is stable in 10mM Tris-HCl pH7.2 at concentration up to 14mg/ml. Both pCIB-HIS and pLAH31 expressed soluble PyrH, but the protein yield differed greatly, due possibly to different promoter strength (*Ptac* is described as a strong constitutive promoter in *Pseudomonas* by Kirner [21] citing unpublished studies). This is

clearly evident from the two SDS PAGE of Fig. 2.7. The gels show PyrH after the first step of purification by $Ni^{2+}$ affinity chromatography.

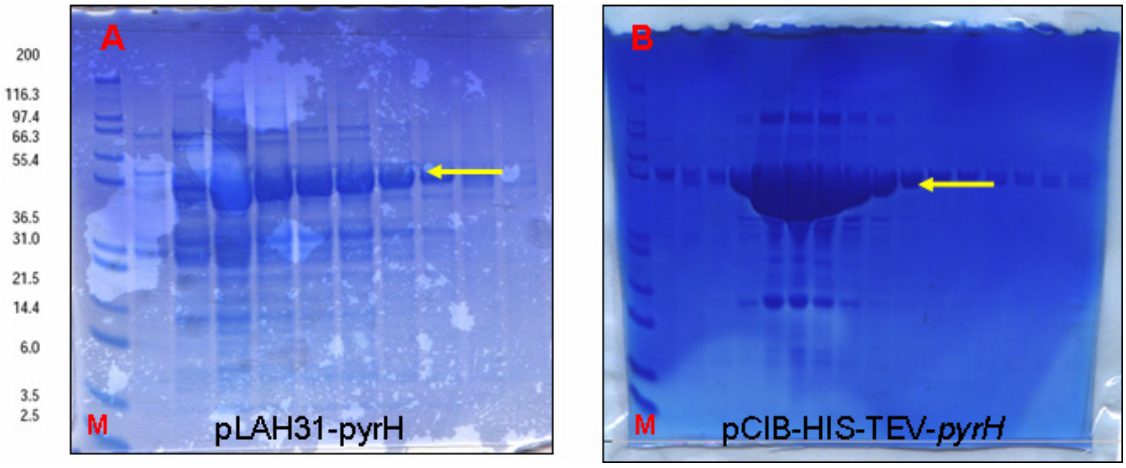



**Figure 2.7**
NuPage 4-12% Bis-Tris gels showing the fractions eluted from the first $Ni^{2+}$ chelating resin column (affinity chromatography) during the purification of PyrH from A: pLAH31 and B: pCIB-HIS. M: molecular weight marker (*Mark 12, Invitrogen, Cat. No. LC5677*) The yellow arrow indicates PyrH.

The dramatic difference is clear. A more quantitative measurement is given in Table 2.6.

**TABLE 2.6**

**Total protein amount at each purification step**

| PyrH: | pLAH31-*pyrH* | pCIB-HIS-TEV-*pyrH* |
|---|---|---|
| LYSATE ON $Ni^{2+}$ COLUMN | 900mg - **100%*** | 1200mg - **100%*** |
| OUTPUT ELUTED PYRH FRACTIONS FROM $Ni^{2+}$ COLUMN | 60mg - **6.6%*** | 195mg - **16.25%*** |
| AMOUNT COLLECTED AFTER GEL FILTRATION COLUMN | 2mg - **0.2%*** | 82,5mg - **6.8%*** |

* compared to the amount loaded on the first column

I was not able to cleave the N-terminal from the overexpressed TEV-PyrH. After incubation at room temperature with TEV protease for more than 2 days the TEV-PyrH fractions from the first $Ni^{2+}$ column were reloaded onto a $Ni^{2+}$ column. This second $Ni^{2+}$ column should remove not only the His tagged protease but also non specific binding proteins. In principle only cleaved PyrH protein would be expected to flow through while the contaminant proteins and the TEV protease would bind to resin. Only a small

fraction of PyrH was cleaved as shown in Fig. 2.8 while the majority of it bound again to the column by means of the uncleaved hexahistidine tag.

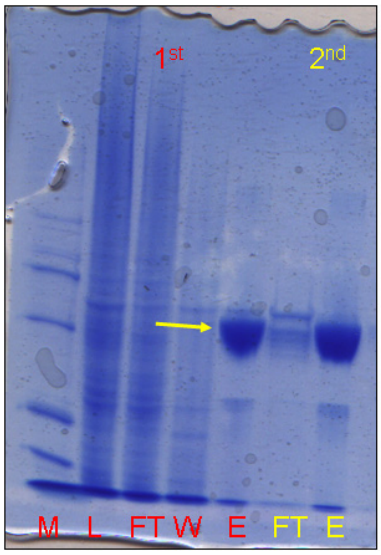



**Figure 2.8**
NuPage 7% Tris-Acetate gel showing TEV-PyrH during the first (red) and after incubation with TEV protease second (yellow) $Ni^{2+}$ chelating resin column. M: marker (Par 2.4.5.10, A); L: total lysate loaded; FT: flow-through; W: wash; E: eluted TEV-PyrH. The yellow arrow indicated PyrH.

### 2.6.3 PyrH Oligomeric State

The PyrH oligomeric state was assessed by analytical gel filtration chromatography. PyrH behaviour was consistent with the behavior of a dimeric protein. PrnA, as well, has been already described as a dimeric protein [15]. An example is shown in Fig. 2.9.

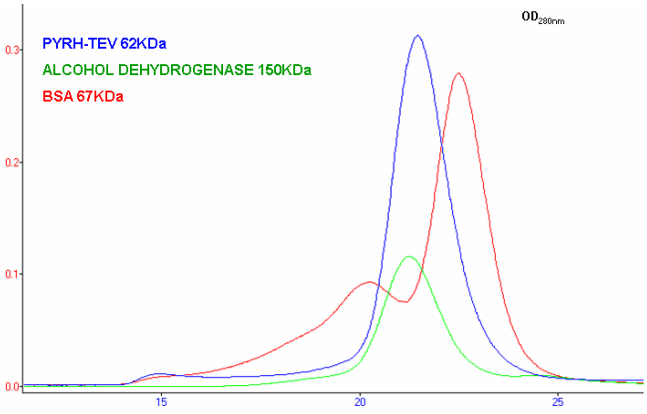



**Figure 2.9**
Analytical gel filtration column Superose 12 HR10/30® was used together with protein weight marker from Sygma. In blue PyrH trace, in red BSA and in green alcohol dehydrogenase. PyrH retention time is between BSA and alcohol dehydrogenase ones consistently with a dimeric protein with a $M_w$ of about 124KDa. (Y axes $OD_{280nm}$; X axes time in minutes).

### 2.6.4 PyrH Enzymatic activity

In order to maximize the possibility of obtaining, biochemically meaningful co-crystals we assayed the purified enzyme with D- and L-tryptophan isomers to be sure the enzyme was able to turn over both of the tryptophan enentiomers. Fig. 2.10A shows a time course experiment using D-tryptophan as substrate. PyrH appears to be active on both of the tryptophan isomers. Despite the non quantitative nature of the assay it is evident from the chromatograms peaks intensity in Fig. 2.10B that the L-isomer is a better substrate than the D-one.

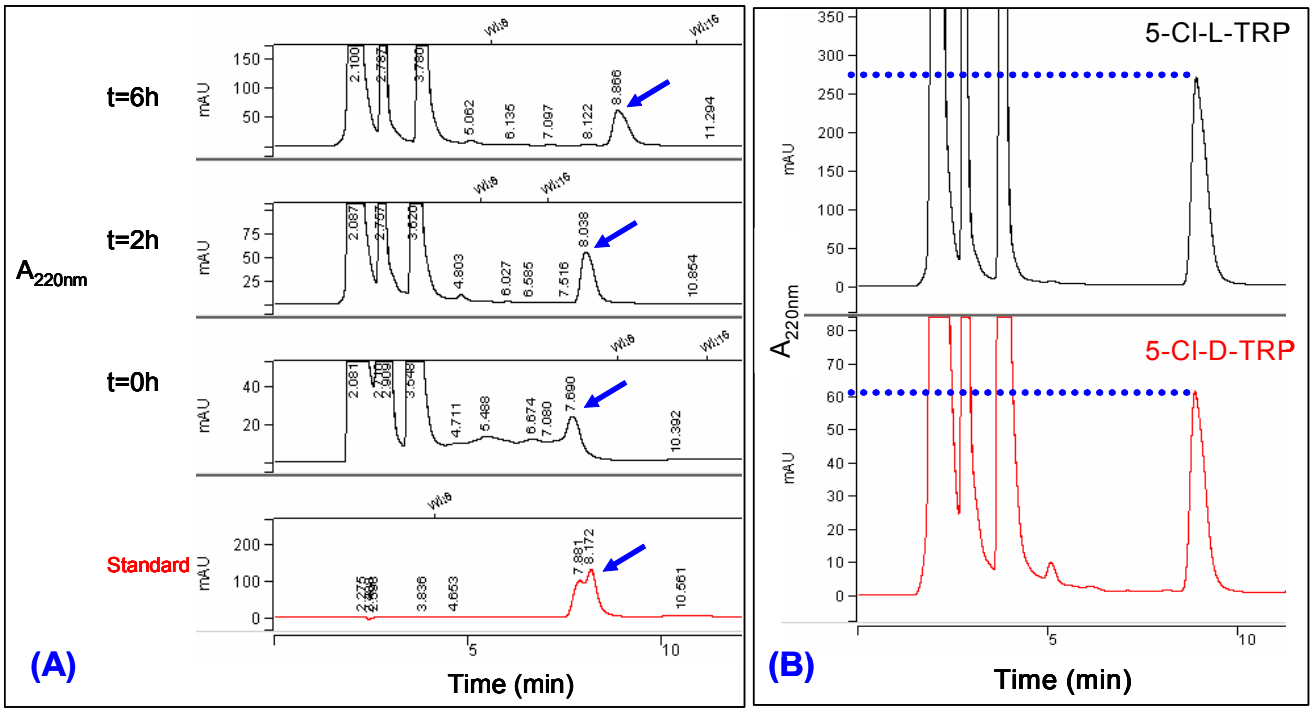



**Figure 2.10**

A: HPLC chromatograms of PyrH reactions built with D-tryptophan as substrate. The reaction was carried out as described in materials and methods. Samples were taken at t=0h, 2h, 6h. In red the chromatography trace of chemically synthesized 5-DL-Cl-tryptophan from Apin Chemicals Ltd, Abingdon, UK (Cat N# 32606c). The blue arrow indicates the 5-Cl-tryptophan peak. Small variation in the retention times are ascribed to the non optimal HPLC configuration. B: comparison of the intensity of the Cl-tryptophan peak at the 6h time point for 5-Cl-D-TRP 60mAU and 5-Cl-L-TRP 275mAU.

### 2.6.5 PyrH crystallization trials

PyrH was concentrated at 10mg/ml (See Fig. 2.11A). FAD (Sigma) was added to a final concentration of 1mM and crystal trial were set up on sitting drop plates (Hampton Research CrystalClear Strips$^{TM}$ 96 wells plates) using crystallization sparse matrix kits from a range of different providers (2µl protein solution plus 2µl of precipitant). PyrH from pCIB-HIS was crystallized using condition number 22 of the Hampton Crystal Screen 2 (0.1 M MES pH6.5, 12%w/v PEG 20K). The crystals were hollow needle shaped and fragile (See Fig 2.11B). Despite the crystal yellow color, indicating the presence of FAD, no reasonable protein diffraction pattern was obtained using our in house X-ray generator (Rigaku 007 rotating anode).

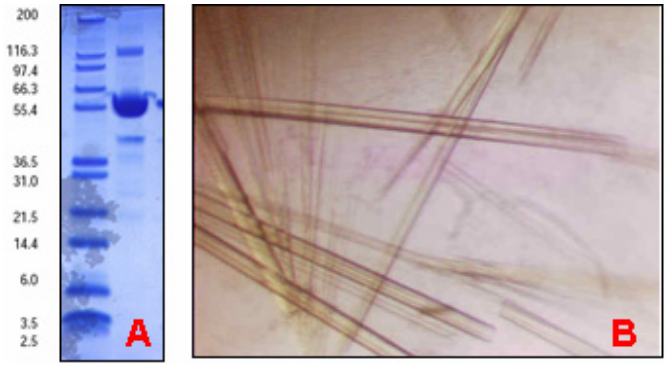



**Figure 2.11**
A: NuPage 4-12% Bis-Tris gels showing in lane 1 protein ladder; in lane 2 PyrH used for crystallization. B: PyrH crystals obtained with crystallization condition N#22 from Hampton Crystal Screen 2, Hampton Research.

Crystal trials with PyrH from pLAH31 did not give us any lead to pursue. PyrH was recloned into the pCIB-HIS vector together with a TEV protease signature at the protein N-terminal. TEV-PyrH protein carrying a 40aa long tag resisted TEV protease cleavage. Crystal trials with the TEV-PyrH protein failed to reproduce the Hampton Crystal Screen 2 number 22 condition rods. Three months after we set up a Hampton Natrix screen, we found bipyramidal crystals into the condition 11 drop (0.01M $MgCl_2$, 0.05M Na Cacodylate pH 6.0, 1.0 M $Li_2SO_4$ * $H_2O$) (See Fig 2.12A). We went back to the original plates containing PyrH with the shorter 32aa long N-terminal tag and in some of the Natrix number 11 drops we found crystals (See Fig 2.12B). Table 2.7 summarizes our findings.

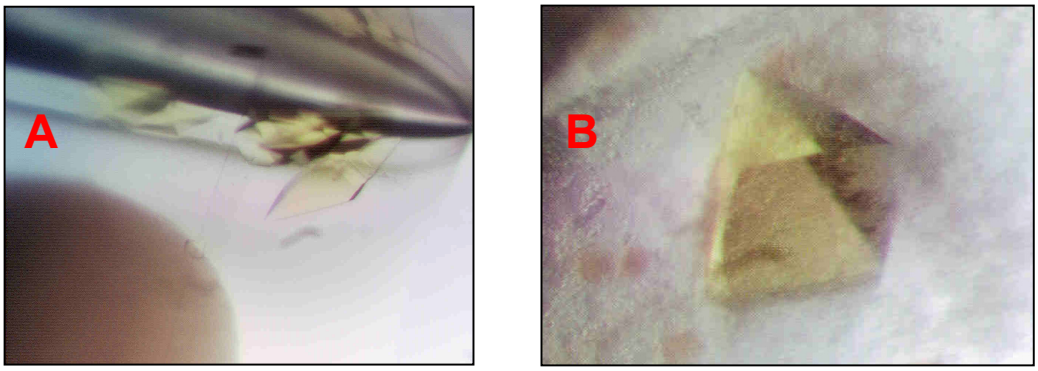



**Figure 2.12**
A: TEV-PyrH crystals obtained with crystallization condition N#11 from Natrix Screen, Hampton Research B: one of the few PyrH crystal found with crystallization condition N#11 from Natrix Screen.

**Table 2.7**

| | | Crystals | |
|---|---|---|---|
| **Construct** | **Protein feature** | **N#22 C.Screen 2** | **N#11 Natrix** |
| pLAH31-*pyrH* | 14aa N-tag | none | none |
| pCIB-HIS-*pyrH* | 32aa N-tag | needles | bypiramidal |
| pCIB-HIS-TEV-*pyrH* | 40AA N-tag<br>TEV uncleavable | none | bypiramidal |

### 2.6.6 PyrH crystal optimization

Given the starting crystallization conditions being 0.01M $MgCl_2$, 0.05M Na-Cacodylate pH 6.0, 1.0 M $Li_2SO_4$ * $H_2O$ we expanded the pH and $Li_2SO_4$ concentration to optimize the crystallization conditions.

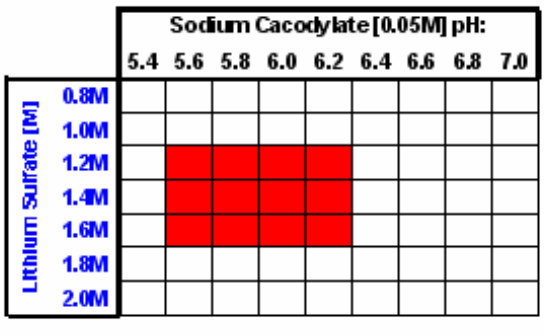



**Figure 2.13**
Optimization grid for crystallization condition N#11 from Natrix Screen, Hampton Research. [$Li_2SO_4$] was increased up to 2M and pH up to 7. Condition given better bigger crystals are highlighted in red.

Figure 2.13 shows the extended grid used to optimize TEV-PyrH crystallization conditions. Between pH 5,6 and pH 6,2 and with a $Li_2SO_4$ concentration between 1.2M and 1.6M we were able to grow crystal in less than 1 week (See Fig. 2.14)

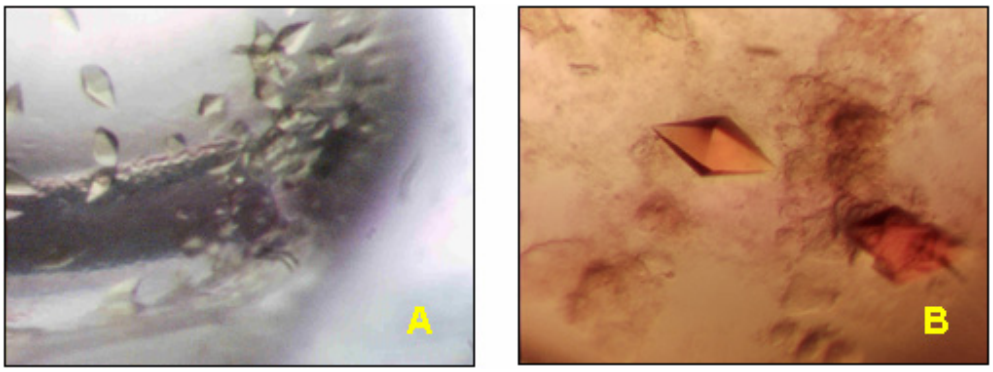



**Figure 2.14**
TEV-PyrH crystals obtained in 5 days optimizing crystallization condition N#11 from Natrix Screen

We did not obtain any improvement in crystal diffraction increasing $MgCl_2$ concentration and we were not able to obtain any crystal, after swapping Na-Cacodylate arsenic containing buffer with the less toxic buffer MES. Co-crystallization was attempted saturating the protein with substrate or the product analogue 5-DL-Br-TRP (Sigma). PyrH has been report to catalyze bromination of the tryptophan five position in presence of NaBr [74].

### 2.6.7 PyrH data collection

Our in house X-ray generator (Rigaku 007 rotating anode) was used to screen initial PyrH crystals to test their diffracting quality and determine cell dimensions and space group. A single PyrH crystal was cryoprotected with 20% glycerol added to the mother liquor and 10 minutes exposure with 0.5° oscillation produced the diffraction image in Figure 2.15 A1, with spots to 3.2 Å at the edge. A higher resolution data set was collected at The Daresbury Synchrotron Radiation Source on Station 10.1. The dataset was collected using a smaller oscillation 0.2° due to long cell dimension that gave rise at wider oscillation to too many overlaps Figure 2.15A2. A 2.4Å resolution dataset was obtained. 600 images were indexed in MOSFLM [86], which identified the crystal as primitive

tetragonal. Data reduction suggested the crystal lattice belonged to the Laue group P422. Examination of systematic absences in SCALA[87] indicated that the crystal belongs to space group $P4_32_12$ or $P4_12_12$. Data were index in $P4_32_12$, the same space group of PrnA crystals), unit-cell parameters and statistics are shown in Table 2.8.

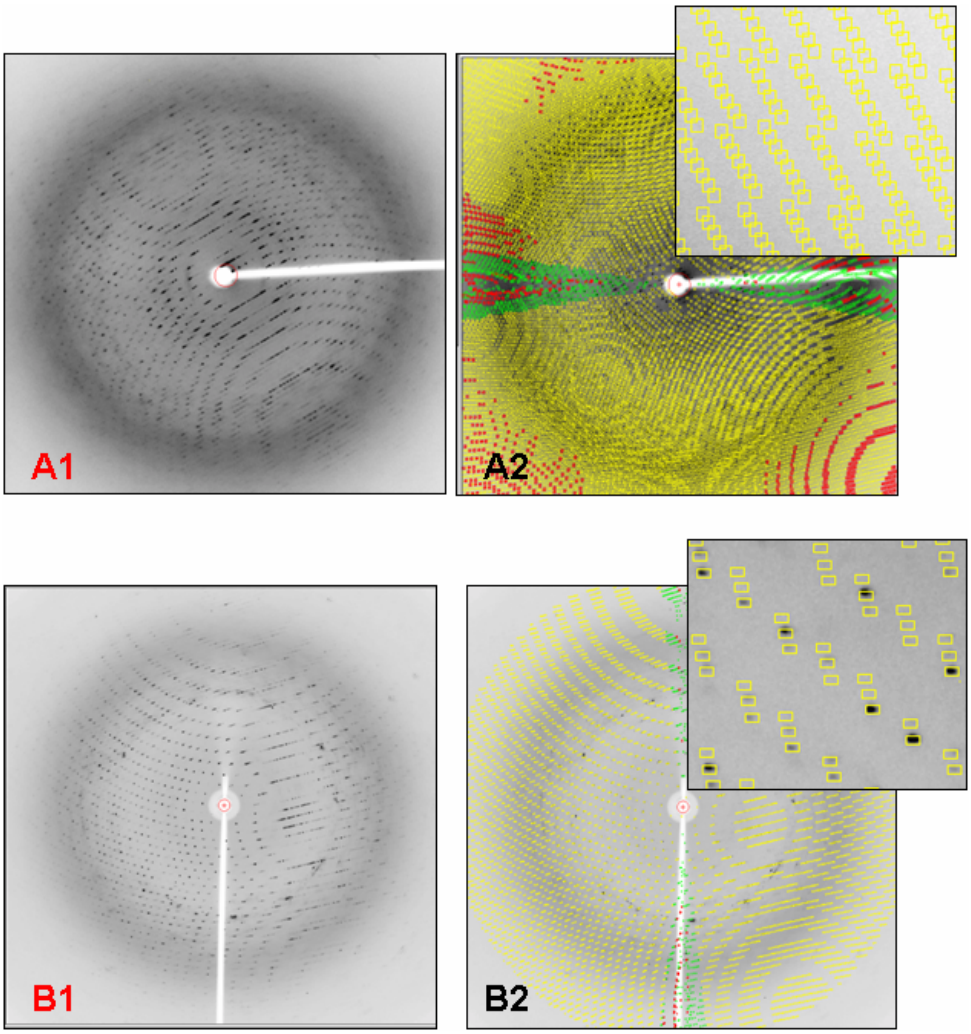



**Figure 2.15**
Differences in MOSFLM spot finding between images collected in house with 0.5°oscillation (A1 image; A2 image processed with MOSFLM with the superimposed yellow box indicating partial reflections, red overlapping ones) and data collected at the synchrotron with 0.2° oscillation (B1 image; B2 image processed with MOSFLM with the superimposed yellow box indicating partial reflections, red overlapping ones).

**Table 2.8**

| DATA COLLECTION | PyrH | |
|---|---|---|
| | **Overall** | **OuterShell** |
| Low resolution limit (Å) | 50 | 2.53 |
| High resolution limit (Å) | 2.4 | 2.4 |
| Wavelength | 0.980 Å | |
| Unit-Cell | a=b=137.6 Å, c=307.9 Å, $\alpha$=ß=$\gamma$=90.0 ° | |
| Space Group | $P4_32_12$, | |
| Total number unique reflections | 113154 | 15413 |
| Multiplicity (%) | 7.0 | 4.6 |
| Completeness (%) | 97.6 | 92.4 |
| Rmerge | 0.094 | 0.340 |
| I/$\sigma$ (Mean((I)/sd(I)) | 16.9 | 3.7 |

### 2.6.8 PyrH structure solution and refinement

The PyrH structure was solved by molecular replacement with PHASER [88; 89] using a PrnA monomer (PDB accession code 2APG) as a search unit. A solution was found for 4 monomers in the asymmetric unit cell, consistent with the Matthews coefficient calculation results [90 91] (See Table 2.9), and resolution range between 70 Å and 4.5 Å.

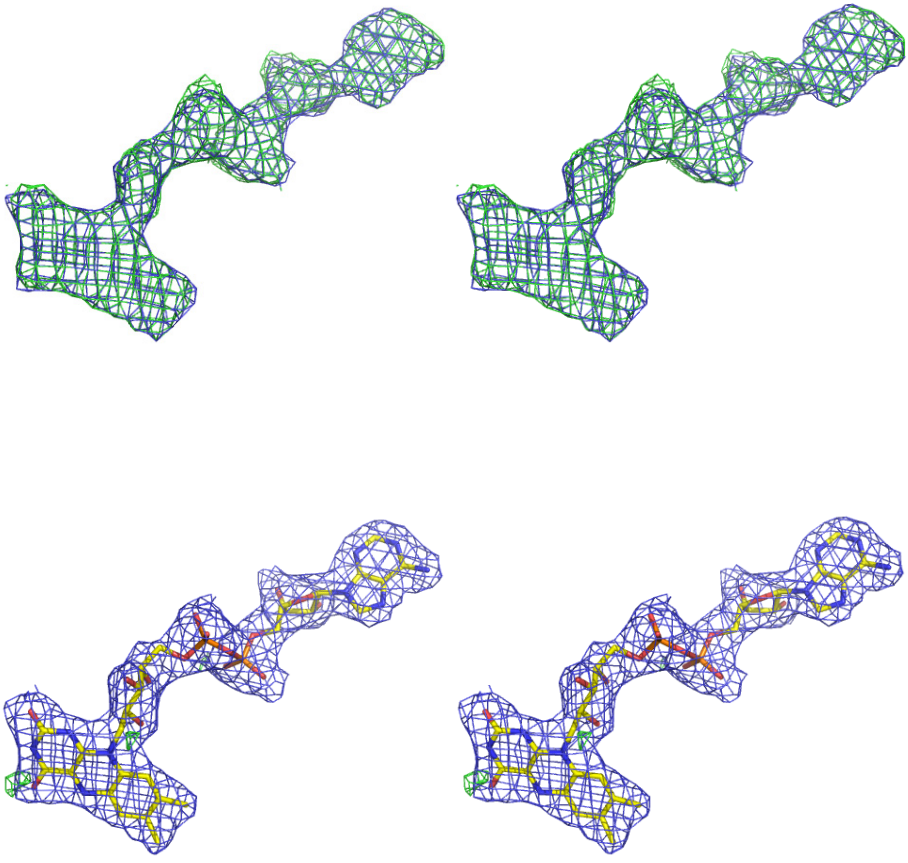
**Table 2.9**

| SPACE GROUP | CELL VOLUME Å | $N_{molecule}$ / Asymmetric Unit | Matthews Coefficient | % SOLVENT | P* |
|---|---|---|---|---|---|
| $P4_32_12$, | 5822474 | 4 | 2.90 | 57.62 | 0.25 |

**(P*: probability across all resolution ranges)**

CHAINSAW (CCP4 package) was used to help to mutate residues that differ between PrnA and PyrH in the solution model. Refinement was carried out using cycles of manual refinement with WinCoot [51] and Refmac5 [92; 93] (CCP4 package): final $R_{cryst}$ was 0.217 and

final $R_{free}$ was 0.296. The FAD electron density was present and clear in all the four monomers, an example is shown if Fig. 2.16.



**A**

**B**

**Figure 2.16**
PyrH $2F_O$-$F_C$ at 1σ in blue and $F_O$-$F_C$ at 3σ in green electron density maps around the FAD cofactor: A: PyrH monomer B unbiased density; B: PyrH monomer B with modelled FAD cofactor.

## 2.7 Discussion

### 2.7.1 PyrH Expression and Purification

A protocol for the purification on PyrH from three different constructs has been established. The differences in protein yields obtained using the two different promoters T5 and $P_{tac}$ confirmed previous unpublished results claiming good constitutive activity of the latter one [21] in *Pseudomonas* species. It has been proved that the enzyme is active on both D- and L-tryptophan expanding our chance of finding co-crystallization conditions. Interestingly we were not able to cleave the histidine tag at the protein N-terminal. This problem could be related to the reported inactivity of histidine tagged HalB halogenase respect to the untagged version [94]

### 2.7.2 Overall Structure

As expected from the high sequence homology between PrnA and PyrH the two proteins structure resulted to share the same overall fold. The PyrH asymmetric unit cell contains four monomers not related by crystallographic symmetry, see Fig. 2.17.

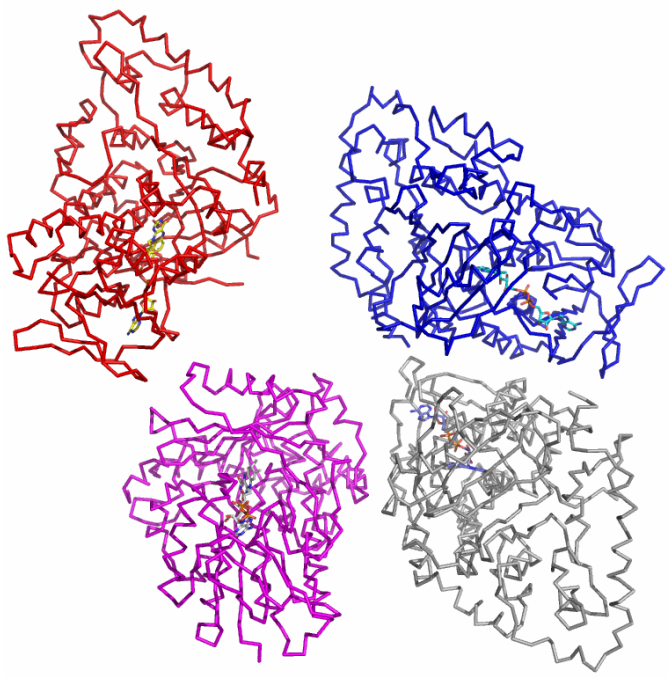


**Figure 2.17**
The four monomers of PyrH asymmetric unit. Monomer A: red carbon α trace yellow FAD. Monomer B: blue carbon α trace cyan FAD. Monomer C: purple carbon α trace gray FAD. Monomer D: gray carbon α trace blue FAD.

Monomer B is the one displaying the best electron density. No major differences are observed between the four subunits other than in small interconnecting loop regions. Structural superposition of monomer A, C and D to monomer B using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50] resulted in a Z-score of 17.2 with a RMSD of 0.60Å, a Z-score of 19.4 with a RMSD of 0.62Å and a Z-score of 21.2 with a RMSD of 0.51Å. Figure 2.18 shows the overlapped four monomer carbon $\alpha$ traces.

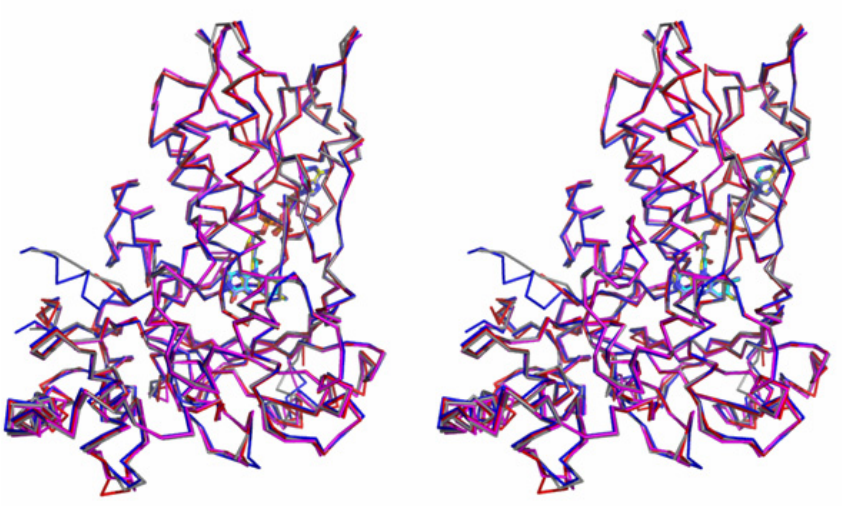


**Figure 2.18**
Ribbon representation of PyrH monomers superposed using the SSM algorithm of WinCOOT [51]. Monomer A: red carbon $\alpha$ trace yellow FAD. Monomer B: blue carbon $\alpha$ trace cyan FAD. Monomer C: purple carbon $\alpha$ trace gray FAD. Monomer D: gray carbon $\alpha$ trace blue FAD.

Analysis of protein protein interfaces in the crystal using the web based PISA Server [95] software (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) help to identify the two PyrH dimers that constitute the asymmetric unit cell consistently with gel filtration results shown in Paragraph 2.6.3 . The first PyrH dimer if formed by momers A and B the second by chain C and D. The PISA software calculated an interface area buried between the two monomers of 1700Å$^2$ for the AB dimer and 1827Å$^2$ for the CD dimer. Figure 2.19 shows the PyrH dimer AB. The interface between the two monomers is defined mainly by the substrate binding domain together with helix $\alpha$12. As expected the PyrH dimer superpose with PrnA dimer with in a Z-score of 20 and a RMSD of 1.66Å for 992 aligned residues (See Fig. 2.20).

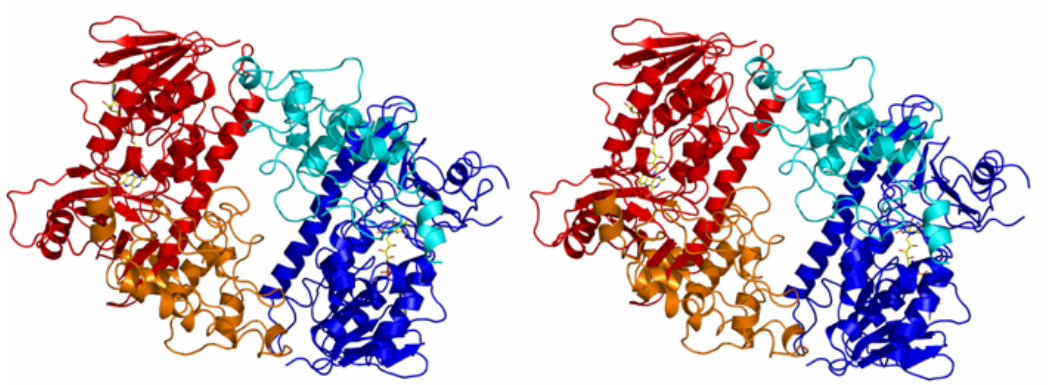


**Figure 2.19**
Cartoon representation of PyrH AB dimer. Monomer A: Flavin binding domain in blue, tryptophan binding domain in cyan. Monomer B: Flavin binding domain in red, tryptophan binding domain in orange. FAD in yellow.

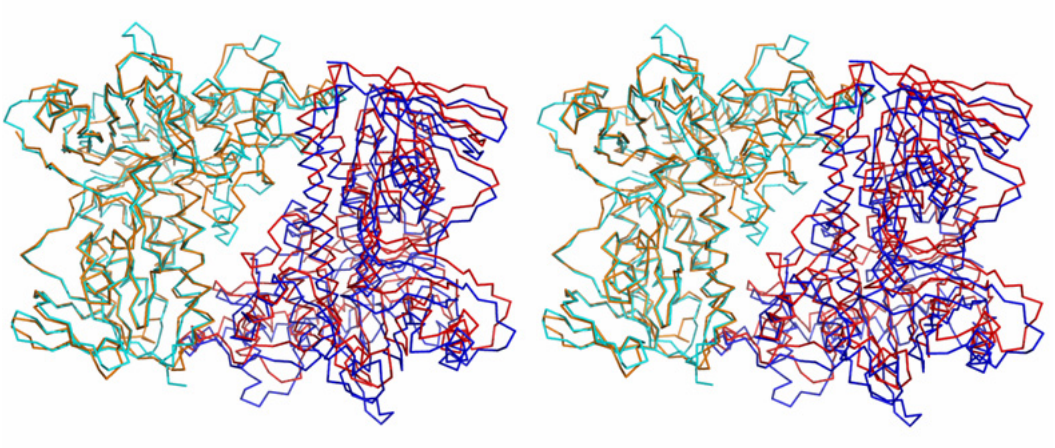


**Figure 2.20**
Ribbon representation of PyrH and PrnA dimers superposed using the SSM algorithm of WinCOOT [51]. PyrH: monomer A in orange; monomer B in red. PrnA: monomer A in cyan; monomer B in blue.

PyrH structure covers the entire *Streptomyces rugosporus* protein coding sequence from the first Methionine (aa 1) to the last residue Glutamine 511 which corresponds to the TEV-PyrH construct residues from aa 41 to aa 551. Monomer A contains two gaps between aa 185 and 204 and between aa 356 and 360. Monomer B contains a gap between aa 186 and 196. Monomer C contains a gap between aa 185 and 204. Monomer D contains a gap between aa 185 and 201. All the gaps are in interconnecting loop regions. The FAD density is clear for all the four monomers. The secondary structure was assigned, using STRIDE [96] web based software (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py) and confirmed by visual inspection of the model, (See Fig. 2.21 and 2.22).

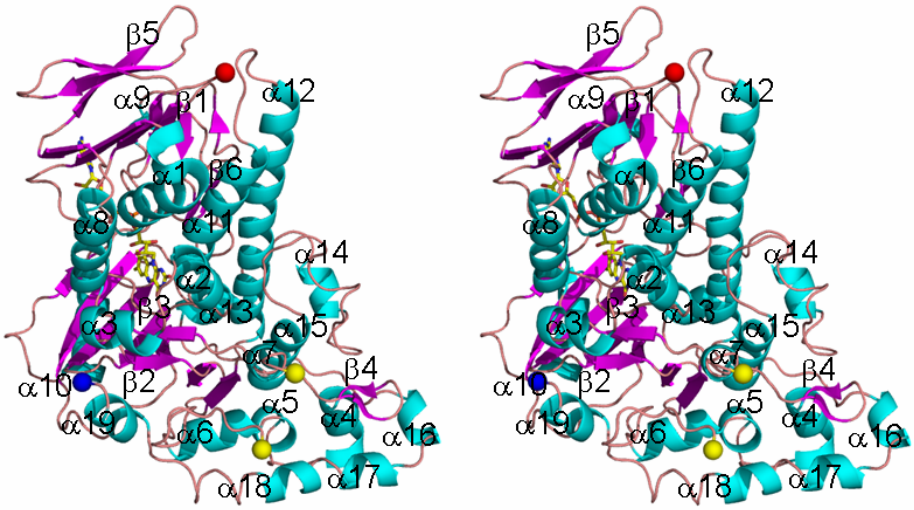



**Figure 2.21**
PyrH monomer B structure cartoon representation. Each of the identified secondary structure is labeled (α = helix, β = sheet). FAD group is in yellow. The N terminal is highlighted by a red sphere the C-terminal with a blue one. The two yellow spheres sign the boundary of the disordered loop between aa 186 and aa 196.

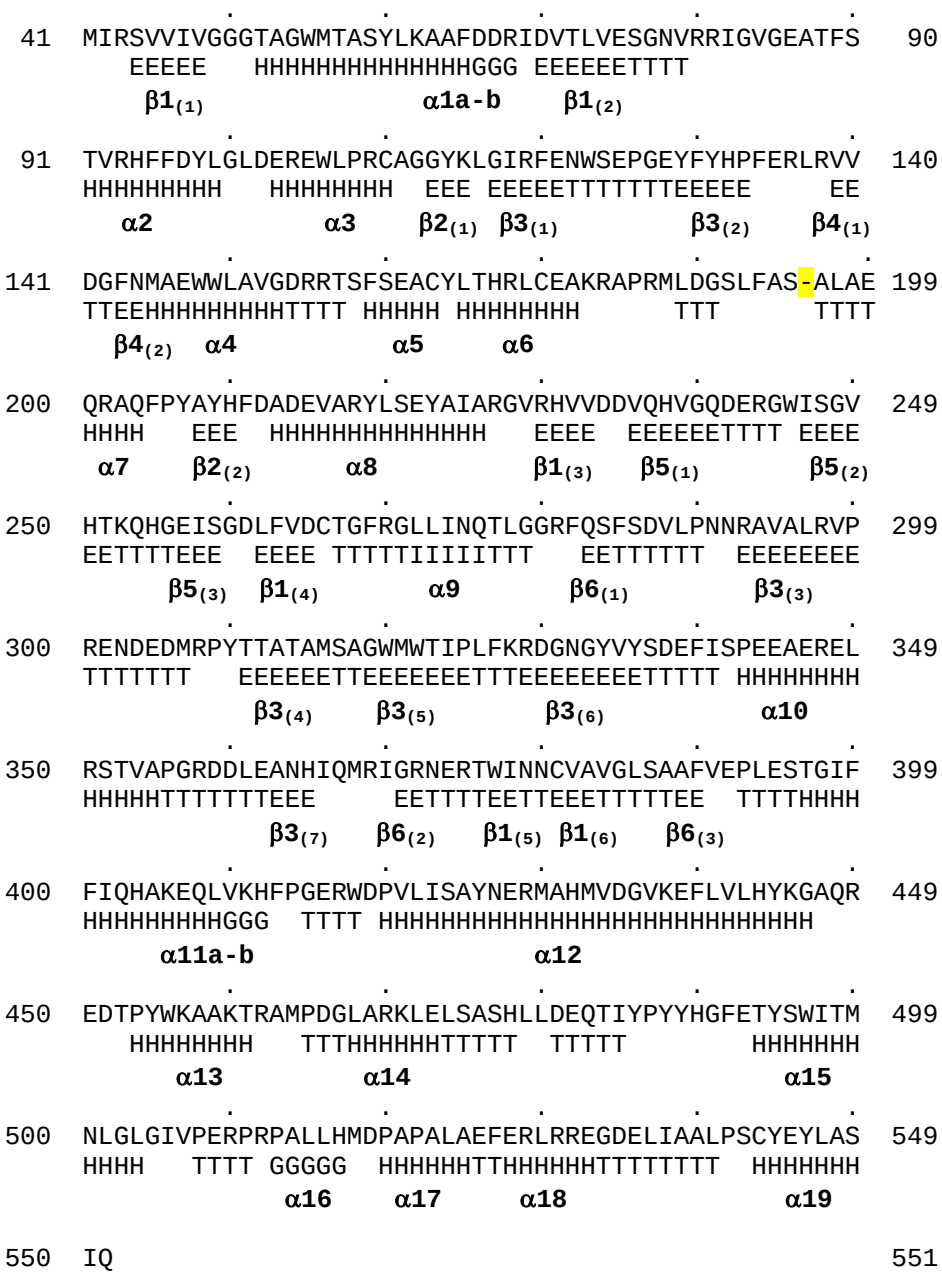



**Figure 2.22**
PyrH, monomer B secondary structure assignment output file from the web based version of STRIDE [96] (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py). Residues are numbered as in the refined structure. T = turns, H = α-helix, G = $3_{10}$-helix, I = Pi-helix, E = β-strand. Shaded in yellow is the gap containing the disordered region from aa 186 to aa 196

### 2.7.3 PyrH and PrnA comparison

Structural superposition of PyrH monomer B on PrnA model using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50] resulted in a Z-score of 16.7 with a RMSD of 1.35Å for 472 aligned residues. The structure based superposition is shown is Figure 2.23. Conserved important residues were identified by manual inspection of PyrH/PrnA superposed models in WinCOOT [51]. In Figure 2.24 are presented both the ribbon and the cartoon models of the two superimposed structure. The overall protein fold is clearly conserved; major differences are present in the "pyramidal" substrate binding domain. The FAD cofactor is superposed in the two models.

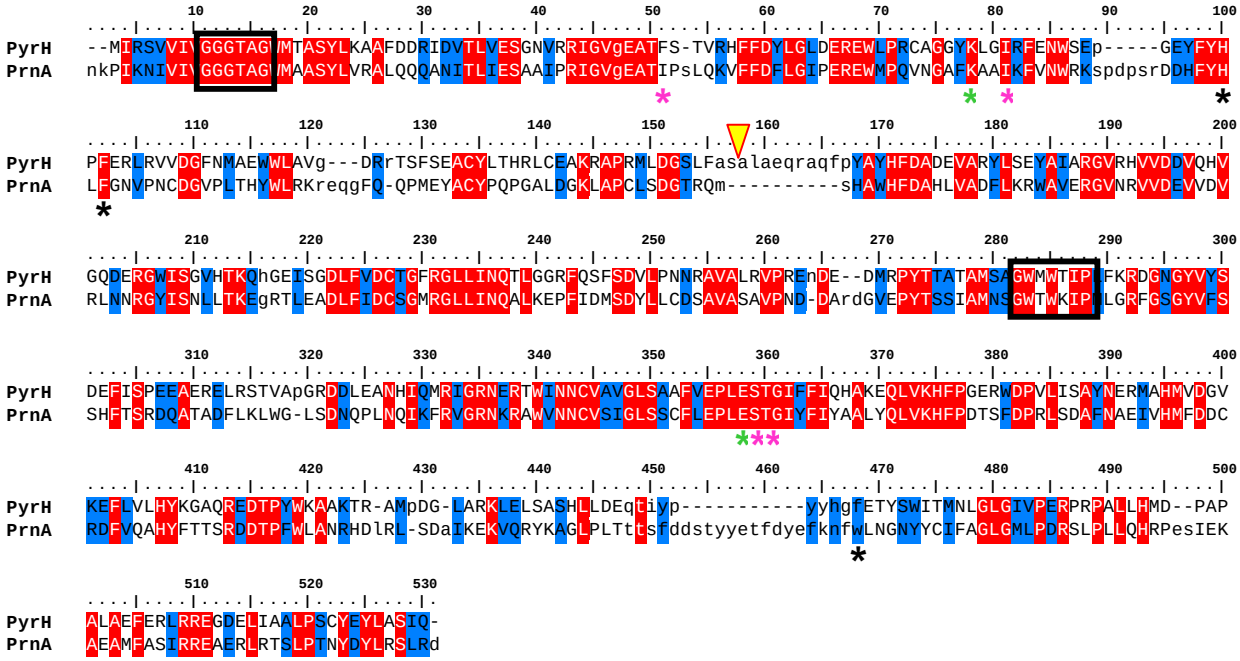
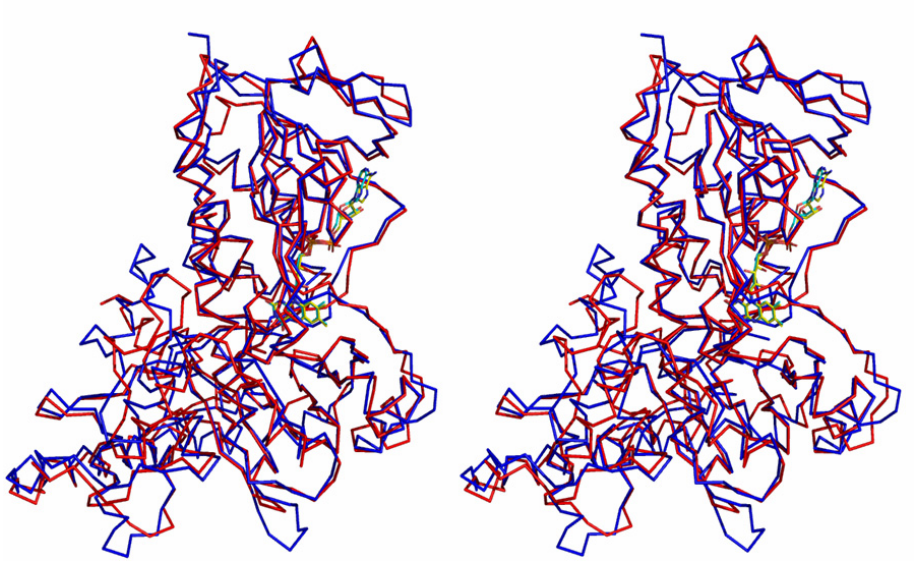
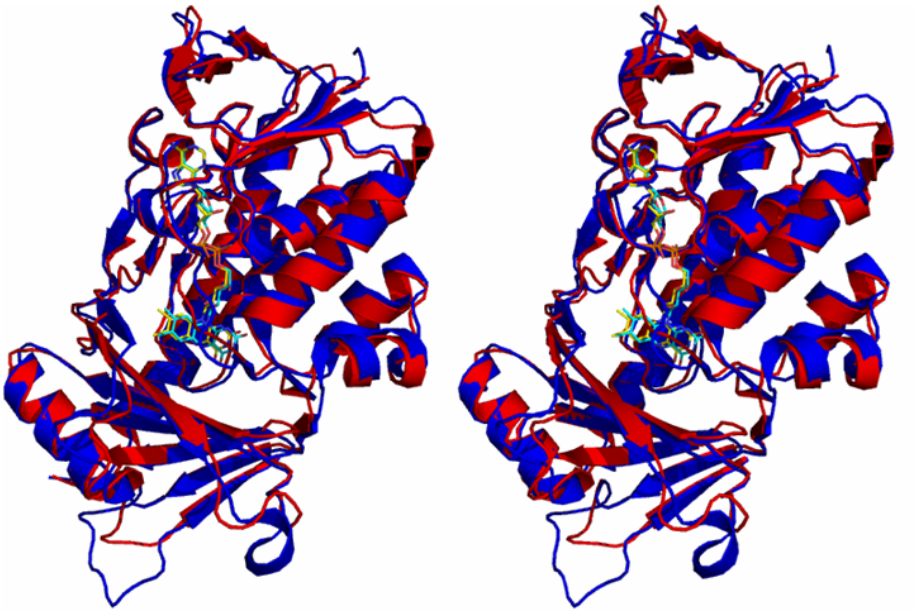


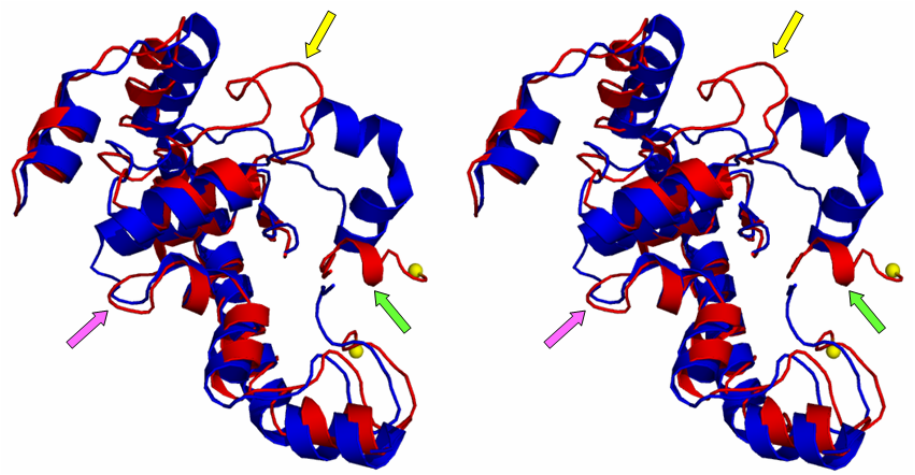

**Figure 2.23**
Sequence alignment based on structural superposition of PyrH on PrnA using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50]. In capital superposed residues; shaded in red identical residues, in blue similar ones. The black boxes indicate the GxGxxG and WxWxIP motifs involved in FAD binding and protection. Highlighted by a star are aa essential for PrnA reaction mechanism: in purple aa constituting the HOCl; channeling tunnel; in green E347 and K79; in black the tryptophan binding residues. Shaded in red identical residues, in blue similar ones. The yellow triangle indicates the position of the unique gap (disordered loop) present in PyrH protein model.



A



B

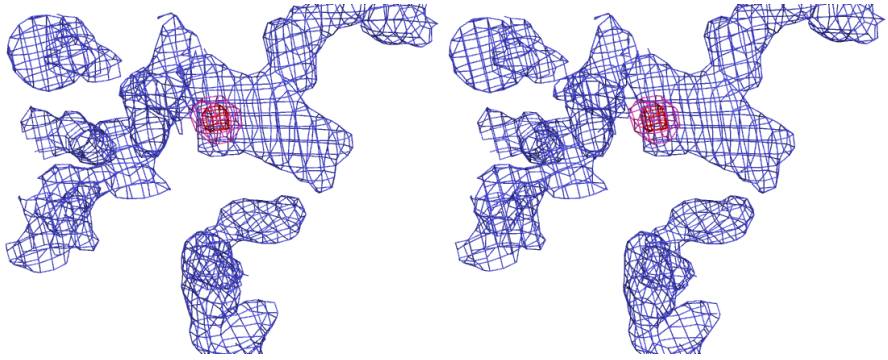


**C**

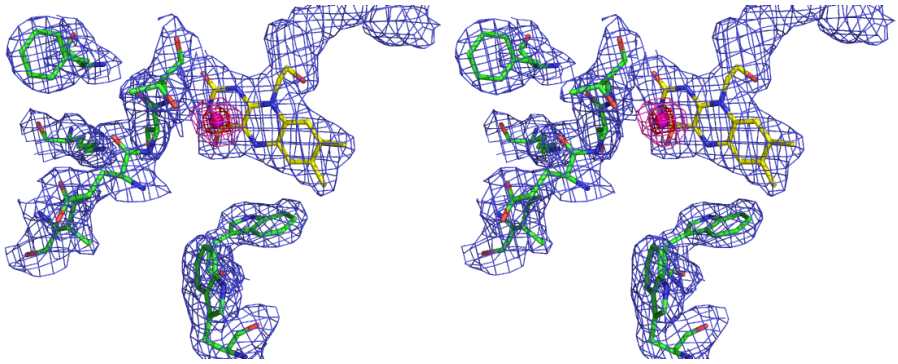
**Figure 2.24**
Ribbon (A) and Cartoon (B, C) representation of PyrH (red , yellow FAD) superposition to PrnA (blue, cyan FAD), using the SSM algorithm of WinCOOT [51]. In A entire protein, in B FAD binding domain, in C substrate binding domain. Despite small differences the overall protein fold is clearly conserved. Note the conserved position of the FAD prosthetic group. Major differences are observed in the less conserved substrate binding module. In panel C: the magenta arrows indicates the, shorter than PrnA, PyrH loop connecting helices $\alpha$4 and $\alpha$5. The green arrow indicates the PyrH region we were not able to locate in the electron density map: the gap between helices $\alpha$6 and $\alpha$7. Two yellow spheres define the gap boundaries. The yellow arrow indicates the loop between helices $\alpha$14 and $\alpha$15; this region is more extended in PrnA where it hosts an extra helix.

### 2.7.3.2 PyrH the active site

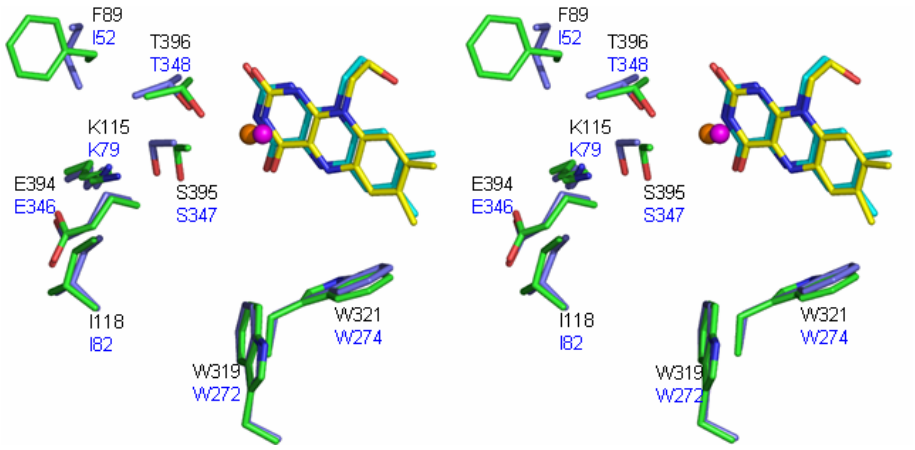
The features forming the FADH binding site in PrnA structure are clearly conserved in the PyrH model, see Fig. 2.25. The two PyrH tryptophan, W319 and W321, which are thought to be essential, overlap to PrnA ones. The FAD isoalloxazine ring is also superposed in the two models. In PyrH structure next to the FAD we were able to locate a chlorine ion, magenta sphere in Fig. 2.25C, already identified in PrnA model, orange sphere. Most of the residues which define the putative HOCl tunnel from FAD to the tryptophan binding region are conserved, notably the important E346/K79 PrnA couple (E394/K115 in PyrH). At the far end of the tunnel I52 is replaced by F89 representing the only difference in the HOCl channeling region. This residue at the end of the tunnel could be responsible for the different orientation of the tryptophan substrate with respect to the incoming hypochlorous acid, thus leading to halogenation of the indole ring at position 5 instead of at 7 observed in PrnA.



**A**



**B**




**C**

**Figure 2.25**

**A**: PyrH $2F_O$-$F_C$ at 1σ electron density map around the protein tunnel region in blue. In red the $2F_O$-$F_C$ at 4σ and in magenta the $2F_O$-$F_C$ at 3σ around the Cl ion. **B**: the same as **A** with both FAD and the residues defining the tunnel region modelled into the electron density. The 0FAD cofactor isoalloxazine ring is in yellow, PyrH clorine ion as a magenta sphere. **C**: Superposition of PyrH and PrnA HOCl tunnel region. PyrH residues in green, PyrH FAD cofactor isoalloxazine ring in yellow, PyrH clorine ion magenta sphere. PrnA residues in blue, PrnA FAD cofactor in cyan, PrnA clorine ion orange sphere. PrnA I52 is not conserved in PyrH structure, where is replaced by F89. This residues is already part of the substrate binding region.

### 2.7.3.3 PyrH the substrate binding region

Analysis of the tryptophan binding site in the superposed structures reveals that two of the three aromatic residues involved in the stacking of the tryptophan indole ring in PrnA protein, H101, F103 are conserved in PyrH protein with H132, F134. W455 in PrnA is replaced by F491 in PyrH. The presence in the substrate binding region of PyrH of F89 is thought to prevent L-tryptophan from binding in the same orientation observed in PrnA model, see Fig. 2.26.

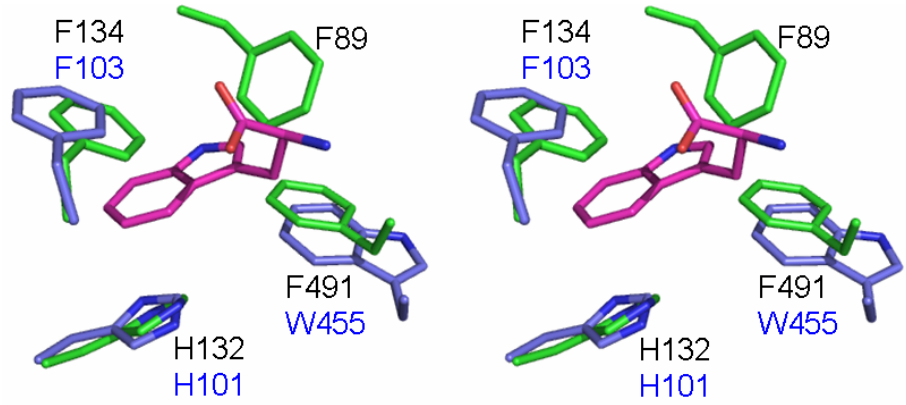



**Figure 2.26**
Superposition of PyrH and PrnA tryptophan binding region. PyrH residues in green, PrnA residues in blue, L-tryptophan from PrnA substrate bounded structure in magenta. All the aromatic residues involved in L-tryptophan indole ring stacking in Prna model are conserved in PyrH with the only exception of W455 replaced by F491. In PyrH protein presence of F89 would prevent L-tryptophan binding in the way observed in PrnA structure. The phenylalanine ring would clash with the tryptophan main chain.

None of the PrnA residues involved in hydrogen bond formation with the L-tryptophan carbon $\alpha$ nitrogen and carboxyl group is conserved in PyrH structure, see Fig 2.27.

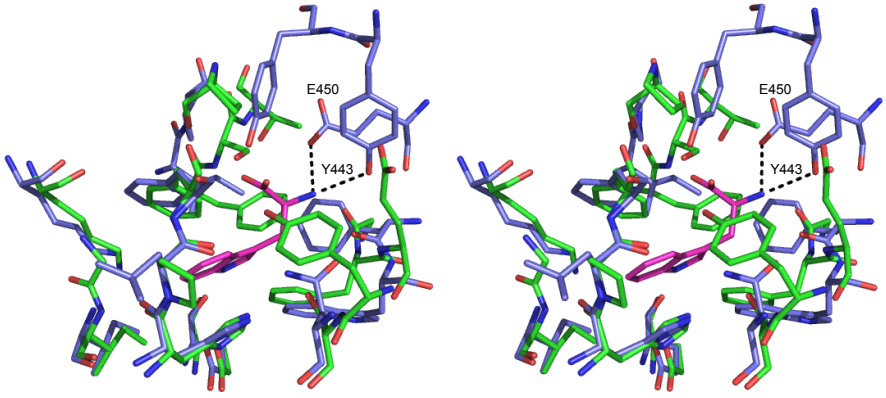
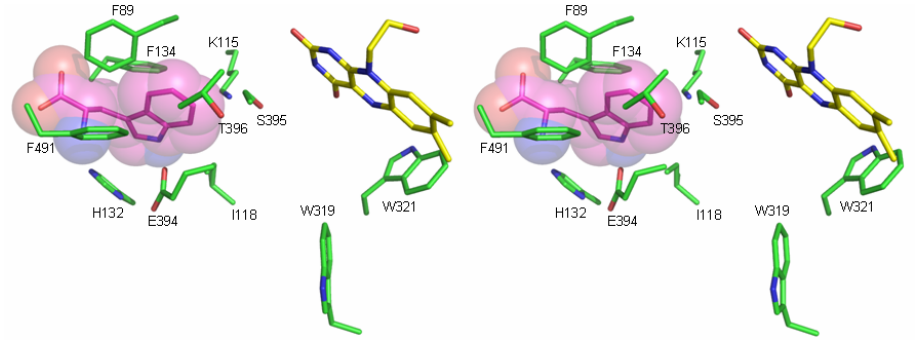
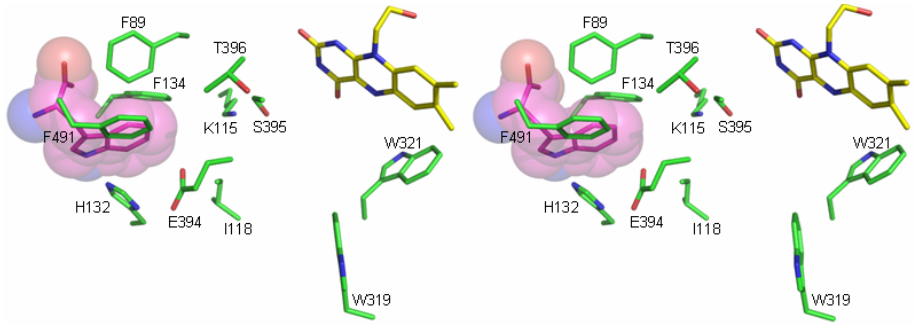



**Figure 2.27**
Superposition of PyrH and PrnA tryptophan binding region. PyrH residues in green, PrnA residues in blue, L-tryptophan from PrnA substrate bounded structure in magenta. The L-tryptophan carbon $\alpha$ nitrogen interacts directly with E450 and Y443, while the carbon $\alpha$ carboxyl group interacts via water molecules (not shown) with other carboxyl group. The superposed PyrH structure is not conserved in this region suggesting a different tryptophan binding mode for the latter halogenase.

Attempts to co-crystallize PyrH with its substrate L and D tryptophan or its products 5-Cl-DL-tryptophan and 5-Br-DL-tryptophan were unsuccessful. Nevertheless we were able to model L-tryptophan into the PyrH substrate binding region using ArgusLab 4.0 software [97] (http://www.planaria-software.com). The modelling software gave us a results consistent with the proposed enzyme reaction mechanism. Two of the best fitting results examples are shown in Fig. 2.28. In both results the indole ring fifth position results to be facing the FAD molecule at the end of the HOCl tunnel. Furthermore the essential residues K115 and E394 are pointing towards the indole ring. It is not possible to rule out conformational changes in the tryptophan binding site upon substrate binding, however such changes would probably be minimal and influence only the aminoacid side chains position rather than the $\alpha$ carbon main chain due to the rigidity observed in the empty, substrate bond or product bond PrnA structures [15].




**A**




**B**

**Figure 2.28**
A,B: Two results obtained docking of L-tryptophan (purple) into PyrH (residues green, FAD isoalloxazine ring yellow) proposed substrate binding site. For computer-assisted molecular modelling we employed the ArgusLab software [97].

We expect L-tryptophan to bind into PyrH active site with an intermediate position between the ones described in Fig. 2.28. This is evident observing in Fig. 2.29 the two modelled

L-tryptophan positions superposed to 7-Cl-tryptophan from PrnA product bond structure. Given the position of the clorine atom in 7-Cl-tryprophan and the C-5 position of the docked molecules we expect L-tryptophan to bind with the aromatic indole ring stacked between F134 and F491and its C-5 position closer two the important catalytic couple K115/E394.

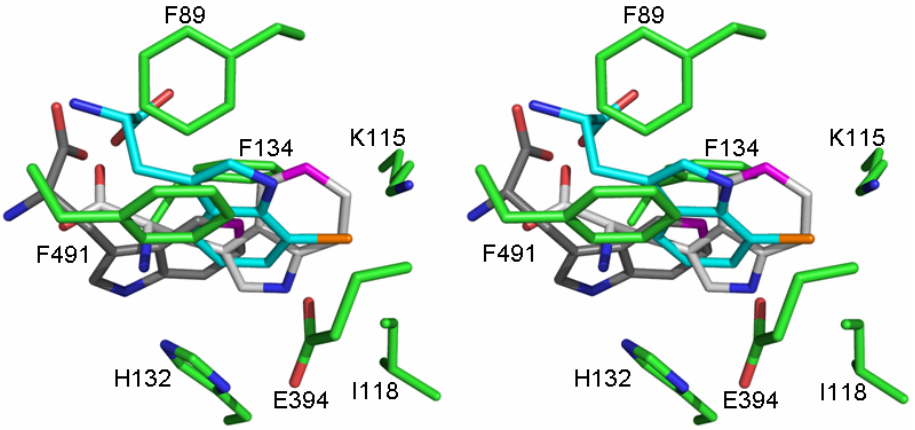



**Figure 2.28**
Superposition of 7-Cl-tryptophan (in cyan, clorine atom in orange) from PrnA product bond structure into PyrH active site with the two L-tryptophan molecules docked into PyrH active site (light and dark gray, C-5 position in magenta).

To further validate the results we decide to repeat the simulation modifying the tryptophan binding site. As shown in Fig. 2.30, F89 side chain was rotated to allow the substrate to sit in the active site in the same way observed in PrnA. Again the results (See Fig. 2.30) show L-tryptophan facing with its carbon-5 position the end of the HOCl tunnel confirming the previous results.

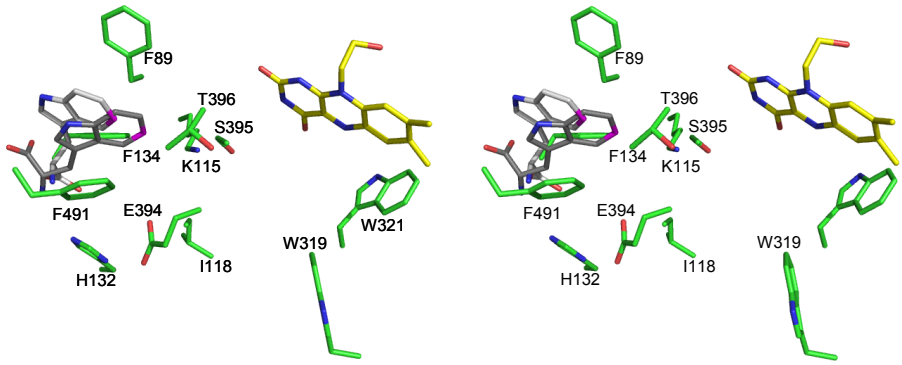



**Figure 2.30**
Docking of L-tryptophan, light and darkg ray C-5 position highlighted in magenta, into PyrH substrate binding site, residues in green. F89 has been rotated respect to the experimental structure (See Fig. 2.26, 2.27, 2.28, 2.29) to allow L-tryptophan to sit in the same position as in PrnA. Despite the F89 has been moved, the two best resulted obtained show modelled L-tryptophan molecules facing the end of the HOCl tunnel with their carbon-5 atom. For Computer-assisted molecular modelling we employed the ArgusLab software [97].

To test the reliability of the modelling software the program was used to fit L-tryptophan into PrnA active site. The result is shown in Fig. 2.30 and confirms the quality of the docking algorithm.

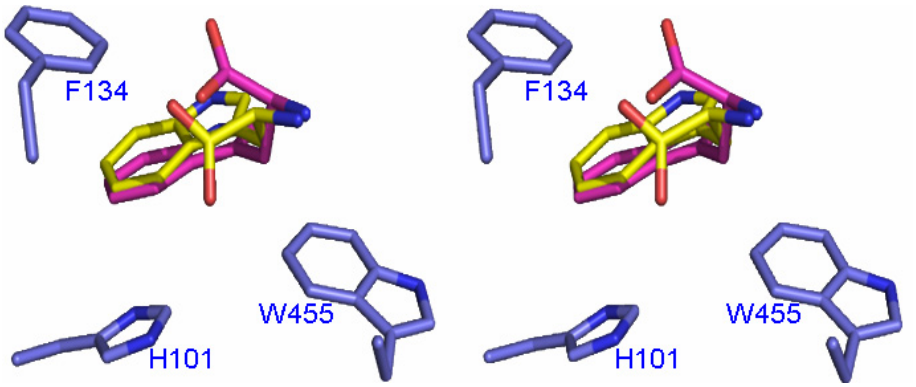



**Figure 2.30**
Docking of L-tryptophan, yellow, into PrnA substrate binding site, residues in blue. The experimentally determinated L-tryptophan is in purple. For computer-assisted molecular modelling we employed the ArgusLab software [97].

## 2.8 Future Work

The PyrH structure was confirmed, as expected, to belong to the same superfamily as PrnA. Success in the co-crystallization experiment with protein substrates and reagents would definitively establish the reaction mechanism of FAD dependent tryptophan halogenases. Surely, at this point, elucidation of the structure of FAD dependent halogenases with different substrate specificity is of primary interest. This structural information will enable us to extend the PrnA reaction mechanism to the entire halogenase family.

Given tryptophan halogenases poor kinetic parameters in *in vitro* conditions further study will be necessary to identify better reaction conditions to exploit this class of enzyme as a biocatalyst. Identification of the physiological flavin reductase partner could be one of the strategies to pursue. Due to the high number of *Pseudomonas* genome sequencing projects recently completed, with respect to the *Streptomyces* genus, it should be easier to identify the partner reductase of *Pseudomonas* halogenases rather than the *Streptomyces* ones.

# Chapter 3

# PrnB Cloning, Expression and Purification

## 3.1 Summary

PrnB is a new unique completely uncharacterized enzyme. It catalyzes the second step of the pyrrolnitrin biosynthetic pathway in *Pseudomonas* and related species. In this chapter we describe PrnB cloning, over-expression and purification from two sources. The protein could not be expressed in soluble form using *E. coli*. Soluble protein was obtained using *Pseudomonas* itself as expression host. The pure protein was red colored and further analysis suggested it contains an heme group. Preliminary crystallization trials with overexpressed PrnB both from *Pseudomonas* and *Myxococcus* failed to produce any relevant hit. Analysis of gel filtration chromatograms and analytical gel filtration of both proteins revealed the presence of aggregation that could be eliminated with the use of reducing agents indicating presence of possible disulfide bonds. Failure to identify the putative cysteines involved in the oligomerization by labeling experiments and overall lower *Myxococcus* PrnB tendency to produce aggregates pointed towards the presence of aspecific disulfide bonds. Mutation of all *Pseudomonas fluorescence* PrnB Cysteines to Serines resulted in a monomeric enzyme that retains its activity and could be crystallized; the enzyme from P. fluorescens could be purified in mg quantities. In Chapter 4 we describe *P. fluorescens* PrnB crystallization and structure determination.

## 3.2 PrnB sequence analysis

Apart from genetic evidences showing *prnB* gene involvement into the second step of the pyrrolnitrin biosynthetic pathway [20; 21; 61], nothing is known about PrnB protein. The reaction catalyzed by this enzyme, the conversion of 7-chloro-L-tryptophan to monodechloroaminopyrrolnitrin coupling the rearrangement of the indole ring to a phenylpyrrole and a decarboxylation, has not been identified in any other metabolic pathways discovered so far. The enzyme has been reported to be active also on tryptophan [20; 21; 61] (See Fig. 3.1).

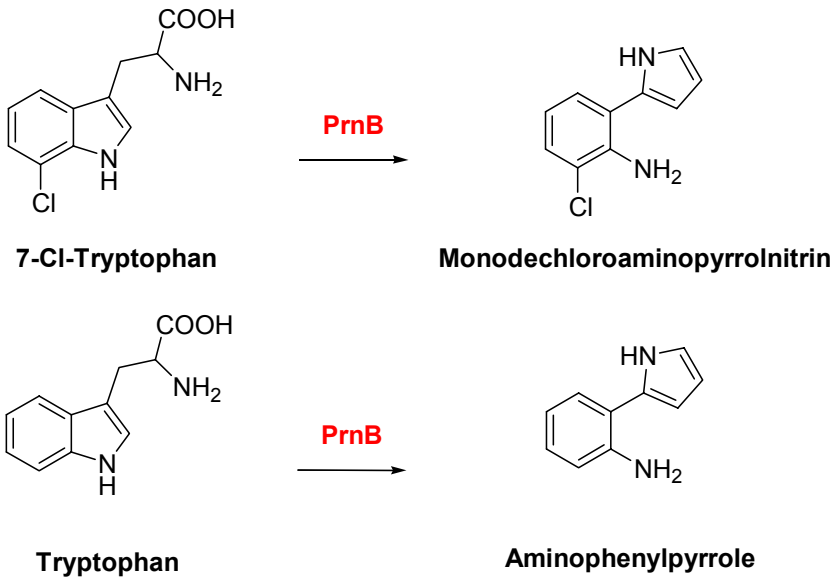



**Figure 3.1**

Analysis of *prnB* gene did not reveal any conserved domain or motif. BLAST searching did not produce relevant meaningful hit. The only interesting result was the identification through the NCBI Conserved Domain Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) of a non statistically significant short sequence match with indoleamine-2,3-dioxygenase (IDO), a heme containing protein involved in tryptophan catabolism [98]. The Secondary Structure prediction algorithm, 3D-pssm (http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html), predicts PrnB to be as IDO an all $\alpha$-helical protein.

## 3.3 Aim of our study

With so little known or predictable about PrnB a structural study was planned. A structure of PrnB would be an excellent starting point to formulate testable hypotheses about its enzymatic mechanism. Information on PrnB mechanism and structure would be very useful in harnessing its unusual activity in biotranformation study. A necessary precondition of structural study is the production of mg quantities of highly pure homogenous protein.

## 3.4 Materials and Methods

All the genetic techniques and biochemistry procedures have been already illustrated in Chapter 2 Materials and Methods section. Only relevant differences are reported in this section.

### 3.4.1 Strains used in this study

**Table 3.1**

| N# | Strain | Relevant genotype | Organism / Use | Supplier / References |
|---|---|---|---|---|
| **1** | BL915 | PRN+ | *Pseudomonas fluorescens / genomic DNA extraction* | 70 |
| **2** | Mx f147 | PRN+ | *Myxococcus fulvus / genomic DNA extraction* | 67 |
| **4** | TOP10 | *F- mcrA D(mrr-hsdRMS-mcrBC) f80lacZDM15 DlacX74 deoR recA1 araD139 D(ara-leu)7697 galU galK rpsL (Str$^R$) endA1 nupG* | *Escherichia coli* / Cloning | Invitrogen |
| **5** | S17.1 | *Tp$^R$ Sm$^R$ recA, thi, pro, hsdR$^-$M$^+$RP4: 2-Tc:Mu: Km Tn7 λ pir* | *Escherichia coli* / Conjugation | Biomedal |
| **6** | BL915 ΔORF1-4 | PRN$^-$ (ΔprnA, ΔprnB, ΔprnC, ΔprnD) *km$^R$* | *Pseudomonas fluorescens* / Expression | 20; 70 |
| **7** | KT2440 | | *Pseudomonas putida* / Expression | 99; 100 |
| **8** | DB3.1™ | F- gyrA462 endA1 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(SmR) xyl-5 λ- leu mtl1 | *Escherichia coli / Cloning, manteining of Gateway ccdB containing plasmids* | Invitrogen |
| **9** | BL21 | *E. coli* B F– ompT hsdS (rB– mB–) dcm+ Tetr gal (DE3) endA Hte | *Escherichia coli / Expression Strain* | Novagen |
| **10** | Rosetta | *E. coli* B F- ompT hsdSB (rb- mB-) gal dcm lacY1 (DE3) pRARE6 (CmR) | *Escherichia coli / Expression Strain* | Novagen |

### 3.4.2 Plasmids

**Table 3.2**

| N# | Plasmid | Antibiotic Resistance / Host | Type of Vector / Characteristic | Supplier / References |
|---|---|---|---|---|
| **1** | pCIB-HIS (~21Kbp) | Tetracycline / *E.coli*; *Pseudomonas sp.* | HIS tag expression, *Ptac* promoter Broad host range gram -/+ vector (~21Kbp) | 20; 21; 74; 77; 78 |
| **2** | pFastBac HT A | Ampicillin, / *E.coli* | cloning for Bacmid generation; high copy number ideal for subcloning (4856bp) | Invitrogen |
| **3** | pCR2.1-TOPO | Ampicillin, Kanamycin / *E.coli* | Cloning (3900bp) | Invitrogen |
| **4** | pZERO-Blunt | Kanamycin / *E.coli* | Cloning (3300bp) | Invitrogen |
| **5** | pDONR221 | Kanamycin *ccdb*$^+$ / *E.coli* | GATEWAY cloning (4762bp) | Invitrogen |
| **6** | pDEST17 | Ampicillin, / *E.coli* | GATEWAY HIS tag expression (6354bp) | Invitrogen |
| **7** | pEXP3-DEST | Ampicillin, / *E.coli* | GATEWAY HIS-Lumio tag expression (4607bp) | Invitrogen |
| **8** | pIVEX-MBP-DEST | Ampicillin, / *E.coli* | GATEWAY HIS-MBP-tag expression (6403bp) | Naismith Lab |
| **9** | pETG-41A | Ampicillin, / *E.coli* | GATEWAY HIS-MBP-tag expression (8219bp) | EMBL |

### 3.4.3 Bacterial Growth and Treatments

#### *3.4.3.1 Media*

MD1 medium for*Myxococcus*:

PMM-Agar: 3,0 g/l Peptone from casein
2,0 g/l Mg $SO_4$ X $7H_2O$
0,5 g/l $CaCl_2$ X $2H_2O$

dissolve in $H_2O$, pH 7,2; autoclaved.

*Myxococcus* cells liquid cultures were incubated on an orbital shaker at approximately 180 to 210rpm at 30˚C for 6 days before genomic DNA extraction.

### *3.4.3.2 Antibiotics*

The following antibiotics were used when appropriate:

| Antibiotic | Stock at | Used at |
|---|---|---|
| Ampicillin Sodium | 50 mg/ml in $H_2O$ at -20°C | 100 µg/ml |
| Kanamycin Sodium sulphate | 25 mg/ml in $H_2O$ at -20°C | 50 mg/ml |
| Spectinomycin | 50 mg/ml in $H_2O$ at -20°C | 100 µg/ml |
| Streptomycin | 50 mg/ml in $H_2O$ at -20°C | 30 µg/ml |
| Chloramphenicol | 34 mg/ml in EtOH at -20°C | 34µg/ml |
| Tetracycline | 5 mg/ml in EtOH at -20°C | 15µ g/ml *E. coli*<br>30 µg/ml *Pseudomonas sp*. |

## 3.4.4 Genetic Techniques

Gateway™ BP and LR clonase reaction, were performed by standard procedures or as recommended by the suppliers of the reagents used [101; 102]

### *3.4.4.1 Isolation of Genomic DNA from Pseudomonas and Myxococcus*

Genomic DNA was extracted from *Pseudomonas* using the *QIAGEN Blood & Cell Culture DNA Kit* by using the Bacteria method. 10ml of an O/N culture were processed on a genomic DNA purification column. Genomic DNA quality was assessed running DNA gels [103].

### *3.4.4.2 Polymerase Chain Reaction – PCR*

PCR has been used to amplify the *prnB* gene from genomic and plasmid DNA. The reaction was performed using the *GeneAmp® PCR System 2400* thermal cycler *(Applied Biosystem)*. A standard reaction contained the following reagents:

| Reagent | Amount - Stock [C] | Final [C] | PROGRAM | |
|---|---|---|---|---|
| dNTPs | 1µl -10mM | 0.2mM | 98°C X 3'00'' | |
| Oligos (forward and reverse) | 1µl - 100µM | 0.5µM | 98°C X 45'' | 30 cycles |
| Polymerase buffer | 5µl - 10X | 1X | 45°C X 45'' | |
| Polymerase* | 1µl - 2U/µl | 2U | 72°C X 1.30'' | |
| Template DNA | 1µl – 50ngr/µl | 50ngr | 72°C X 10'00'' | |
| $H_20$ | Up to 50µl | - | | |

**Table 3.3 *OLIGOS:***

| N# | Name | Description | Sequence |
|---|---|---|---|
| 1 | 5'-prnB | Amplification of *prnB* from *P.fluorescens* BL915 genomic DNA | GTGGAACGCACCTTGGACCGGGTAGGCGTATTC |
| | 3'-prnB | | TCAGGATTCGTCGAGCGCGGCGCGGAC |
| 2 | 5' prnB-TEV | *P.fluorescens* prnB amplification: 5' contains TEV protease sequence in and 5' of prnB; 3' contains *attB2* site and 3' of prnB; in bold ATG and STOP codons | GAAAACCTGTATTTTCAGGGC**ATG**GAACGCACCTT GGACCGGGTAGGCGTATTC |
| | 3' prnB-GAT | | GGGGACCACTTTGTACAAGAAAGCTGGGTC**TCA**GG ATTCGTCGAGCGCGGCGCGGAC |
| 3 | 5' GAT-TEV | *P.fluorescens* *prnB* amplification: 5' contains *attB1* site in blue TEV protease sequence in yellow; 3' contains *attB2* site in blue and 3' of prnB in bold ATG and STOP codons | GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGAAA ACCTGTATTTTCAGGGC |
| | 3' prnB-GAT | | GGGGACCACTTTGTACAAGAAAGCTGGGTC**TCA**GG ATTCGTCGAGCGCGGCGCGGAC |
| 4 | 5'-prnB-B | Amplification of *prnB* of *P.fluorescens* BL915 5' contains flanking BamHI site 3' contains flanking HindI in bold GTG and STOP codons site | ATTCGTAGCGGATCCA**GTG**GAACGCACCTTGGACC GGGTAGGCGTATTC |
| | 3'-prnB-H | | GTATGCTACCAAGCTT**CTATCA**GGATTCGTCGAGC GCGGCGCGGAC |
| 5 | 5'-prnB-B-TEV | Amplification of *prnB* of *P.fluorescens* BL915 5' contains flanking BamHI site and TEV , in yellow, sequence 3' contains flanking HindIII site in bold ATG and STOP codons | ATTCGTAGCGGATCCA**GAAAACCTGTATTTTCAGG GC**ATGGAACGCACCTTGGACCGGGTAGGCGTATTC |
| | 3'-prnB-H-TEV | | GTATGCTACCAAGCTT**CTATCA**GGATTCGTCGAGC GCGGCGCGGAC |
| 6 | 5'-prnB-Myxo-Short | Amplification of *prnB* from genomic DNA of *M. fulvus* Mx f147 | ATGAATCCGGGGCAGAACTTTTCATC |
| | 3'-prnB-Myxo-Short | | CCGCGCGGTGTGCTTCAAGGCCGGCTC |
| 7 | 5'-prnB-Mx-External | Amplification of *prnB* containing fragment from genomic DNA of *M. fulvus* Mx f147 | CAAGGGGCTCGGTGGCGCTTTCACGAG |
| | 3'-prnB-Mxt-External | | GAATGAAGGGTGTCAACGGTGGGCTC |
| 8 | 5' prnB-B-Mx | Amplification of *prnB* of *M. fulvus* Mx f147 5' contains flanking BamHI site 3' contains flanking HindIII site in bold ATG and STOP codons | ATTCGTAGCGGATCCA**ATG**AATCCGGGGCAGAACT TTTCATC |
| | 3' prnB-H-Mx | | GTATGCTACCAAGCTT**CTATCA**CCGCGCGGTGTGC TTCAAGGCCGGCTC |
| 9 | 5'-prnB-B-TEV-Mx | Amplification of *M. fulvus* Mx f147 *prnB* 5' contains flanking BamHI site and TEV sequence in yellow 3' contains flanking HindIII site | ATTCGTAGCGGATCCAGAAAACCTGTATTTTCAGGG CATGAATCCGGGGCAGAACTTTTCATC |
| | 3'-prnB-H-TEV-Mx | | GTATGCTACCAAGCTT**CTATCA**CCGCGCGGTGTGC TTCAAGGCCGGCTC |

Reagent concentrations and procedure were chosen as recommended by the manufacturers of *BIO-X-ACT (BIOLINE Cat. No. BIO-21065); Thermalace (INVITROGEN Cat. No. E0200); Vent Polymerase (NEW ENGLANDS BIOLAB Cat. No.M0254L); Pfu Polymerase (PROMEGA Cat. No.M774B); Taq Polymerase (ROCHE Cat. No.1-146-173).*

### *3.4.4.3 Site Directed Mutagenesis*

*M. fulvus* PrnB mutant was created using the QuickChange II XL Site-Directed Mutagenesis Kit (Strategene Cat. N# 200521). Three round of mutagenesis on plasmid pFastBac-HT-A carrying the *M. fulvus* TEV-*prnB* cloned into BamHI HindIII site were necessary to mutate the Cysteine triplet C129, C130 and C171 using primer couples 1,2 and 3 shown in Table 3.4. After every cycle of mutagenesis, the plasmid DNA was sequenced to verify the introduced mutation. After the last mutation the TEV-*prnB* BamHi HimdIII insert was cloned into plasmid pCIB-HIS using the same resctriction sites. The expression construct was transferred to *E.coli* S17.1 and then by conjugation to *Pseudomonas fluorescens* BL915 ΔORF1-4.

The *P. fluorescens PrnB* mutant was created using the QuickChange Multi Site Site-Directed Mutagenesis Kit (Strategene Cat. N# 200514). Only one round of mutagenesis on plasmid pFastBac-HT-A carrying the *P. fluorescens* TEV-*prnB* cloned into BamHI HindIII site was necessary to mutate the cysteine triplet C21, C160 and C176. As suggested by the kit instruction we attempted to mutate all the three cysteines using both forward and reverse primers, number 5 and 6 in Table 3.4. We obtained satisfactory results only with the reverse primer series. After confirming the mutations introduction by sequencing mutation the TEV-*prnB* BamHi HimdIII insert was cloned into plasmid pCIB-HIS using the same resctriction sites. The expression construct was transferred to *E.coli* S17.1 and then by conjugation to *Pseudomonas fluorescens* BL915 ΔORF1-4.

**Table. 3.4**

| N# | Name | Sequence |
|---|---|---|
| 1 | 5'-C129S-Myxo | CTGGAGCGCCGGTACACCTCCTGCCGCGACGAG |
| | 3'-C129S-Myxo | GAACCTCGTCGCGGCAGGAGGTGTACCGGCGCTC |
| 2 | 5'-C130S-Myxo | GAGCGCCGGTACACCTGCTCCCGCGACGAGGTTC |
| | 3'-C130S-Myxo | GGTGAACCTCGTCGCGGGAGCAGGTGTACCGGCG |
| 3 | 5'-C171S-Myxo | CTCGTTCGCGCCAATGTCCGACGAAATCACAAGC |
| | 3'-C171S-Myxo | GGTGGCTTGTGATTTCGTCGGACATTGGCGCGAAC |
| 4 | 5'-C129S-MyxoC130S | GAGCGCCGGTACACCTCCTCCCGCGACGAGGTTC |
| | 3'-C129S-MyxoC130S | CCTCGTCGCGGGAGGAGGTGTACCGGCGCTC |
| 5 | 5'-C21S-fluorescens | ACGCTGCCGTGGCGGCCTCCGATCCGCTGCAGGC |
| | 5'-C60S-fluorescens | GCGCGGCCTGCCCTCCGGCTGGGGTTTCGTCGAAGC |
| | 5'-C175S-fluorescens | CGAGTTCGCGCAAAGGTCCGACGAGCTGGAAGCC |
| 6 | 3'-C21S-fluo | CGCGCCTGCAGCGGATCGGAGGCCGCCACGGCAGC |
| | 3'-C60S-fluo | GACGAAACCCCAGCCGGAGGGCAGGCCGCGCACCGG |
| | 3'-C175S-fluo | CTTCCAGCTCGTCGGACCTTTGCGCGAACTCGG |

**QuickChange II XL PROGRAM (*prnB Myxococcus*)**

| | |
|---|---|
| 95°C X 1'00'' | |
| 95°C X 50'' | 18 cycles |
| 52°C X 50'' | |
| 68°C X 12'00'' | |
| 68°C X 10'00'' | |

**QuickChange Multi Site PROGRAM (*prnB P.fluorescens*)**

| | |
|---|---|
| 95°C X 1'00'' | |
| 95°C X 1'00'' | 30 cycles |
| 50°C X 1'00'' | |
| 65°C X 12'00'' | |
| 65°C X 5'00'' | |

#### *3.4.4.4 DNA Sequencing*

*prnB* containing plasmid DNA was sequenced at the University of Dundee Sequencing Service and analyzed with the *Contig Express feature of The Vector sequence analysis software (InforMax).* Plasmid DNA and DNA primers were sent to the Sequencing Service according to the Unit Standard Procedures.

| N# | Name | Description | Sequence |
|---|---|---|---|
| **1** | M13F | Sequencing primers for pCR2.1 TOPO, forward | GTAAAACGACGGCCAGTG |
| **2** | M13R | Sequencing primers for pCR2.1 TOPO, reverse | GGAAACAGCTATGACCATG |
| **3** | M13F-GAT | Sequencing primers for pDONR221, forward | GTAAAACGACGGCCAG |
| **4** | M13R-GAT | Sequencing primers for pDONR221, reverse | CAGGAAACAGCTATGAC |
| **5** | prnB-SEQ | Sequencing primers for *P.fluorescens* prnB inside the gene (from 530bp to 550bp from the 1st GTG) | AGCTGGAAGCCTATCTGCAG |
| **6** | prnB-SEQREV | Sequencing primers for *P.fluorescens* prnB inside the gene (from 1086bp to 1060bp from the 1st GTG) | TCAGGATTCGTCGAGCGCGGCGCGGAC |
| **7** | 5'-pFastBac HT A-For | Sequencing of DNA cloned into pFastBac HT A polylinker (from 3995-4015bp) | TATTCCGGATTATTCATACC |
| **8** | 3'-pFastBac HT A-Rev | Sequencing of DNA cloned into pFastBac HT A polylinker (from 4319-4301bp) | GTTTCAGGTTCAGGGGGAG |

### 3.4.5 Protein Biochemistry

#### *3.4.5.1 Recombinant Protein Expression in E.coli*

For the production of recombinant proteins, *E. coli* expression strains (BL21, Rosetta) carrying the appropriate expression plasmids were grown to an $A_{600}$ between 0.6 and 0.9 *(Spectrophotometer: Pharmacia Biotech Ultrospect 1000)*. The protein expression was induced by adding 0.1 to 1mM Isopropyl β-D-thio-galactopyranoside (IPTG). Alternatively cells were incubated for 10 to 20min at 0˚C or 42˚C (cold heat and heat shock) before induction with IPTG to promote protein expression. After the induction cells were incubated on a shaker from 2 hours to O/N at temperature ranging from 18 to 37˚C. Bacteria were

pelleted by centrifugation 13000xg (25min, 4 ˚C; *Beckman Coulter™ Avanti™ J-20 XP*). The cell pellets were stored at –20˚C until required [104; 105].

#### *3.4.5.2 Purification of His-Tagged Fusion Proteins from E.coli Using Ni $^{2+}$ Chelating Matrix*

Bacterial cell pellets were completely resuspended in 2ml/g binding buffer containing *Complete EDTA free Protease Inhibitor Cocktail Tablets (ROCHE Cat. No.1873580)*. After adding DNAse (final conc. 5μg/ml) and Lysozyme (final conc. 1mg/ml) the cell suspensions were incubated 20min at room temperature before being sonicated or passed twice through a Cell Distruptor, *The Basic Z Constant System Ltd*., to achieve complete cell disruption. The lysate was centrifuged at 48000xg (40min, 4˚C; *Beckman Coulter™ Avanti™ J-20 XP*) to pellet insoluble fractions and debris. Insoluble fractions were resuspended in 8M urea for SDS analysis. Before loading onto the $Ni^{2+}$-column, the soluble fractions were syringe filtered (w/0.22μm). The column *HisTrap HP™ 5ml (Amersham Biotech Pharmacia)* connected to a *P1 peristaltic pump* (*Amersham Biotech Pharmacia)* was regenerated before use as follows: washed with 20ml water, stripped with 20ml 50mM EDTA and after a further wash step with 20ml water, loaded with 10ml 400mM $NiSO_4$ and washed again. After equilibration with 15ml of binding buffer, the protein solution was applied to the column and the flow through retained for analysis. The column was washed with washing buffer until all the unbound proteins were removed from the column. The protein content of the flow through was monitored with Bradford assay. Bound proteins were eluted applying the elution buffer containing 250mM imidazole. The protein content of eluted fractions, collected by a *FRAC100 fraction collector (Amersham Biotech Pharmacia)*, was checked by Bradford assay and analyzed by SDS PAGE. The fractions, which contained the purified protein were concentrated *(VIVASCIENCE VivaSpin Concentrators)* by centrifugation (3000 rpm, 4 ˚C) and afterwards dialyzed against an appropriate buffer to remove imidazole. Buffer compositions are listed in Table 3.5.

**Table 3.5**

| **Binding Buffer:** | **Washing Buffer:** | **Elution Buffer:** |
|---|---|---|
| $KH_2PO_4$ 20mM pH7.4 | $KH_2PO_4$ 20mM pH7.4 | $KH_2PO_4$ 20mM pH7.4 |
| NaCl 0.1M | NaCl 0.1M | NaCl 0.1M |
| Imidazole 20mM | Imidazole 35mM | Imidazole 250mM |

### *3.4.5.3 PrnB Enzymatic Activity Assay*

PrnB enzymatic activity in *Pseudomonas* crude extract were assayed in van Peé's Lab as previously described [20; 21]. To asses PrnB activity *in vivo* whole cell assay/feeding experiments [40; 106] were carried out. Over-night *Pseudomonas* culture in HNB medium containing the appropriate antibiotic were diluted 1 to 20 in fresh medium. Cells were grown for 24 hrs at 30°C and 200rpm on orbital shaker incubators. Sterile filtered 7-DL-Chloro-Tryprophan was added to a final concentration of 1mM. Cells were grown for additional 2 days. Medium alone or sonicated medium plus bacteria was extracted twice with one volume of Ethyl Acetate. The extracted mixture was dried under vacuum and resuspended in Methanol:$H_2O$ 65:35. The extract were analysed by isocratic HPLC, Methanol:$H_2O$ 65:35, using a Varian OmniSpher C-18 250 * 4.6 mm column. Flow rate was 1ml/min and detection at 220nm. The retention time of PrnB product MDA was comparable with previously described data by van Peé *et al.* [40; 106]. Mass spectrometry analysis of the putative MDA peak by St-Andrews University Mass Spec Services confirmed the presence of a chlorinated compound of the right size (192Da), with the characteristic 3:1 ratio pattern of $^{35}Cl$:$^{37}Cl$, see Fig. 3.2.

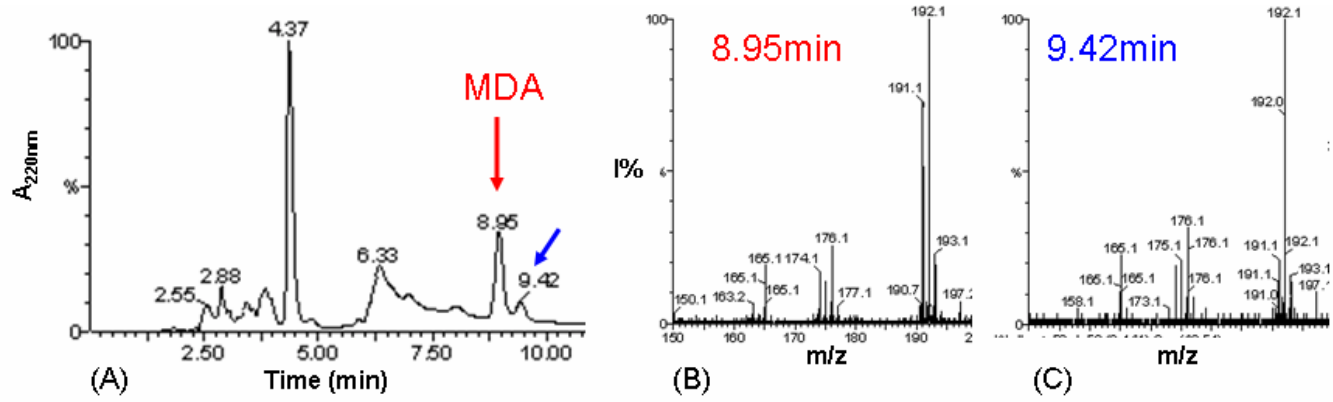



**Figure 3.2**
Mass spectrometry analysis of the MDA peak. A: C-18 elution chromatogram of a broth extract containing MDA, highlighted by a red arrow, the peak run at 8.95min. At 9.42min is present a contaminant peak. B: Mass spectrometry analysis of the 8.95min peak, negative ionization. C: Mass spectrometry analysis of the 9.42min peak, negative ionization. The MDA peak shows the presence of a compound of 191Da with a typical 3:1 ration pattern of chlorine containing compound compatible with the MDA mass of 192Da. The 192Da mass at 8.95min results from contamination by the compound of 192Da present in the 9.42min peak.

## 3.5 Results

### 3.5.1 Cloning of *P. fluorescens* BL915 *prnB* gene

Genomic DNA was extracted from *P. fluorescens* BL915. *PrnB* gene was amplified from genomic DNA using primers 5'-prnB/3'-prnB. Different polymerases were tested. The best results were obtained with the *BIO-X-ACT* polymerase. *BIO-X-ACT* polymerase has a non-template dependent terminal transferase activity that adds a single deoxyadenosine to the 3' ends of PCR products. We clone the gene into the pCR2.1 TOPO vector using the *TOPO TA CLONING KIT*. *E. coli* TOP10 cells were transformed with 5μl of the reaction mix. Plasmid DNA was extracted from Ampicillin resistant clones. Plasmids containing DNA fragment of the right size were sequenced (pCR2.1TOPO*prnB* with primers M13R/M13R). Clones containing the right DNA fragments were isolated.

### 3.5.2 prnB expression in *E.coli*

After we obtained the right *prnB* gene we decided to take advantage of the GATEWAY SYSTEM and subcloned it into a GATEWAY DONOR vector inserting a TEV protease cleavage site immediately upstream of the first protein aa. Two sets of PCR amplification were carried out, by using *Vent Polymerase*, with two different series of primers to avoid the use of extremely long oligonucleotides following the scheme in Fig. 3.3 After the second PCR the four genes were cloned into pDONR221 by using the *GATEWAY BP clonase*.

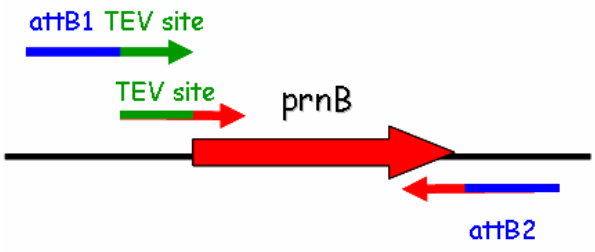



**Figure 3.3**
Amplification strategy for prnB gene into the GATEWAY donor vector pDONR221. Two round of amplification were carried out to avoid the use of long oligonucleotides.

*E. coli* TOP10 cells were transformed with 5μl of the reaction mix. Plasmid DNA was extracted from Kanamycin resistant clones. Plasmids containing DNA fragments of the right size were sequenced (pDONR221*prnB* with primers M13F-GAT/M13R-GAT). Clones containing the right DNA fragments were isolated.

### 3.5.3 Expression of PrnB in *E. coli*

By using *GATEWAY LR clonase mix prnB* was transferred into plasmids pDEST17, pEXP3-DEST, pIVEX-MBP-DEST and pETG-41A, which was provided by the EMBL Protein Expression and Purification Core Facility, Heidelberg. *E. coli* TOP10 cells were transformed with 5μl of the reaction mix. Plasmid DNA was extracted from Kanamycin resistant clones. Plasmids containing DNA fragment of the right size were transformed in *E.coli* BL21 and Rosetta strains. Protein expression trials were set up for each strain. Results are summarized in the table below (Table 3.2). Two examples are shown in the next picture (See Fig.3.4).

**TABLE 3.6**

| Plasmid | Strain | Result |
|---|---|---|
| pDEST17-*prnB* | BL21 | Insoluble protein |
| | Rosetta | Insoluble protein |
| pEXP3-*prnB* | BL21 | Insoluble protein |
| | Rosetta | Insoluble protein |
| pIVEX-MBP-*prnB* | BL21 | Insoluble PrnB protein without MBP tag |
| | Rosetta | Plasmid instable |
| pETG-41A-*prnB* | BL21 | Soluble MBP-PrnB |
| | Rosetta | Soluble MBP-PrnB |

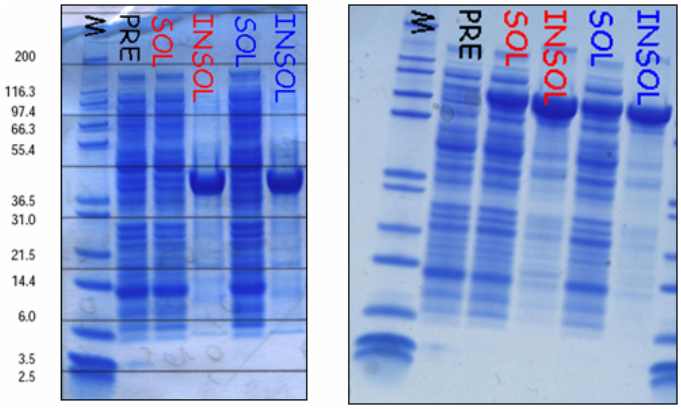



**Figure 3.4**
**a:** SDS PAGE of protein crude extracts from BL21 pDEST17*prnB* cells. Cells were grown until $A_{600}$=0.8, incubated at 42°C (red) or 0°C (blue) for 15min. After IPTG in 0.1mM final concentration was added cells were shifted at 18°C and grown for further 16hours. M: Markers/ladder Mark12; PRE: soluble fraction prior IPTG addition; SOL: soluble fraction; INSOL: insoluble fraction. **b:** SDS PAGE of protein crude extracts from Rosetta pETG-41A-*prnB* cells. Cells were grown until $A_{600}$=0.8, incubated at 42°C (red) or 0°C (blue) for 15min. After IPTG 0.5mM final concentration was added cells were shifted at 30°C and grown for further 4hours. M: Markers/ladder Mark12; PRE: soluble fraction prior IPTG addition; SOL: soluble fraction; INSOL: insoluble fraction.

MBP-PrnB was purified on a *HisTrap HP™ 5ml (Amersham Biotech Pharmacia)* $Ni^{2+}$ chelating column and digested by using TEV protease. Unfortunately, as soon as the PrnB protein was digested and separated from the MBP tag it became unstable causing precipitation of the digestion solution (See Fig 3.5).

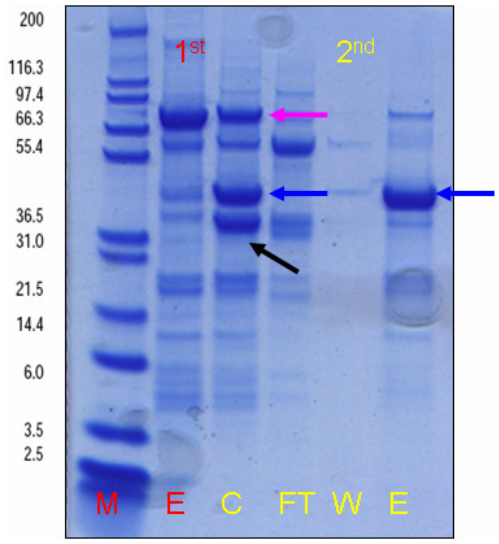



**Figure 3.5**
NuPage 4-12% Bis-Tris gel showing MBP-PrnB after the first (red) and after incubation with TEV protease second (yellow) $Ni^{2+}$ chelating resin column. M: marker (Par 2.4.5.10). E: eluted MBP-PrnB C: cleave FT: flow-through; W: wash; E: elution. The purple arrow indicate MBP-PrnB; the blue one the MBP tag; the black one PrnB. Note that despite concentrating the second $Ni^{2+}$ column FT to an increase in the amount of the contaminant running just below MBP-PrnB there is no clear PrnB band. The PrnB fragment is lost between the cleavage and the 2nd elution step due to precipitation.

### 3.5.4 Cloning of *prnB* in pCIB-HIS and Expression in *Pseudomonas*

The failure to express soluble PrnB in *E.coli* led us to investigate its expression in *Pseudomonas*. PrnA was successfully expressed in *Pseudomonas* despite being completely insoluble in *E.coli*, like PrnB. The *prnB* ORF was cloned into a pCIB-HIS (Table 3.7) vector for *Pseudomonas* overexpression. The gene was amplified by PCR using *Vent Polymerase* and primers containing suitable restriction sites. PCR fragments were cloned into the BamHI HindIII sites of plasmid pFastBac-HT-A. The obtained clones were verified by sequencing. BamHI HindIII fragments of clones containing the correct DNA sequence, were inserted using the same restriction sites to plasmid pCIBHIS as summarized in the following table:

**Table 3.7**

| Gene | Primers and Restriction Site | Plasmids |
|---|---|---|
| *prnB P.fuorescens* | 5'-prnB-B; BamHI / 3'-prnB-H; HindIII | pCIBHIS-*prnB* |
| '' | 5'-prnB-B-TEV; BamHI / 3'-prnB-H-TEV; HindIII | pCIBHIS-TEV-*prnB* |

*E. coli* TOP10 cells were transformed with 5μl of the reaction mix. Plasmid DNA was extracted from tetracycline resistant clones. Plasmids containing DNA fragments of the right size were transformed in *E.coli* S17.1 strain. Plasmid DNA was transferred by conjugation from *E.coli* S17.1 to *P. fluorescens* BL915 ΔORF1-4. Plasmid DNA was extracted from tetracycline resistant clones. Clones carrying the correct plasmids were chosen for protein expression.

### 3.5.5 Activity of prnB in *Pseudomonas* Cells Extracts

Once obtained expression of HIS-PrnB from *Pseudomonas* in *PRN*$^-$*, prnABCD*$^-$ *P. fuorescens* BL915 ΔORF1-4 strain, van Peé's lab was able to show activity of the overexpressed enzyme in crude cell extracts as shown in Fig. 3.6. This result shows that the overexpressed protein was properly translated and folded because the retention of its enzymatic activity.

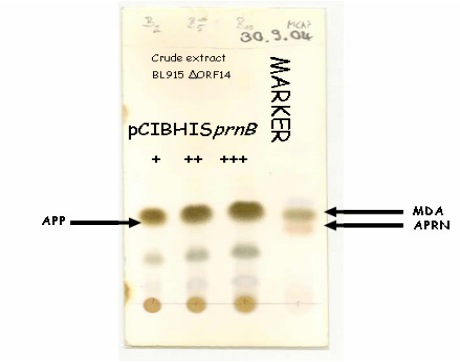



**Figure 3.6**
*Pseudomonas* raw cell extracts were feeded with 7-chloro-TRP, the culture extracted with an equal volume of ethyl acetate, the organic phase dried under vacuum and the residue dissolved in methanol. Thin layer chromatography was performed on silica coated plates with toluene/dioxan (9:1) as the mobile phase pyrrolnitrin (PRN), monodechloroaminopyrrolnitrin (MDA), aminopyrrolnitrin (APRN) and aminophenylpyrrole (APP) were visualized with Erlich's reagent. Increasing amounts of *prnB* extract was incubated with 7-chloro-TRP. Marker: MDA and APRN. MDA production was evident. The lower lighter band is thought to be APP.

### 3.5.5 PrnB activity in feeding experiment

In order to asses all the constructs activity we used the feeding experiment assay developed by van Peé [40; 106]. All the constructs resulted active. Fig 3.7 shows one example.

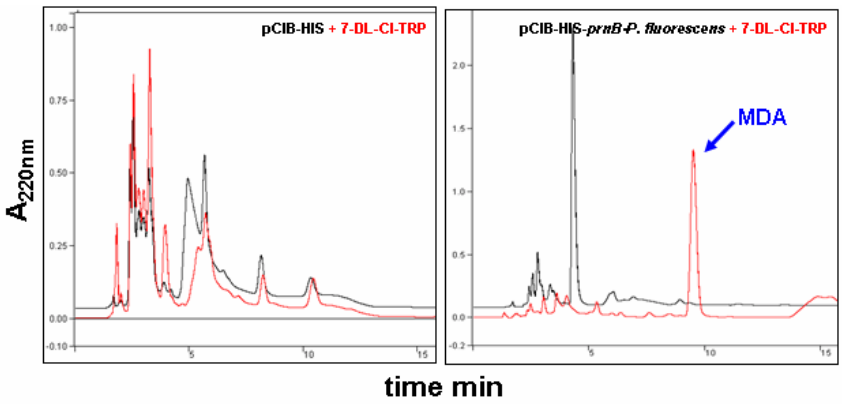



**Figure 3.7**
Ethyl acetate extract HPLC chromatograms from *Pseudomonas* broth culture carrying empty pCIB-HIS or pCIB-HIS-PrnB from *Pseudomonas* in the presence (red) or absence (black) of 7-DL-Cl-Tryptophan. The blue arrow indicates the monodechloroaminopyrrolnitrin (MDA) peak.

### 3.5.6 PrnB purification protocol

PCIB-HIS-*prnB* and pCIB-HIS-TEV-*prnB* from *Pseudomonas* gave over-expression of soluble HIS-tagged PrnB. The overexpressed protein run on SDS PAGE between the

55.4KDa and the 36.5KDa markers compatibly with the expected proteins $M_w$: HIS-PrnB *Pseudomonas* 43628Da; HIS-TEV-PrnB *Pseudomonas* 44512Da; Furthermore protein identity was confirmed with peptide fingerprinting by the *St Andrews University BMS MASS SPECTROMETRY AND PROTEOMICS FACILITY*. Tagged PrnB was purified by a three step procedure from fresh or frozen bacteria cell pellet, consisting of a first Nickel chelating affinity chromatography (See Fig. 3.8A) followed by an anion exchange column and a gel filtration column. The last step was used to polish the protein preparation and exchange the protein buffer to Tris-HCl, PrnB is stable in 10mM Tris-HCl pH7.2 at concentration up to 16mg/ml.

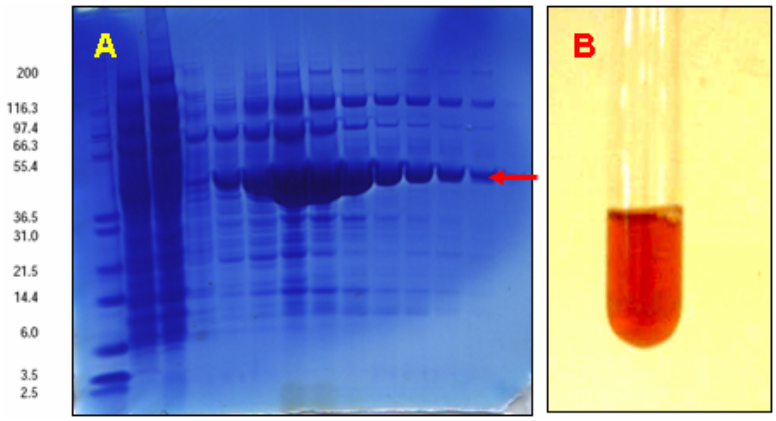



**Figure 3.8**
SDS PAGE showing recombinant hexahistidine tagged PrnB fractions collected after the first affinity column purification. A): *P.fluorescens* PrnB, the protein is highlighted by a red arrow. B): PrnB shows a red color typical of heme containing proteins.

Presence of the TEV cleavage site did not influence the yield or the purification procedure. PrnB resulted to be a heme containing protein (See Fig 3.8B). To characterize PrnB heme cofactor the iron porphyrin was solvent extracted from protein lyophilized by freeze-driyng. Mass Spec analysis of the extracted phase identified a molecule of 615Da compatible with the $M_w$ of heme b (protoporphyrin IX) (See Fig 3.9).

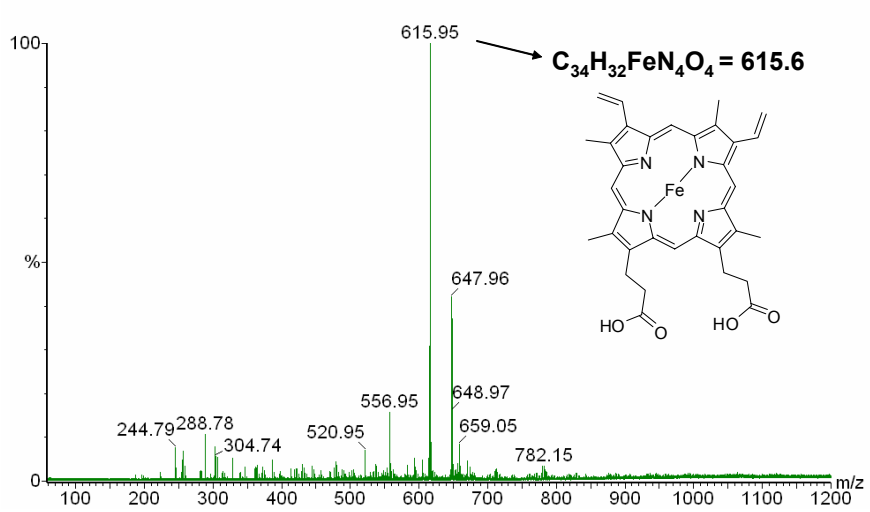



**Figure 3.9**
Mass spectrometry analysis of acetone extracted lyophilized PrnB identified a compound of 615Da, compatible with the presence of heme b as the iron porphyrin cofactor.

Although TEV protease was able to cleave TEV-PrnBs tag, DTT appears to bind tightly to the protein. DTT is an essential component of the TEV protease cleavage buffer, without it the cleavage times become longer. The protein had to be incubated a RT for over a week changing the TEV protease once a day. The resulting PrnB untagged protein failed to give crystals possibly due to aspecific degradation occurred during the long cleavage treatment (See Fig. 3.10) and routine cleavage was abandoned.

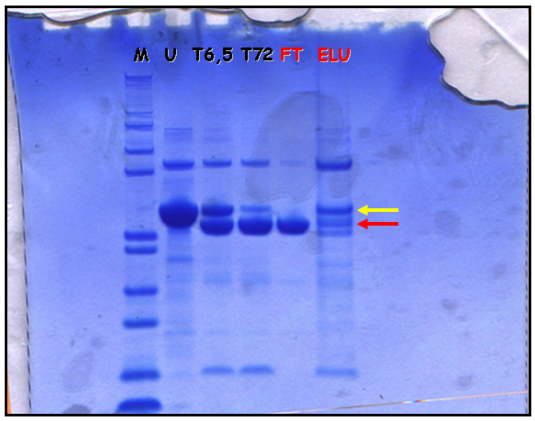



**Figure 3.10**
NuPage 4-12% Bis-Tris gel showing *P.fluorescens* PrnB after the first (black) and after incubation with TEV protease second (red) $Ni^{2+}$ chelating resin column. M: marker (Par 4.4.5.13). U: eluted uncleaved TEV-PrnB; T6,5: after 6.5h of incubation with TEV-protease; T72: after 72h more of incubation with TEV-protease cleave; FT: flow-through after the second $Ni^{2+}$ chelating resin column; ELU: elution after the second $Ni^{2+}$ chelating resin column. The yellow arrow indicate uncleaved TEV-PrnB; the red cleaved TEV-PrnB.

Once *Pseudomonas* over-expression of *P. fluorescens* PrnB revealed the presence of heme as cofactor a second problem emerged. Gel filtration chromatograms of *P. fluorescens* PrnB (See Fig. 3.11) showed us presence of a heterogeneous protein population. Addition of reducing agents, as DTT to the purification buffers prevented protein aggreagation, with an unfortunately heme poisoning side effect. The protein aggregation tendency is better shown by non reducing SDS PAGE gels, where PrnB runs with a ladder effect.

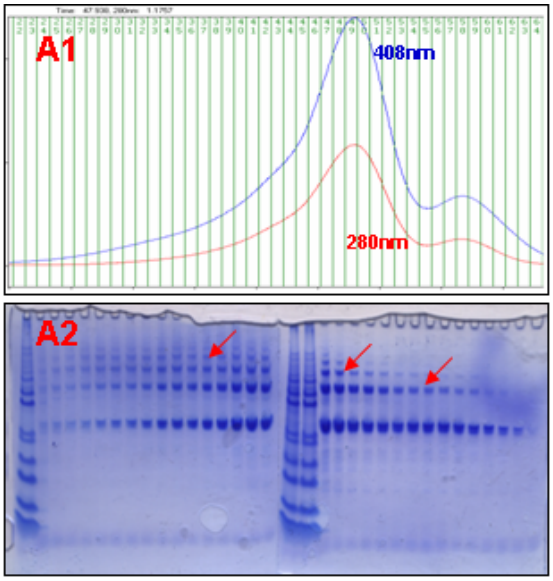



**Figure 3.11**
Gel filtration chromatogram of PrnB from *P. fluorescens* (A1) together with the SDS PAGE of the separated fractions (A2). Chromatograms show the detection at 408nm PrnB Soret peak and 280nm, protein. Red arrows indicated protein identified as *P.fluorescens* PrnB by Mass Spect ID.

### 3.5.7 Cloning and Expression of *M. fulvus* Mx147 *prnB* gene

In order to overcome this problem and try to obtain an homogeneus protein preparation more suitable to crystallization we decide to clone and express the least conserved of the known PrnB genes: the *Myxococcus fulvus* one. Genomic DNA was extracted from *M. fulvus* Mx f147. *PrnB* gene was amplified from genomic DNA using primers 5'-prnB-Myxo-Short and 3'-prnB-Myxo-Short. Different polymerases were tested. Due to the poor quality of the results obtained we decided to amplify a bigger DNA fragment containing *prnB*, using primers 5'-prnB-Myxo-External and 3'-prnB-Myxo-External, see

Fig. 3.12. The reaction product was used to amplify the prnB gene with using primers 5'-prnB-Myxo-Short and 3'-prnB-Myxo-Short. The best results were obtained with the *Vent* polymerase.

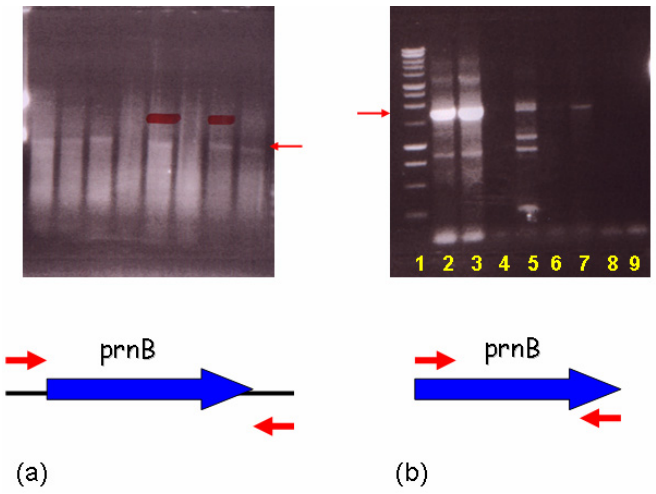



**Figure 3.12**
DNA gels showing *Myxococcus fulvus* Mx f147 *prnB* gene amplification from genomic DNA. A: red arrow shows the amplification product obtained using primer extern to prnB note the high amount of background due to aspecific product. B: red arrow shows differences obtained using different polymerases for PrnB amplification: M1 DNA ladder (*1kbp DNA Ladder, Promega, Cat. No. G5711)*; 1,2: *Vent* polymerase; 3,4*: Thermalace* polymerase; 5,6: *Taq* polymerase; 7,8: *Pfu* polymerase.

The *prnB* ORF was cloned into a pCIB-HIS vector for *Pseudomonas* overexpression. The gene was amplified by PCR using *Vent Polymerase* and primers containing suitable restriction sites. PCR fragments were cloned into the BamHI HindIII sites of plasmid pFastBac-HT-A. The obtained clones were verified by sequencing. BamHI HindIII fragments of clones containing the correct DNA sequence, were inserted using the same restriction sites to plasmid pCIBHIS as summarized in the following table:

**Table 3.9**

| Gene | Primers and Restriction Site | Plasmids |
|---|---|---|
| *prnB M.fulvus* | 5' prnB-B-Mx; BamHI / 3'-prnB-H-Mx; HindIII | pCIBHIS-*prnB-Myxo* |
| " | 5'-prnB-B-TEV-Mx; BamHI / 3'-prnB-H-TEV-Mx; HindIII | pCIBHIS-TEV-*prnB-Myxo* |

PCIB-HIS-*prnB* and pCIB-HIS-TEV-*prnB* from *Myxococcus* gave over-expression of soluble HIS-tagged PrnB. The overexpressed protein run on SDS PAGE between the

55.4KDa and the 36.5KDa markers compatibly with the expected proteins $M_w$: HIS-PrnB *Myxococcus* 45275Da; HIS-TEV-PrnB *Myxococcus* 46127Da. Furthermore proteins identity was confirmed with peptide fingerprinting by the *St Andrews University BMS MASS SPECTROMETRY AND PROTEOMICS FACILITY*. Tagged *M. fulvus* PrnB was purified with the same procedure adopted for *P. fluorescens* PrnB. Fig. 3.13 shows a comparison between *P. fluorescens* and *M. fulvus* PrnB after the first purification step.

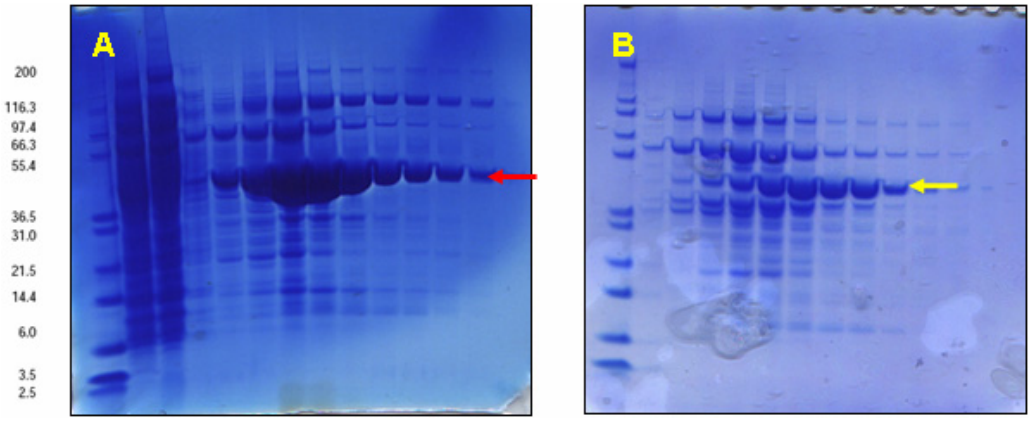



**Figure 3.13**
SDS PAGE showing recombinant hexahistidine tagged PrnB fractions collected after the first affinity column purification. A): *P.fluorescens* PrnB, the protein is highlighted by a red arrow. B): *Myxococcus fulvus* PrnB, the protein is highlighted by a yellow arrow. The difference in xepression is evident if the amount of PrnB is compared with the amount of contaminant protein binding the resin.

Despite differences in protein yield between the *Pseudomonas* and *Myxococcus* gene (See Table 3.10), the amount of recombinant protein obtained after the purification was enough to set up reasonable amount of crystallization trial.

**Table 3.10**

**Total protein amount at each purification step**

| PrnB: | *P. fluorescens* | *M. fulvus* |
|---|---|---|
| LYSATE ON $Ni^{2+}$ COLUMN | 650mg - **100%*** | 1800mg - **100%*** |
| OUTPUT ELUTED PRNB FRACTIONS FROM $Ni^{2+}$ COLUMN | 39mg - **6%*** | 58mg - **3%*** |
| AMOUNT COLLECTED AFTER GEL FILTRATION COLUMN | 5mg - **2%*** | 5mg - **0.27%*** |

* compared to the amount loaded on the first column

Gel filtration chromatograms showed an overall lower tendency for *M. fulvus* PrnB to produce aggregates, but even in this case the final protein preparation was not homogeneous, see Fig. 3.14.

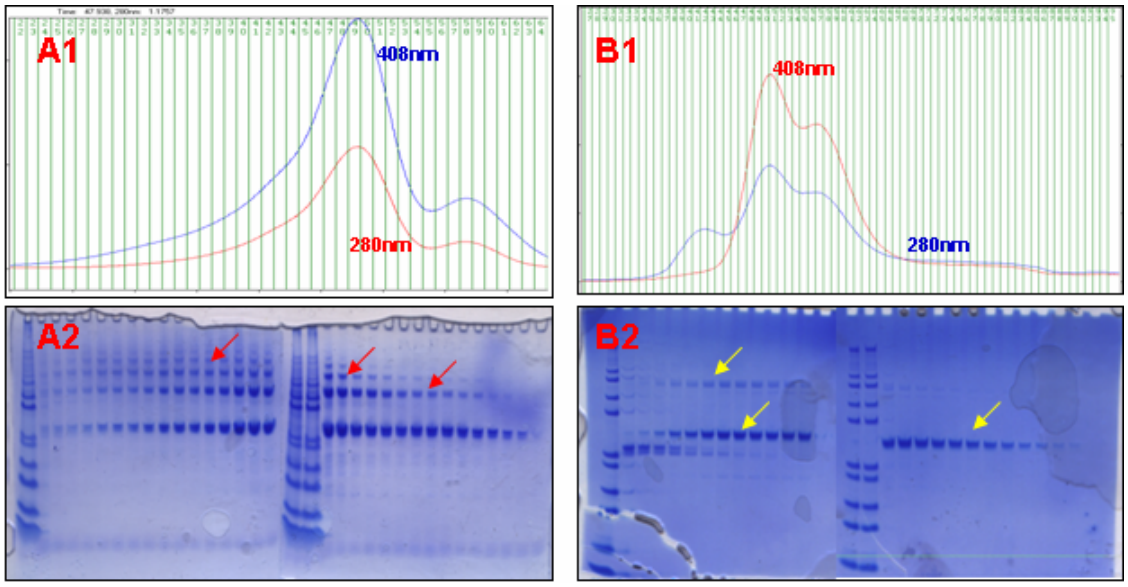



**Figure 3.14**
Gel filtration chromatogram of PrnB from *P. fluorescens* (A1) and *M. fulvus* (B1) together with the SDS PAGE of the separated fractions (A2, B2). Chromatograms show the detection at 408nm PrnB Soret peak and 280nm, protein. Red arrows indicated protein identified as *P. fluorescens* PrnB by Mass Spect ID. Yellow arrows indicated protein identified as *M. fulvus* PrnB by Mass Spect ID. Note how the laddering effect is more pronounced in *P. fluorescens* PrnB, but the oligomerization is present in *M. fulvus* too as indicated by gel filtration chromatogram.

Reducing agents were able to reduce the aggregations problem but poisoned the heme. This result suggested the formation of aspecific disulfide bonds as already observed in *P. fluorescens* PrnB. Analysis of all the known PrnB homologues sequences showed us differences in the number and pattern of cysteines present across the homologues; with the only conserved cysteine being *P.fluorescens* BL915 Cys175 (See Fig. 3.15). *Myxococcus fulvus* PrnB cysteine distribution, with two of the three cysteines next to each other, can explain its lower tendency to aspecific aggregation.

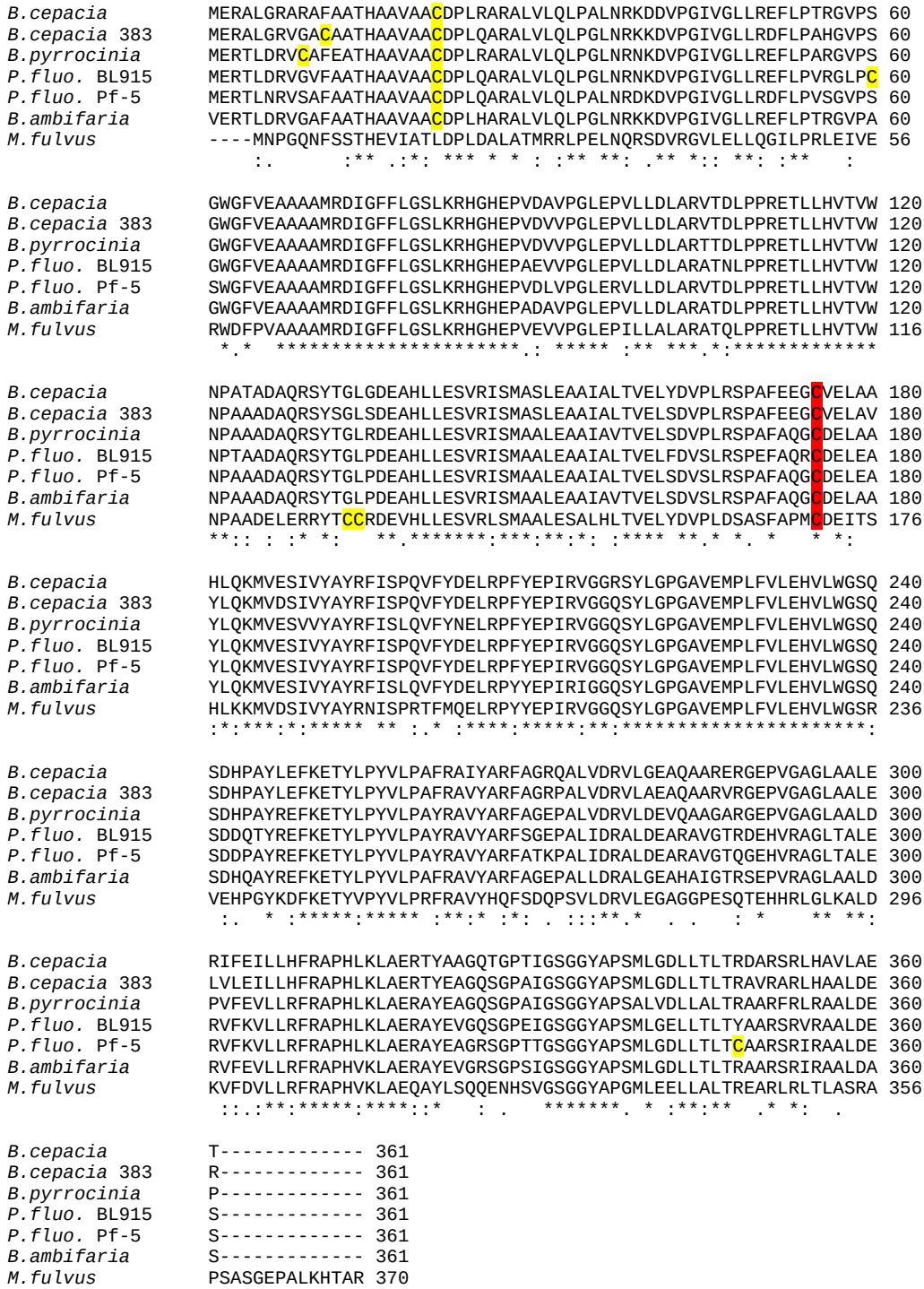



**Figure 3.15**
Clustal alignment of PrnB protein homologues. In yellow non conserved cysteine residues. In red the conserved *P. fluorescens* BL915 C175 residue. * identifiy 100% conserved residue.

### 3.5.8 PrnB Site Directed Mutagenesis

Although *Myxococcus* PrnB showed less protein aggregation it gave again a heterogeneous protein population, difficult to separate for crystal trial experiments. For crystallization we wish to prevent disulphide bond formation. A literature search showed that aggregation by disulphide bond formation has already been encountered during the crystal structure solution of the heme containing neuroglobin protein class [107; 108]. We had to test if the presence of disulfide bonds was a physiological protein feature or resulted because of the overexpression/purification procedure. Furthermore we had to check if the conserved cysteine residue is essential for enzymatic activity. Both *M. fulvus* C129S C130S C171S PrnB and *P. fluorescens* C21S C60S C175S PrnB proteins were successful generated and purified (See Fig. 3.16).

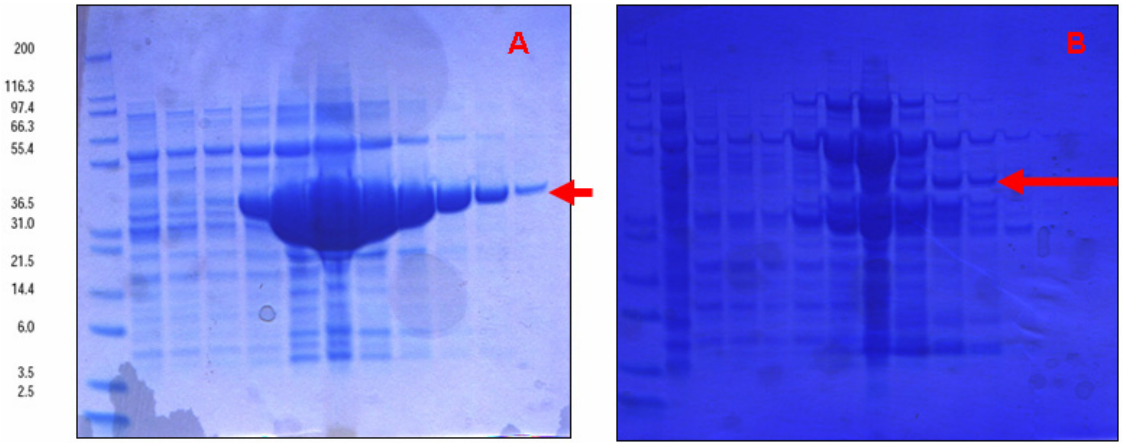



**Figure 3.16**
SDS PAGE showing recombinant hexahistidine tagged triple cysteine mutants PrnB fractions collected after the first affinity column purification. A): *P.fluorescens* C21S C60S C175S PrnB, the protein is highlighted by a red arrow. B): *Myxococcus fulvus* C129S C130S C171S PrnB, the protein is highlighted by a red arrow. The difference in expression is evident if the amount of PrnB is compared with the amount of contaminant protein binding the resin.

The *Myxococcus fulvus* protein was expressed with a much lower yield compared to the wild type protein. No major differences were detected between *P.fluorescens* triple mutants PrnB and the wild type protein. Table 3.11 show a more detailed protein yield comparison between the wild type and mutant PrnBs.

**Table 3.11**

**Total protein amount at each purification step**

| | *P. fluorescens* | | *M. fulvus* | |
|---|---|---|---|---|
| **PrnB:** | **WT** | **Cys ► Ser** | **WT** | **Cys ► Ser** |
| **LYSATE ON $Ni^{2+}$ COLUMN** | 650mg - **100%*** | 1700mg - **100%*** | 1800mg - **100%*** | 1560mg - **100%*** |
| **OUTPUT $Ni^{2+}$ COLUMN** | 39mg - **6%*** | 270mg - **16%*** | 58mg - **3%*** | 35mg - **2.2%*** |
| **OUTPUT GEL FILTRATION COLUMN** | 5mg - **2%*** | 100mg - **5%*** | 5mg - **0.27%*** | 0.250mg - **0.01%*** |

* compared to the amount loaded on the first column

The *Myxococcus fulvus* PrnB produced was not enough for other experiment than its molecular weight assessment on analytical gel filtration. *P.fluorescens* PrnB was used for crystallization trial (See Chapter 4). PrnB oligomeric state was assessed by analytical gel filtration as expected both PrnB mutants run were consistent with the behaviour of a monomeric protein. An example is shown in Fig 3.17. Most importantly the protein preparations appear to be homogeneous, see Fig. 3.18.

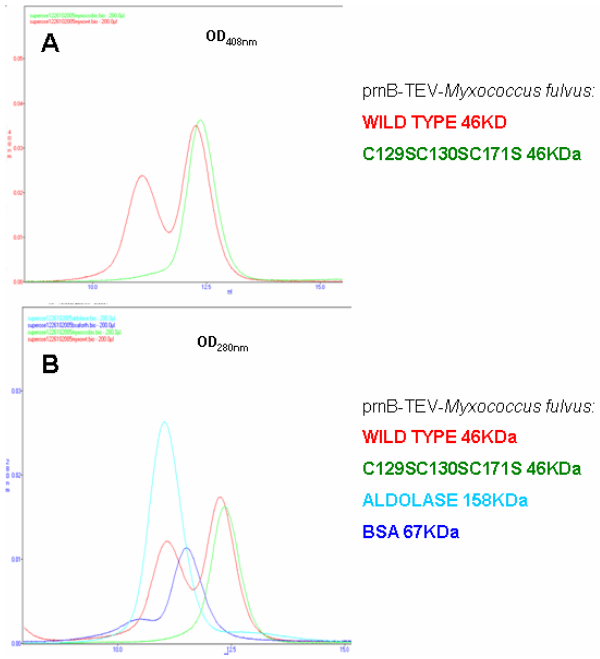



**Figure 3.17**
Analytical gel filtration column Superose 12 HR10/30® was used together with protein weight marker from Amersham Bioscience. Panel A shows differences between wild type *M. fulvus* PrnB (red trace) and the triple cysteine mutant (green trace). Panel B shows both of them together with Aldolase 158KDa marker and BSA 67KDa marker.

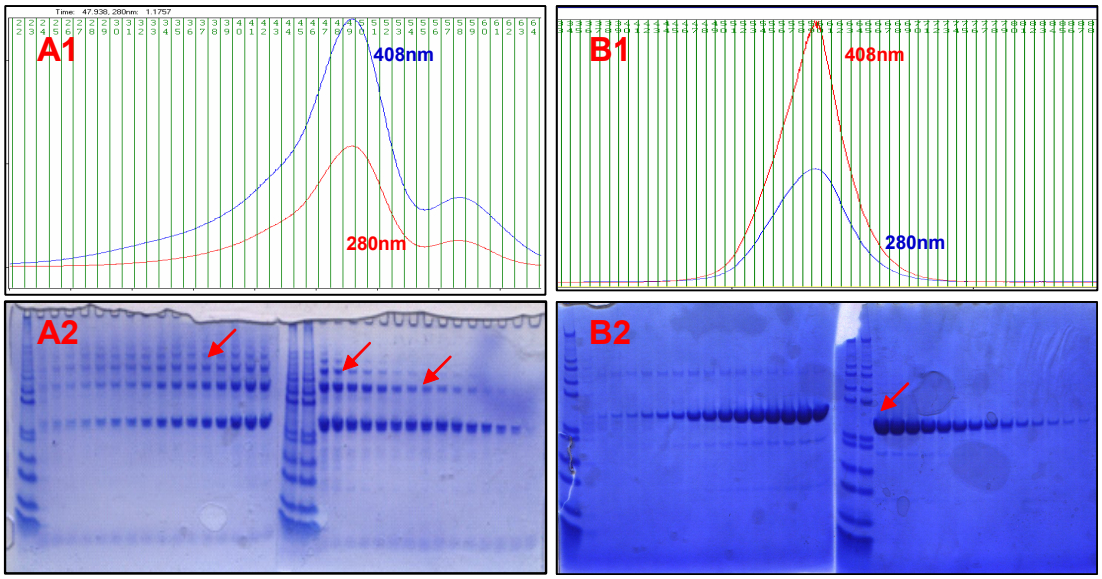



**Figure 3.18**
Preparative preparations gel filtration chromatogram of PrnB from *P. fluorescens* wild type (A1) and triple mutant (B1) together with the SDS PAGE of the separated fractions (A2, B2). Chromatograms show the detection at 408nm PrnB Soret peak and 280nm, protein. Red arrows indicated protein identified as *P. fluorescens* PrnB by Mass Spect ID.

In order to role out involvement of any of the mutated cysteines in the enzymatic activity of PrnB mutants we tested them using the feeding experiment assay developed by van Peé [40; 106]. All the constructs resulted active as shown in Fig 3.19. These results exclude any physiological role of the oligomeric protein aggregates and any involvement of the conserved *P. fluorescens* Cys175 into the enzyme catalytic cycle.

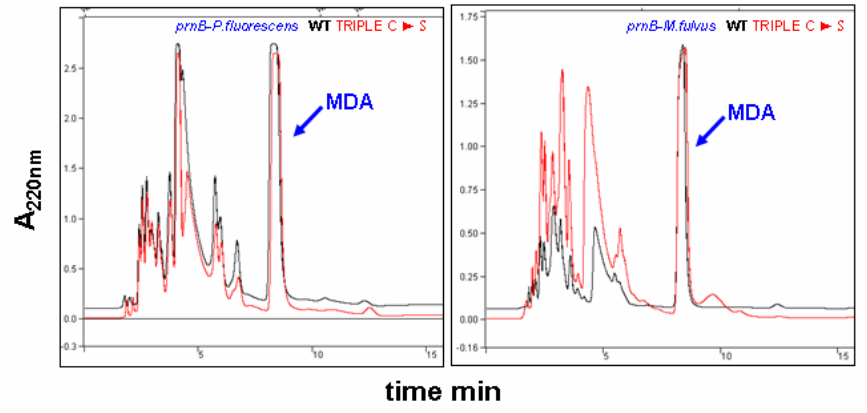



**Figure 3.19**
Ethyl acetate extract HPLC chromatograms from *Pseudomonas* broth culture carrying wild type pCIB-HIS-PrnB (black trace) or triple cysteine mutant pCIB-HIS-PrnB (red trace) from *Pseudomonas* and *Myxococcus* in the presence of 7-DL-Cl-TRP. The blue arrow indicates the characteristic monodechloroaminopyrrolnitrin MDA peak produced by both the wild type and the mutated enzyme.

To confirm PrnB as the unique enzyme responsible for 7-Cl-tryptophan biotransformation to monodechloroaminopyrrolnitrin pCIB-HIS-PrnB triple mutant was transferred by conjugation into *Pseudomonas putida* KT2440 [99; 100], a *Pseudomonas* species unrelated to *P. fluorescens*. This strain resulted to express soluble PrnB, see Fig. 3.20.

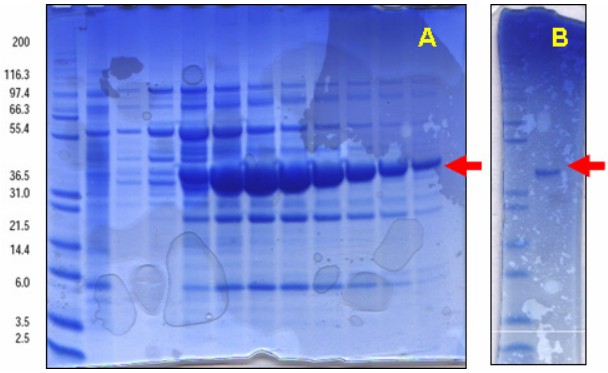



**Figure 3.20**
SDS PAGE showing recombinant hexahistidine tagged triple cysteine mutants PrnB from *P.putida* KT2440 carrying the pCIB-HIS-prnB plasmid . A) fractions collected after the first affinity column purification. B) final ourified protein ready for crystallization trials. The red arrow indicates PrnB.

Feeding experiment with 7-Cl-tryptophan demonstrated the activity of PrnB in this *Pseudomonas* species too, see Fig. 3.21.

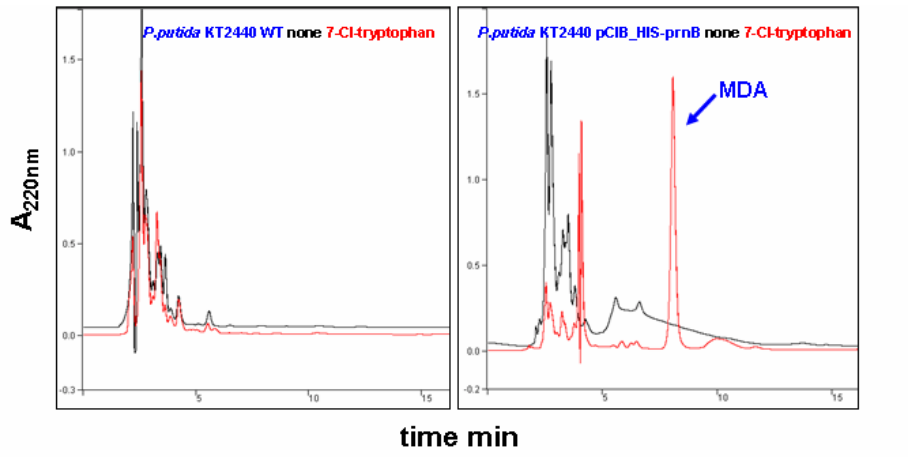



**Figure 3.21**
Ethyl acetate extract HPLC chromatograms from *Pseudomonas putida* KT2240 wyld type broth culture or carrying pCIB-HIS-PrnB triple cysteine mutant alone (black trace) or in the presence of 7-DL-Cl-TRP (red trace). The blue arrow indicates the characteristic monodechloroaminopyrrolnitrin MDA peak produced by active PrnB enzyme.

## 3.6 Discussion

PrnB was successfully over-expressed and purified from *Pseudomonas* culture. The protein resulted to contain a heme cofactor. Given the nature of the heme cofactor and the homology between IDO and PrnB substrates, we went back and reconsider possible similarity between the reactions performed by the two enzymes. IDO is involved in the initial and rate-limiting step of L-TRP catabolism in the kynurenine pathway [109] and given the difference between IDO and PrnB reaction products (See Fig. 3.21) only the comparison of the two crystal structures could give us better clue about a possible partially conserved reaction mechanism.

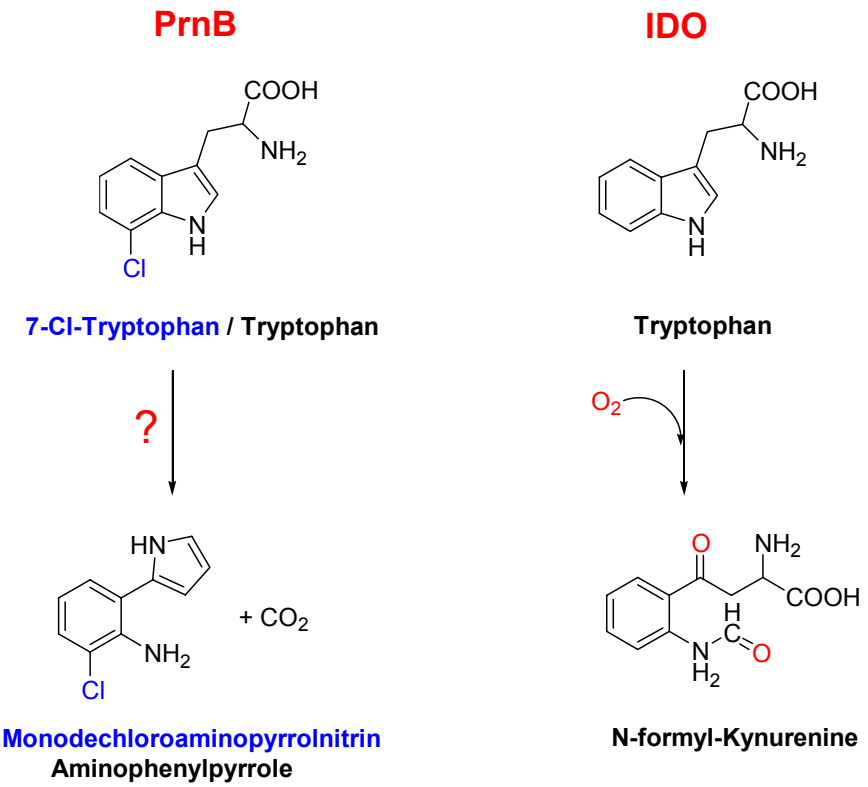



**Figure 3.21**
PrnB and IDO catalyzed reactions.

Over-expression of PrnB into its natural host produce a heme loaded protein while *E.coli* over-expression failed to provide the enzyme with its cofactor resulting in aggregation and precipitation of the unfolded polypeptide. This result underlines once again the limits of *E.coli* as universal tool for protein over-expression even in the prokaryote kingdom.

# Chapter 4

# PrnB Structure Determination and Analysis

## 4.1 Summary

In this chapter I describe the crystallization of purified PrnB from *P. fluorescens* triple cysteine mutant in presence of both D and L tryptophan. The crystals diffract up to 1.7Å with a cell dimension of: a=68.6Å b= 79.5Å c= 92.7Å $\alpha$=$\gamma$= 90.0° ß= 103.8°, and a C2 space group. The structure of PrnB in complex with L-tryptophan has been determined by multiwavelength experiment at the iron edge. The structure of L-tryptophan complex was used to phase the D-tryptophan complex and an additional native dataset. In both structures, tryptophan is bound close to the heme site, but the enantiomers are oriented in two different ways. The structure confirms PrnB is closely relaed to IDO despite only 10% amino acidic sequence similarity. The tryptophan complexes although locating the enzyme active site are probably not relevant to catalysis. Molecular modeling has been used derive a model of enzyme substrate complex. An enzyme mechanism related to that of IDO is proposed.

## 4.2 Introduction

Once PrnB protein suitable for crystallization was obtained we started to set up crystallization trials. Given the fact that enantiomeric pure L or D 7-Cl-tryptophan is not commercially available (even the racemic mixture is not commercialized) we decide to set up co-crystallization experiment with enzyme preincubated with D and L tryptophan. Van Peé *et al*. reported PrnB to be active on tryptophan too [21]. It has to be noted that PrnB substrate stereoselectivity have not been clarified so far and that PrnA, that catalyzes the first step of pyrrolnitrin biosynthesis, is active both on D and L tryptophan enantiomers. Furthermore feeding experiment with *Pseusodomonas c*ulture have demonstrated that both the enantiomer are pyrrolnitrin precursors [61]. Interestingly D-tryptophan has the ability to enhance the antibiotic production. Without a robust *in vitro* assay the involvement of aminoacid racemase in the interconversion of the tryptophan isomers can not be excluded. The PrnB crystal structure could help us to propose a catalytic mechanism for the enzyme and at the same time provide information to set up successful enzyme *in vitro* assays.

## 4.3 Materials and Method/Results

### 4.3.1 prnB crystallization trials

PrnB resulted stable at concentration ranging from 8mg/ml to 16mg/ml. For co-crystallization experiment concentrated protein was incubated with L-tryptophan, D-tryptophan and pure 7-L-Cl-tryptophan at room temperature with gentle agitation for 30min before the excess of TRP was separated by centrifugation. Crystal trials were set up at the Scottish Structural Proteomic Facility with a nanodrop crystallization robot (Cartesian HoneyBee) on sitting drop plate (Hampton Research CrystalQuick™ 96 wells, 4 µl square drop well, 3 drop well) using crystallization sparse matrix kits from a range of different providers (200nl protein solution plus 200nl of precipitant). In Fig. 4.1 a SDS PAGE of PrnB used for crystallization trial is shown.

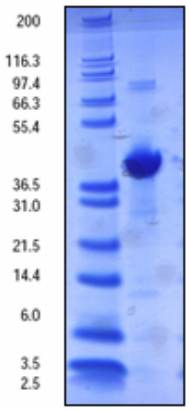



**Figure 4.1**
NuPage 4-12% Bis-Tris gels showing in lane 1 protein ladder; in lane 2 *P. fluorescens* BL915 Triple cysteine mutant PrnB used for crystallization.

Plates were imaged as a part of the Hamilton-Thermo Rhombix System of the the Scottish Structural Proteomic Facility. See Fig. 4.2.

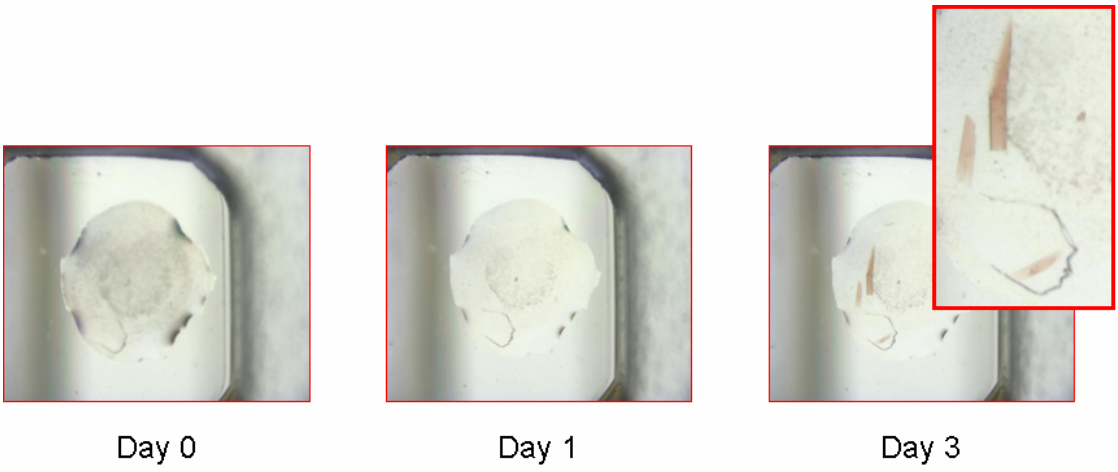



**Figure 4.2**
Images of the condition number 80 from crystal screen THE PEGs (NEXTAL) containing 0.2 M Magnesium sulfate heptahydrate, 20% w/v PEG 3350. Triple cysteine mutated uncleaved TEV-PrnB protein is at 8mg/ml and saturated with L-tryptophan. Crystals appeared three days after the plate was set up. Drop size is 200nl.

Protein crystals were obtained with condition number 80 (0.22 M Magnesium sulfate heptahydrate, 16% w/v PEG 3350) of The PEGs crystal screen (NEXTAL) with PrnB at 8mg/ml and 16mg/ml using both a 1:1 or a 2:1 protein to precipitant ratio in presence of saturating amount of L-TRP. No crystals were obtained incubating the protein with pure 7-L-Cl-TRP, provided by Professor Robert S. Phillips from The University of Georgia, or without tryptophan.

### 4.3.2 PrnB crystal optimization

Crystallization conditions had to be translated and optimized from nanodrop to a bigger drop size. Hanging drop crystal plate produced too many crystals per drop. The best results were obtained using sitting drop plates (Hampton Research CrystalClear Strips$^{TM}$ 96 wells plates) with 1.5µl protein solution plus 1.5µl of precipitant. Expansion of the PEG 3350 and $MgSO_4$ concentration did not change the form of the crystal that tended to grow as cluster of plates/rod as shown in Fig 4.3.

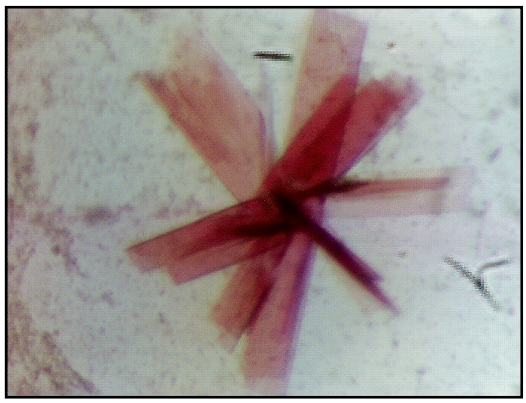


**Figure 4.3**
Images of the condition number 80 from crystal screen THE PEGs (NEXTAL) containing 0.2 M Magnesium sulfate heptahydrate, 20% w/v PEG 3350. Triple cysteine mutated uncleaved TEV-PrnB protein is at 8mg/ml and saturated with L-tryptophan. Crystals appeared one week after the plate was set up. Drop size is 3µl.

Green crystals of the same shape were obtained from PrnB incubated with D-Tryptophan prior to crystallization.

### 4.3.3 PrnB data collection

Given the small size of PrnB crystals and their tendency to grow in clusters we were not able to collect any data in house. The first data were collected at The Daresbury Synchrotron Radiation Source on Station 9.6. Plates were detached from plate clusters obtained from protein incubated with L-tryptophan and cryoprotected with 20% (2*R*,3*R*)-(−)-2,3-butanediol (Cat. N# 237639), Sigma. Two datasets were collected from different crystals. The first dataset with a resolution of 2.4Å consisted of 360 images at 0.5° oscillation. The second dataset with a resolution of 1.75Å consisted of 240 images at 0.5° oscillation. Images were indexed in MOSFLM [86], which identified the higher resolution dataset as $C222_1$ space group and the lower resolution one as C2. Reflection were sorted and merged with SCALA [87] In table 4.2 statistics are reported for the $C222_1$ dataset identified as “native”. PrnB crystallizes in two different space groups in our crystallization condition. In order to solve PrnB structure MAD datasets were collected subsequently at the European Synchrotron Radiation Facility on ID29. A three wavelength MAD dataset was collected on a plate from a crystal grown in presence of L-tryptophan

(Inflection 180 images 1° degree oscillation, Edge 360 images 0.5° oscillation, Remote 180 images 1° oscillation). A further dataset was collected from a plate grown in presence of D-tryptophan (360° images 0.5° oscillation). Both crystals were again cryoprotected with 20% (2*R*,3*R*)-(−)-2,3-Butanediol. Images were index in MOSFLM [86], data were sorted and merged with SCALA [87]. Crystals resulted to belong to space group C2; statistics are reported in table 4.2. Analysis of the Matthews coefficients [91] [90] for both the crystal space groups identified one monomer in the asymmetric unit, see table 4.1. Fig. 4.4 shows the mounted crystal during the MAD experiment and one of the collected images.

**Table 4.1**

**Matthews Coefficient Calculation**

| SPACE GROUP | CELL VOLUME Å | $N_{molecule}$ / Asymmetric Unit | Matthews Coefficient | % SOLVENT | P* |
|---|---|---|---|---|---|
| C2 | 492928.594 | 1 | 2.87 | 57.11 | 0.99 |
| $C222_1$ | 703515.875 | 1 | 2.05 | 14.22 | 0.01 |

**(P*: probability across all resolution ranges)**

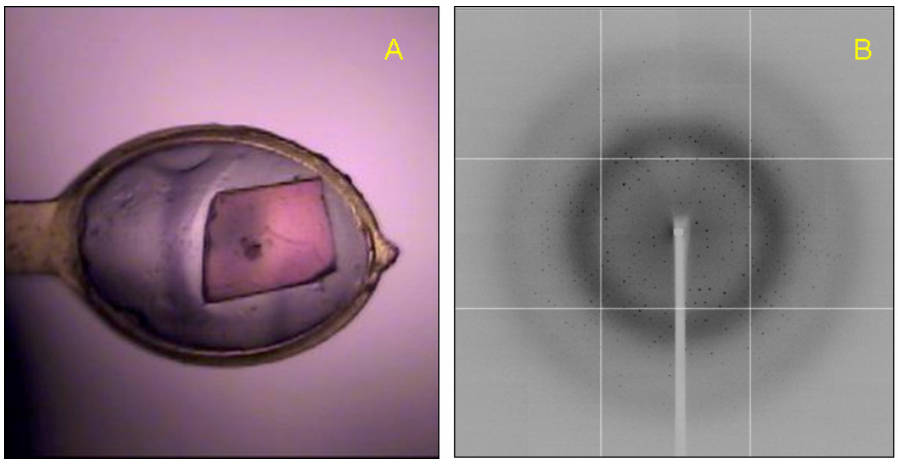



**Figure 4.4**
A: a picture of L-tryptophan PrnB crystal used for the MAD experiment. B: one of the collected images.

**Table 4.2**

| DATA COLLECTION | Inflection | PrnB plus L-TRP Edge | Remote | plus-D-TRP | native |
|---|---|---|---|---|---|
| | **Overall (OuterShell)** | **Overall (OuterShell)** | **Overall (OuterShell)** | **Overall (OuterShell)** | **Overall (OuterShell)** |
| Resolution limit (Å) | 51.10 – 2.00 | 51.10 – 2.00 | 51.10 – 2.00 | 50.97 – 1.79 | 38.26 – 1.75 |
| Highest Shell (Å) | 2.11- 2.00 | 2.11- 2.00 | 2.11- 2.00 | 1.79 – 1.70 | 1.84 – 1.75 |
| Wavelength | 1.74070 | 1.73890 | 1.722 | 0.97620 | 1.4880 |
| Unit-Cell | | a=68.6 Å<br>b= 79.5 Å<br>c= 92.7 Å<br>α=γ= 90.0°<br>ß= 103.8° | | a=68.2 Å<br>b= 79.5 Å<br>c= 92.3 Å<br>α=γ= 90.0°<br>ß= 103.4° | a=67.69 Å<br>b= 80.12 Å<br>c= 129.51 Å<br>α=ß= γ=90.0° |
| Space Group | | C2 | | | C2221 |
| Unique reflections | 31741 (4521) | 31860 (30) | 31836 (4538) | 51494 (7379) | 35216 (4879) |
| Multiplicity (%) | 3.6 (3.7) | 7.1 (7.3) | 3.6 (3.7) | 3.7 (3.8) | 4.4 (4.0) |
| Completeness (%) | 97.0 (94.7) | 97.1 (94.8) | 97.1 (94.8) | 97.5 (96.3) | 98.1 (95.3) |
| Rmerge | 0.059 (0.250) | 0.066 (0.298) | 0.060 (0.324) | 0.076 (0.886) | 0.082 (0.193) |
| I/σ (Mean((I)/sd(I)) | 16.5 (4.9) | 21.5 (6.3) | 15.4 (3.8) | 12.0 (1.5) | 16.0 (4.2) |
| Anomalous Completeness | 75.1 (66.4) | 97.0 (94.8) | 75.8 (67.1) | | |
| f'/f'' | -21 / 13 | -26 / 7 | -17 / 5 | | |
| Δano/σΔano | 1.05 | 1.26 | 1.1 | | |

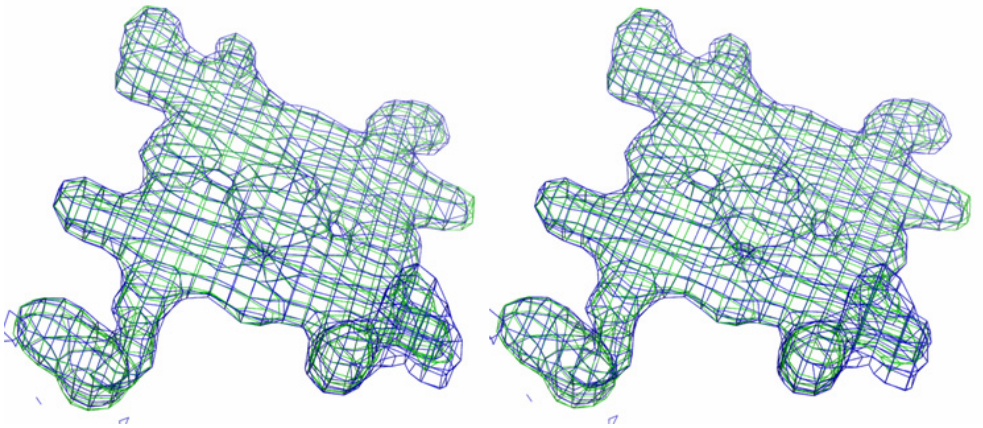
### 4.3.4 PrnB structure solution and refinement

The three-wavelength, pre–merged MAD data, cut to 2.8Å, were used to determine the position of the anomalous scatter Fe atom in SOLVE [110]. The program refined anomalous scattering factors and searched for one single Fe atom, corresponding to one heme/monomer. The SOLVE phases calculated in C2 were used as input to RESOLVE for solvent flattening of the resulting electron density and automatic model building [111; 112]. The phases from RESOLVE was input into ARP/wARP [113] together with PrnB sequence and the intensities from the Edge dataset. ARP/wARP was able to build amino acids 45 to 397 in a single chain with a gap between residues 364 to 378 with a connectivity index of 0.98. ARP/wARP model was refined using cycles of manual refinement with WinCoot [51] and Refmac5 [92; 93] (CCP4 package) against both the Edge dataset and the D-tryptophan containing dataset, See table 4.3 for the details of refinement. The refined D-tryptophan containing model was used to solve the $C222_1$ dataset by molecular replacement with PHASER [88; 89] using PrnB as the search model. A solution was found with 1 monomer per asymmetric unit cell, the model was refined using cycles of manual refinement with WinCoot [51] and Refmac5 [92; 93] (CCP4 package), See table 4.3 for the details of refinement.

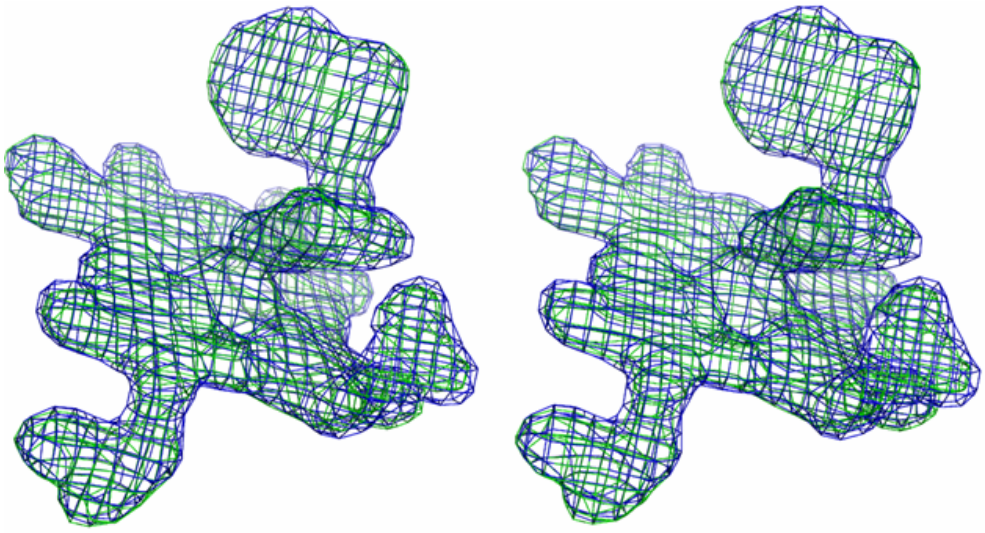
**Table 4.3**

| PrnB: | **plus L-TRP** | **plus-D-TRP** | **native** |
|---|---|---|---|
| $\mathbf{R_{cryst}}$ | 0.169 | 0.194 | 0.183 |
| $\mathbf{R_{free}}$ | 0.216 | 0.214 | 0.231 |

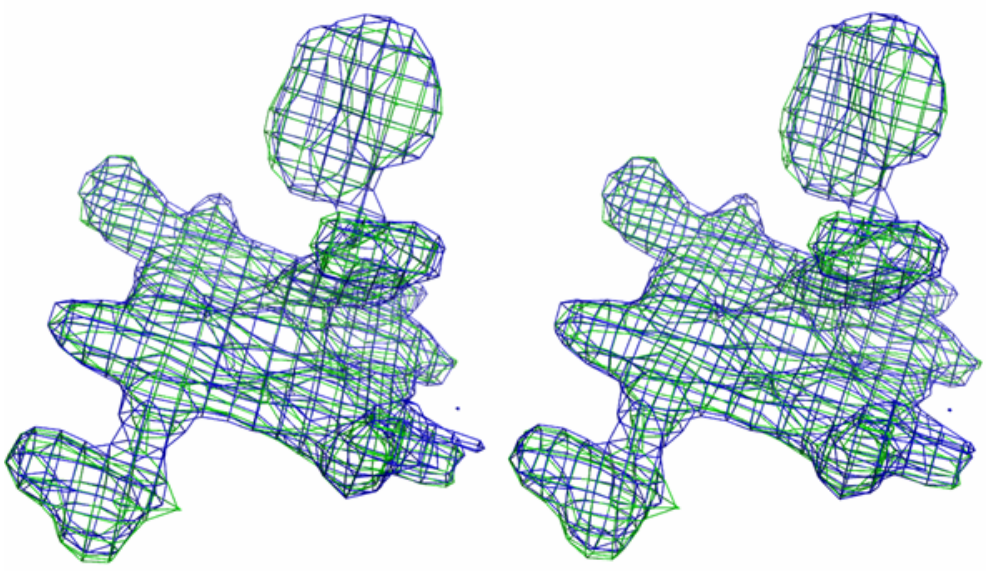
The heme b electron density was clear for all the three models. The green crystal grown with protein incubated with D-tryptophan showed $F_O$-$F_C$ electron density at the sixth position of the heme b iron. This density was modeled as D-tryptophan. Crystal grown in the presence of L-tryptophan and belonging to the C2 space group also showed $F_O$-$F_C$ electron density at the sixth position of the heme b iron. This density was modeled as L-tryptophan. For the $C222_1$ crystal, despite being grown in presence of L-tryptophan, no tryptophan molecules were located next to the heme b in the resulting model. See Fig. 4.5.



**A:** native $C222_1$ crystals



**B:** C2 D-tryptophan containing crystal



**C:** C2 L- tryptophan containing crystal

**Figure 4.5**
Unbiased PrnB $2F_O$-$F_C$ at 1σ in blue and $F_O$-$F_C$ at 3σ in green electron density maps around the heme and tryptophan positions: A: $C222_1$ native crystal; B: C2 D-tryptophan containing crystal; C: C2 L- tryptophan containing crystal. See Fig. 4.11, 4.12 and 4.13 for the stick representation of the modeled heme cofactor and ligands.

## 4.4 Discussion

### 4.4.1 Overall Structure

As predicted PrnB structure consists only of helices. The final refined structure covers the TEV-PrnB construct from residue 45 to 400 that correspond to residues 5 to 360 of the *Pseudomonas* coding sequence. Residues 364-374 in the native model, 366-374 in the D-tryptophan complex and 365-375 in the L-tryptophan complex are disordered. Secondary structure assignment was carried out using STRIDE [96] web base software (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py) and was confirmed by visual inspection of the model, see Fig. 4.6 and 4.7.

```
                .         .         .         .         .
45   LDRVGVFAATHAAVAASDPLQARALVLQLPGLNRNKDVPGIVGLLREFLP   94
       HHHHHHHHHHHHHH TTTTHHHHHHHHHHHHHH  HHHHHHHHHH  T
            α1                  α2          α3
                .         .         .         .         .
95   VRGLPSGWGFVEAAAAMRDIGFFLGSLKRHGHEPAEVVPGLEPVLLDLAR  144
     TTT TTTT HHHHHHHHHHHHHHHHHHHHH   GGG TTTHHHHHHHHHH
                      α4                α5       α6
                .         .         .         .         .
145  ATNLPPRETLLHVTVWNPTAADAQRSYTGLPDEAHLLESVRISMAALEAA  194
     HH       HHHHHTTT  TTTTT TTTT HHHHHHHHHHHHHHHHHHHH
                α7                       α8
                .         .         .         .         .
197  IALTVELFDVSLRSPEFAQRSDELEAYLQKMVESIVYAYRFISPQVFYDE  244
     HHHHHHHH  TTTTHHHHHHHHHHHHHHHHHHHHHHHHHH  HHHHHHH
                            α9                    α10a
                .         .         .         .         .
245  LRPFYEPIRVGGQSYLGPGAVEMPLFVLEHVLWGSQSDDQTYREFKETYL  294
     HGGG   BBBTTBBB    GGG  HHHHHHHHHHTTT  HHHHHHHHHHG
     α10b    β1   β1    α11    α12               α13a-b
                .         .         .         .         .
295  PYVLPAYRAVYARFSGEPALIDRALDEARAVGTRDEHVRAGLTALERVFK  344
     GG  HHHHHHHHHHTTT  HHHHHHHHHHHHTTTTHHHHHHHHHHHHHH
              α14             α15               α16
                .         .         .         .
345  VLLRFRAPHLKLAERAYEVG------PSMLGELLTLTYAARSRVRAALDE  400
     HHHHHHHHHHHHHHHHHH        HHHHHHHHHHHHHHHHHHHHH
                                      α17
```

**Figure 4.6**
PrnB secondary structure assignment output file from the web based version of STRIDE [96] (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py). Residues are numbered as in the refined structure. T = turns, H = α-helix, G = $3_{10}$-helix, B = β-strand. Shaded in yellow is the gap containing the disordered region from aa 365 to aa 376, in purple the three Cysteine residues mutated to Serine for crystallization purpose.

Analysis of protein protein interfaces in the entire crystal using the web based PISA Server [95] software (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) failed to identify any possible relevant protein protein interaction, confirming the PrnB monomeric character assessed by gel filtration. Seventeen helices can be identified in PrnB model. The heme binding helices are at the protein C-terminal, with the heme proximal ligand histidine on helix α16. A loop between helices α10 and α11 is situated on top of the heme prosthetic group. The sixth coordination position of the heme b is occupied by the tryptophan ligand in the two tryptophan containing models and by a water molecule in the native model.

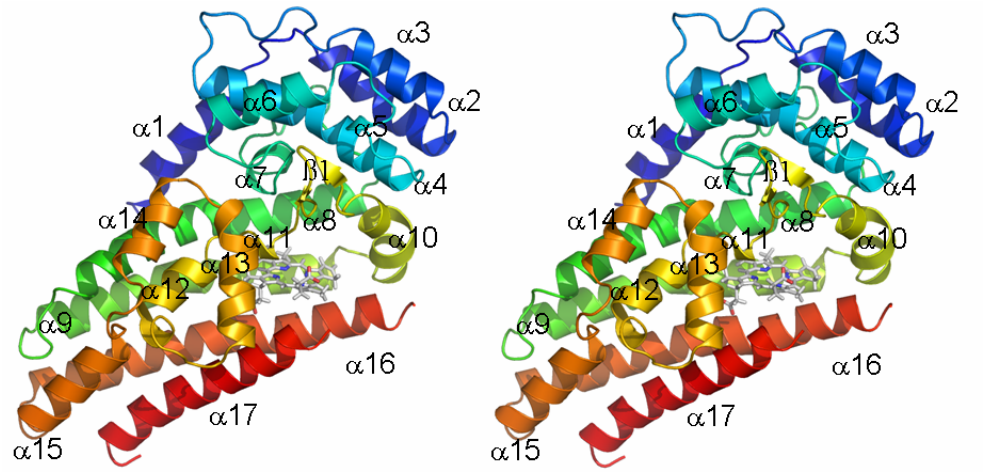



**Figure 4.7**
Cartoon representation of PrnB structure. Each of the 17 identified helices is labeled (α = helix, β = sheet). Heme group is in white. The N terminal is at α1the C-terminal at α17.

Residues 364-376 connecting helix α16 to helix α17 are likely to be part of a flexible loop region outside the heme pocket. The three cysteine residues that were mutated are far from the protein active site. C61 is located in the connecting loop between helices α1 and α2; C100 is located in the connecting loop between helices α3 and α4 and C215 is at the N-terminal of helix α9. C61 and C100 are the two best candidates for aspecific bonding in PrnB concentrated protein solution because of their localization on solvent exposed flexible loops, see Fig. 4.8. Mutation of solvent exposed cysteine in heme containing protein to obtain material suitable for crystallization has precedent in the murine neuroglobin structure [108] (PDB code 1W92) and human neuroglobin structure [107] (PDB code 1OJ6). No major

rearrangements were observed between the two ligand bonded models or between them and the native structure.

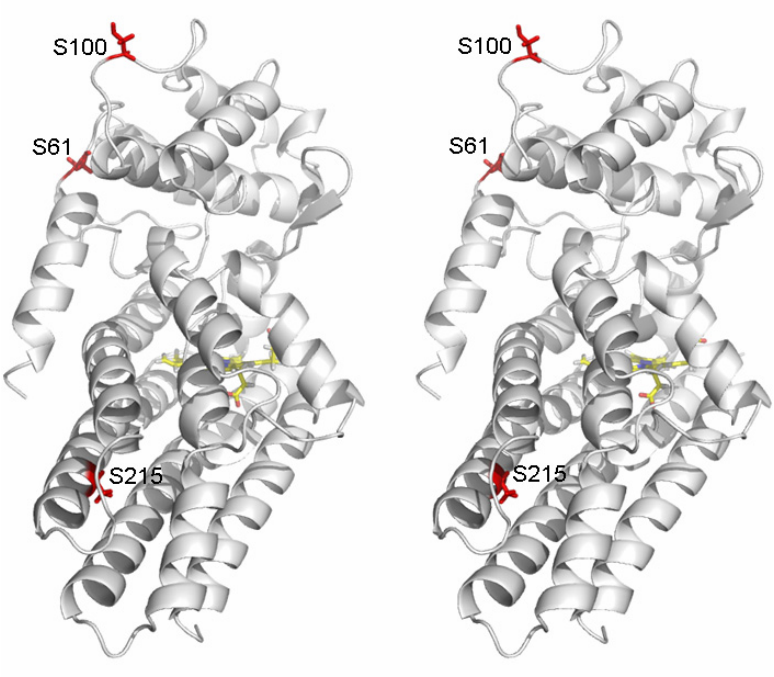



**Figure 4.8**
Three cystein residues had to be mutated to serin residues in order to obtain protein material suitable for PrnB crystallization. In this picture showing a cartoon model (in gray) of PrnB the three mutated serin residues are highlighted in red. Heme b is colored in yellow

The protein electrostatic potential calculated using the APBS (Adaptive Poisson-Boltzmann Solver) Pymol plugin [114] revealed PrnB protein surface as mainly negative with a positive charge patch underneath the heme cavity, see Fig. 4.9. Furthermore Fig. 4.9 helps to identify the substrate entrance toward the big heme distal cavity site.

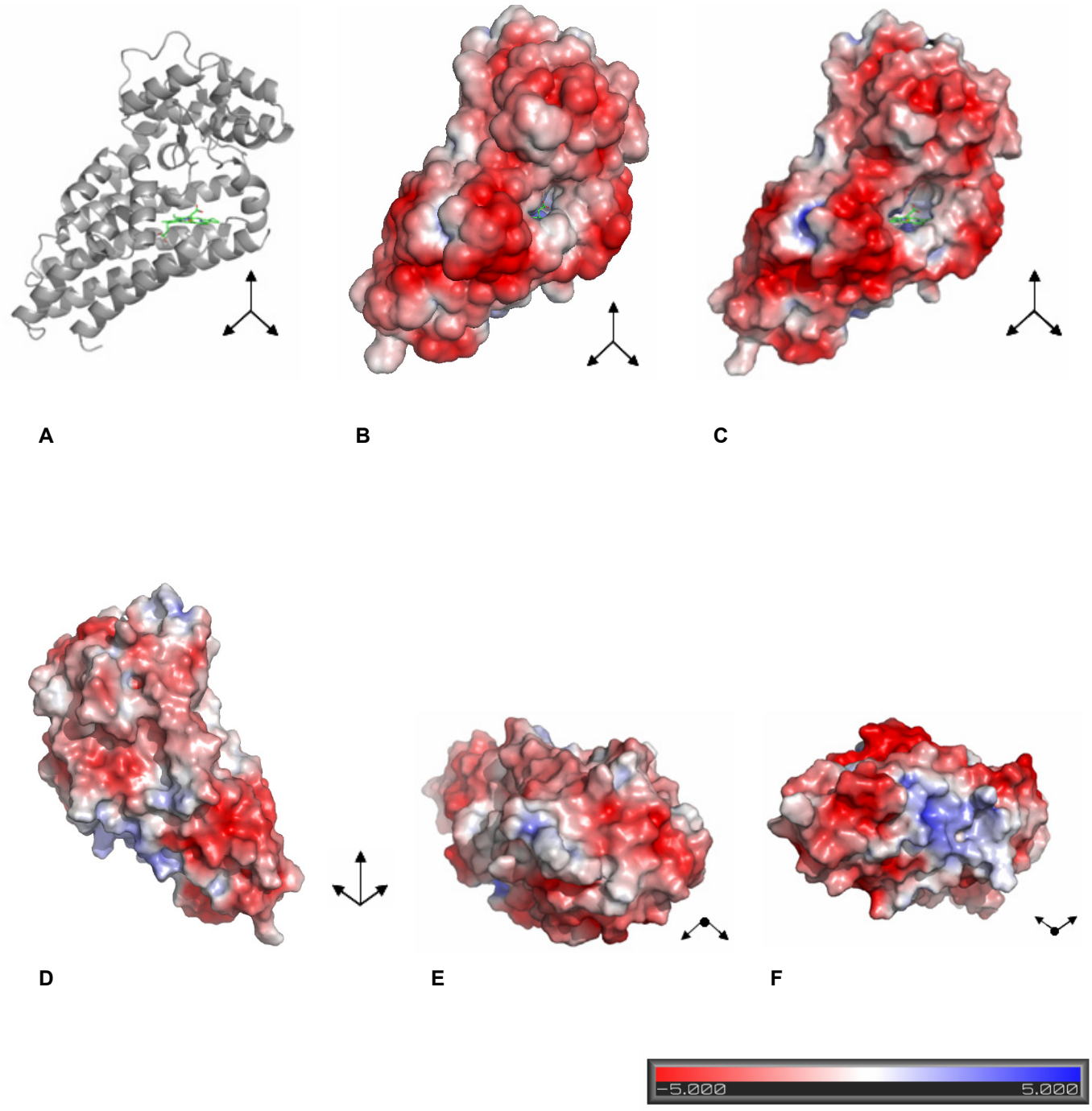



**Figure 4.9**
Electrostatic potential of native PrnB model calculated with the APBS Pymol plugin [114; 115]. A: PrnB cartoon representation, heme in green. B: electrostatic potential on PrnB solvent accessible surface, note the access to the heme cavity, PrnB in the same orientation of A. C: the same as B, but showing the electrostatic potential calculated on the solvent accessible surface projected on PrnB molecular surface. D: as C but showing PrnB backview. E: again same as C, but showing PrnB top view. F: as C but showing PrnB bottom view. Blue and red colors represent positive and negative electrostatic potential, respectively. Hue intensity depicts strength of the potential (in kT/e units) as shown on the scale bar at the bottom of the page.

### 4.4.2 Heme binding

The heme binding pocket is constituted by the α8, σ9, σ15 and σ16 helices bundle together with the perpendicular to the bundle σ10 and σ13 helices and the loop between helices σ10 and σ12. H353 on helix α16 acts as the proximal heme ligand. See Fig. 4.10.

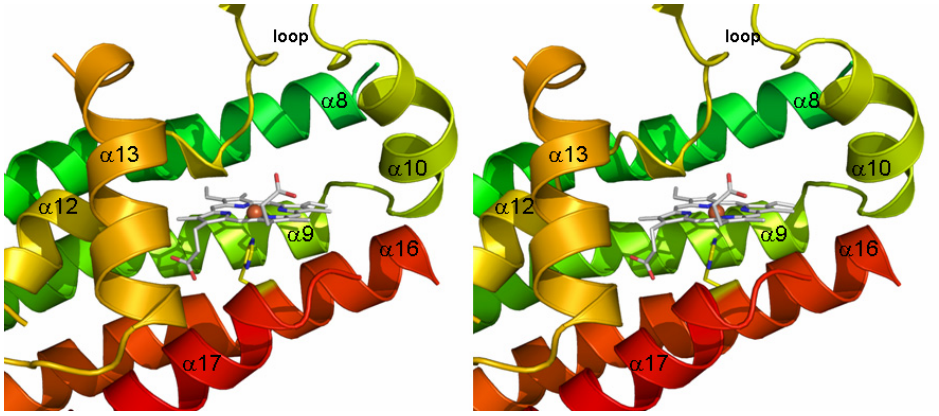



**Figure 4.10**
A close look at the heme binding helices. Histidine 350 the proximal heme ligand is shown as stick in yellow color. Heme in white.

Visual inspection of the model with WinCoot [51] and heme-protein interaction analysis with Ligplot [116] helped to identify residue involved in heme binding. Fig. 4.11 shows the electron density for the heme binding region in the native structure together with residues interacting with it.

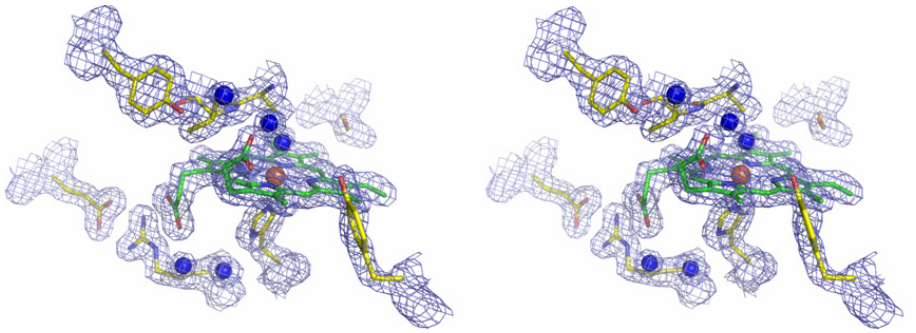


**Figure 4.11**
Electron density and model of the heme region. Density is countered at 1σ. Heme in green, protein residues in yellow, Fe atom in orange waters molecules in blue.

In Fig. 4.12 interaction between the same residues heme prosthetic group and the protein are highlighted by dotted black lines. H353 is the heme fifth ligand.

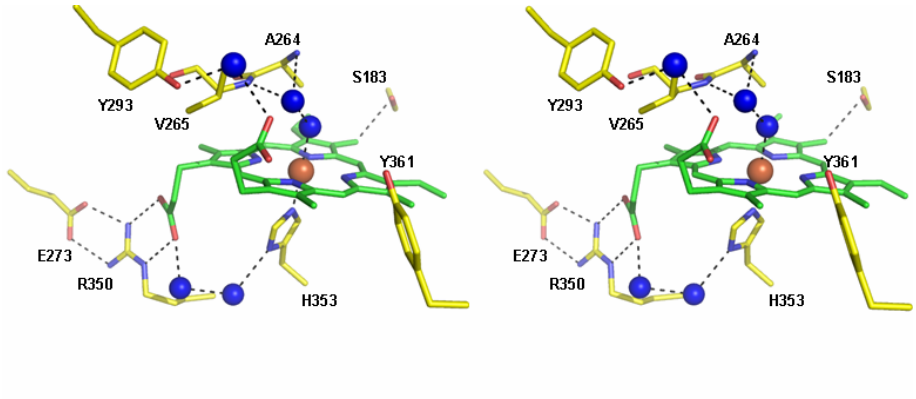
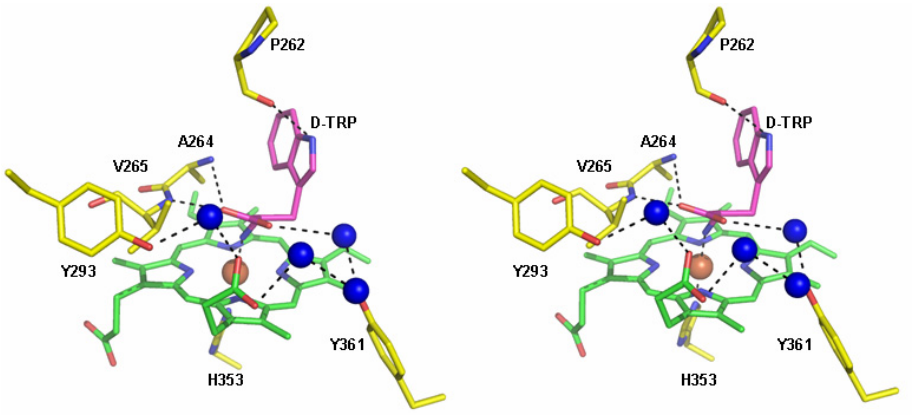


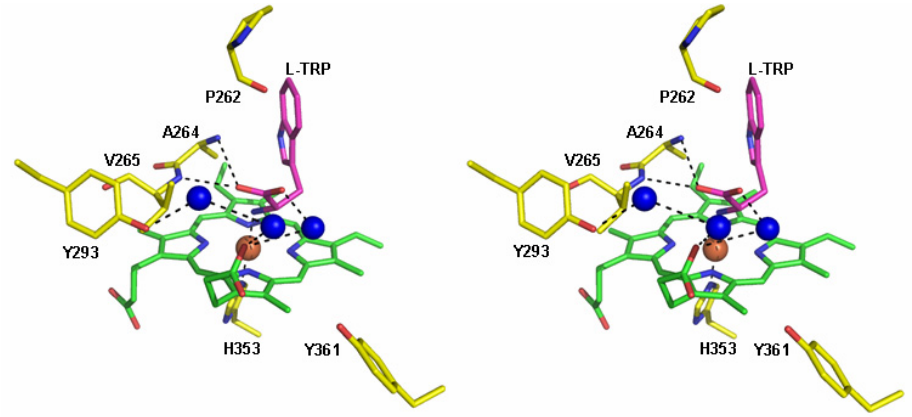

**Figure 4.12**
Stick representation of the heme binding region from the native model. Heme cofactor in green, protein residues in yellow. Water molecules as blue spheres and Iron as a orange sphere. Only for residues V65 and A264 the main protein backbone is shown. In black dotted line interactions between residues waters and the heme.

The sixth position is occupied by a water (or hydroxyl) in the native structure. On the heme proximal side H353 interacts through two waters with the heme 6-propionate. The 6-propionate is hydrogen bounded to R350 ($\alpha$16) that interacts tightly with E273 ($\alpha$12) forming a salt bridge. On the heme proximal side 3 methyl is close S183. The 7-propionate interacts through water with Y293 ($\alpha$13). As already mentioned the iron sixth ligand is a water molecule that through a second water interacts with the peptide bond nitrogens of A264 and V265.

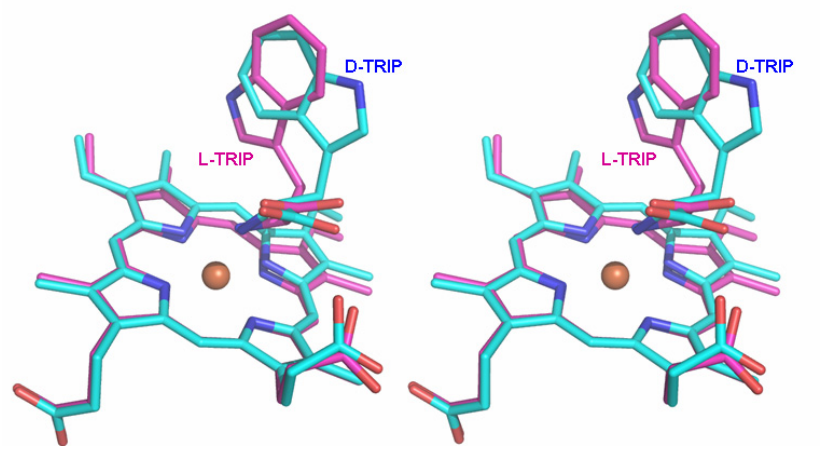
Upon tryptophan binding the sixth proximal water is replaced by the aminoacid amino group, See Fig. 4.13. In the D-tryptophan complex (Fig. 4.13 A) the tryptophan carboxyl group interacts with V265 and A264 amino nitrogens, while the indole nitrogen interacts with the protein backbone carboxyl of P262 part of the loop directly above the heme distal site. Y293 interacts trough one water to the heme 7-propionate, while Y361 is connected by two water molecules to the tryptophan carboxyl group and by two waters to the 7-propionate carboxyl group by one water molecule. Given the different orientation of L and D-tryptophan respect to the heme (see Fig. 4.13 C) in the L-tryptophan complex




A




B




C

**Figure 4.13**

Stick representation of the heme binding region from the D-tryptophan (A) and L-tryptophan (B) bounded models. Heme cofactor in green, protein residues in yellow. Water molecules as blue spheres and Iron as a orange sphere. . Only for residues V65 and A264 the main protein backbone is shown. In black dotted line interactions between residues waters and the heme. Tryptophan ligands are in purple. In C superposition of the heme and tryptophan ligand for the two complexes (L-tryptophan in purple, D-tryptophan in cyan)

(Fig. 4.13 B) the tryptophan carboxyl group is still interacting with V265 an A264, while the indole nitrogen results to be more than 3.5Å from the P262 backbone carboxyl. Y293 is connected to the 7-propionate by two waters. A new interaction arises between the 7-propionate and the tryptophan carboxyl group through one water molecule.

Interestingly L and D-tryptophan phenyl rings overlap each other in the superposed structure (see Fig 4.13C) suggesting the presence of a hydrophobic pocket positioning the aromatic ring above the heme.

### 4.4.3 PrnB the heme distal pocket

A closer look at the residues defining the heme distal pocket, see Fig. 4.14, reveals hydrophobic aminoacids as the major contributors to the heme cavity.

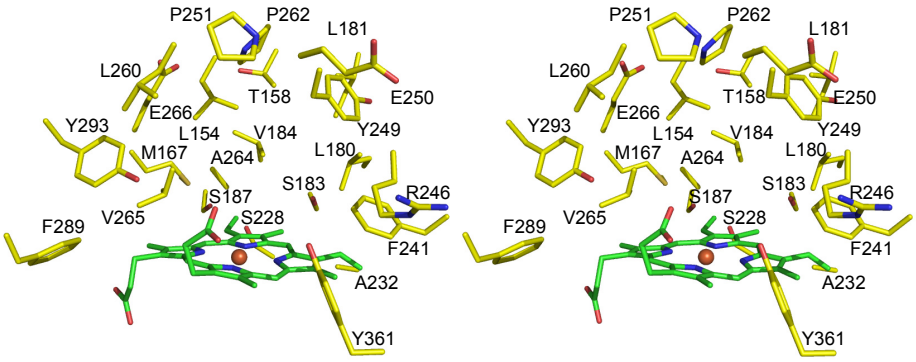
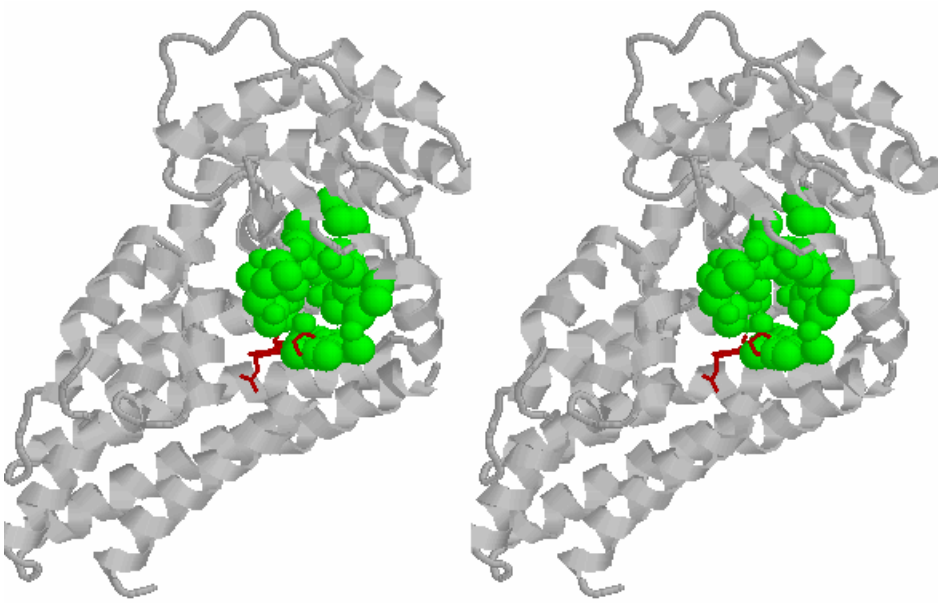



**Figure 4.14**
PrnB heme distal pocket. The side chains of residues defining the heme distal cavity are represented as sticks. Atoms colors: carbon yellow, oxygen red and nitrogen blue. Heme carbons are in green, iron orange.

Apart from the serine residues S183, S187 S228 at the heme plane level and the two tyrosine Y293 and Y361, aromatic residues, that close the substrate access site to the heme distal region most of the remaining side chains have hydrophobic character. R246 beside the substrate entrance is facing away from the heme as well as E250 and E266. M167, is perfectly opposed to heme 7-propionate. Right at the top of the cavity a further tyrosine, Y249, which could participate in hydrogen bonding to the substrate is facing away from the heme. The visual inspection results are confirmed by using the CASTp web based

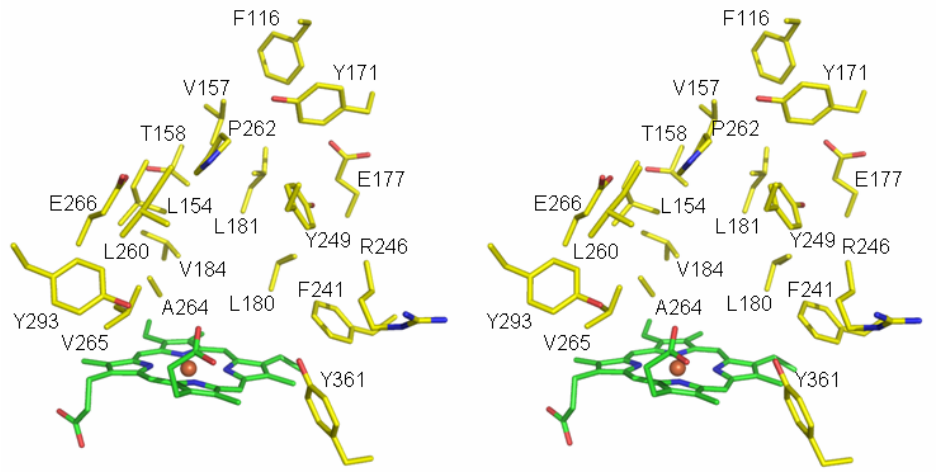
analysis software (http://sts.bioengr.uic.edu/castp/index.php) [117; 118] to calculate the cavity dimensions. The program identified the 21 reidues that together with the heme cofactor that define a cavity with a 420Å$^2$ surface and a volume of 520Å$^3$, see Fig. 4.15.



**A**

```
                .         .         .         .         .
45   LDRVGVFAATHAAVAASDPLQARALVLQLPGLNRNKDVPGIVGLLREFLP   94

                .         .         .         .         .
95   VRGLPSGWGFVEAAAAMRDIGFFLGSLKRHGHEPAEVVPGLEPVLLDLAR  144

                .         .         .         .         .
145  ATNLPPRETLLHVTVWNPTAADAQRSYTGLPDEAHLLESVRISMAALEAA  194

                .         .         .         .         .
197  IALTVELFDVSLRSPEFAQRSDELEAYLQKMVESIVYAYRFISPQVFYDE  244

                .         .         .         .         .
245  LRPFYEPIRVGGQSYLGPGAVEMPLFVLEHVLWGSQSDDQTYREFKETYL  294

                .         .         .         .         .
295  PYVLPAYRAVYARFSGEPALIDRALDEARAVGTRDEHVRAGLTALERVFK  344

                .         .         .         .
345  VLLRFRAPHLKLAERAYEVG------PSMLGELLTLTYAARSRVRAALDE  400
```

**B**




**C**

**Figure 4.15**
**A:** CASTp (http://sts.bioengr.uic.edu/castp/index.php) [117; 118] output generated using a 1,4Å probe. PrnB as cartoon in gray; heme in red stick and as green spheres the cavity identified by the program. **B:** Residues defining the cavity highlighted in green in PrnB structure sequence. **C**: Stick representation of PrnB residues defining the CASTp calculated heme cavity, heme in green, residues side chains in yellow.

In Fig. 4.16 the sphere representation of the amino acids described in Fig. 4.15 helps to identify the hydrophobic pocket that hold in place the tryptophan aromatic ring.

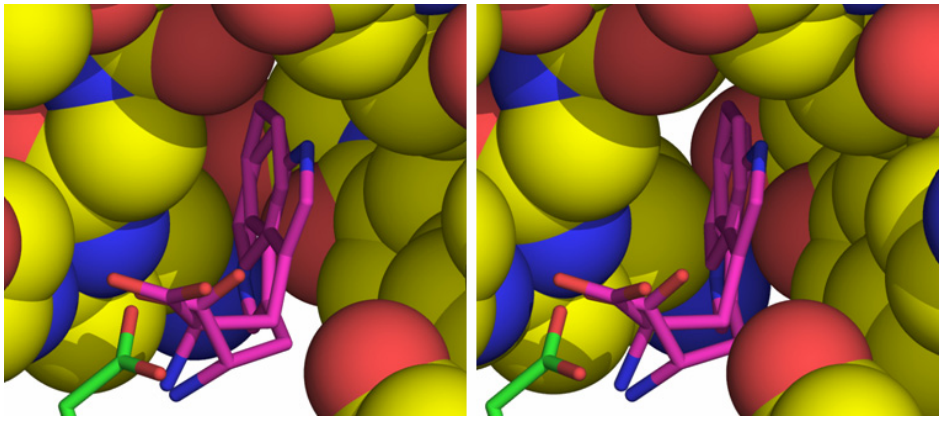


**Figure 4.16**
Sphere representation of the PrnB hydrophobic pocket (carbon in yellow, oxygen in red nitrogen in blue) holding in place L and D-tryptophan ligands (in magenta), the heme 7-propionate is visible (in green).

### 4.4.4 PrnB homologues

Fig. 4.17 shows the sequence alignment of the initial four homologues identified by Hammer *et al*. [71] and four homologues from the DOE Joint Genome Institute Integrated Microbial Genomes webpages together with PrnB main secondary structure elements. Because of the very high sequence homology we can assume that the structure of PrnB would be conserved for all the homologues. *Myxococcus fulvus* sequence with a sequence identity of 59% and sequence similarity of 69% respect to *P. fluorescens* BL915 PrnB sequence is the least conserved of the eight homologues. The higher variation is observed at the protein N and C terminii. All the major residues involved in the heme cofactor binding are conserved as well as the one defining the heme distal cavity. Interestingly the loop, between helix 10 and helix 11, that covers the heme distal site is highly conserved too.

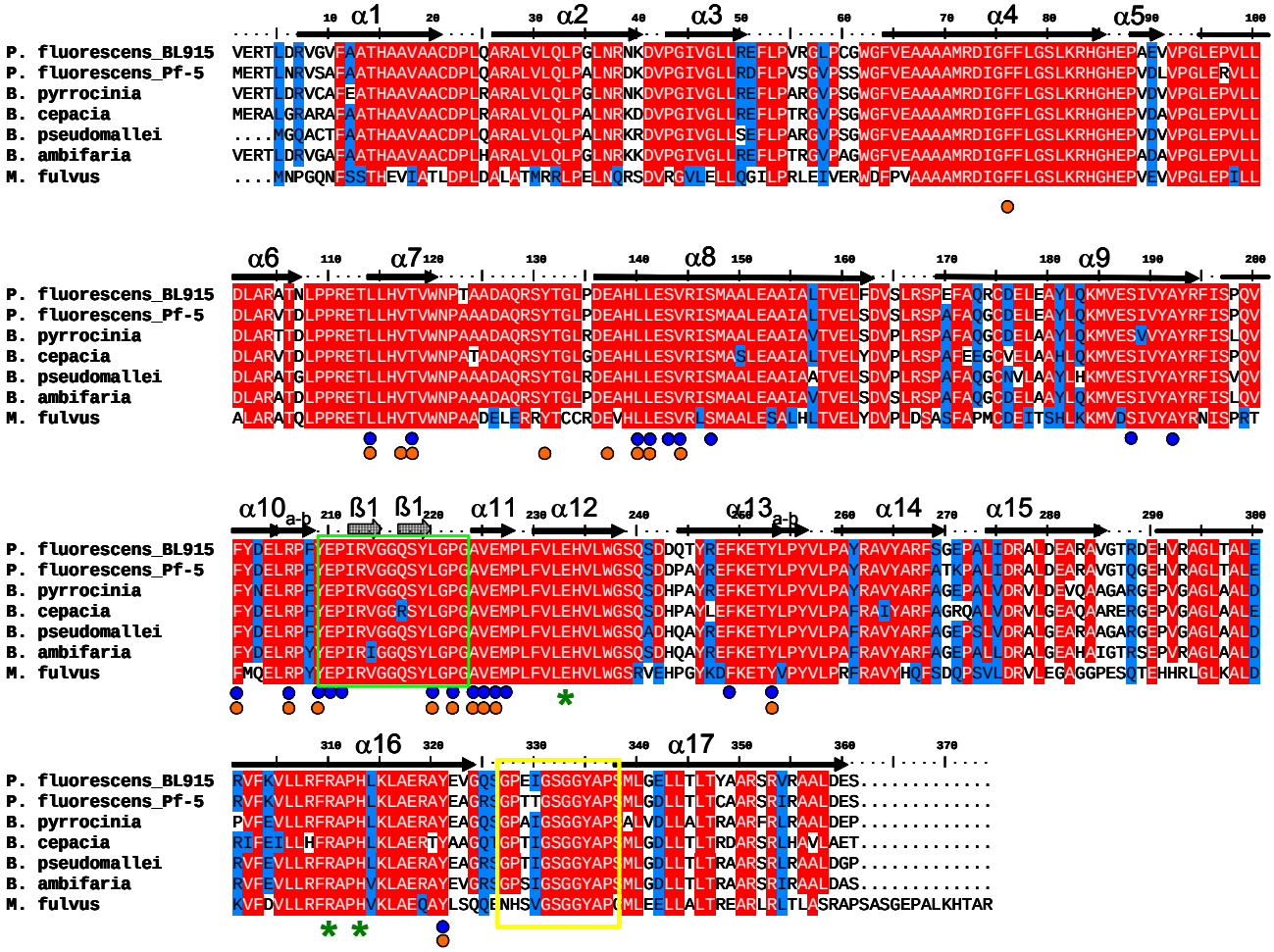


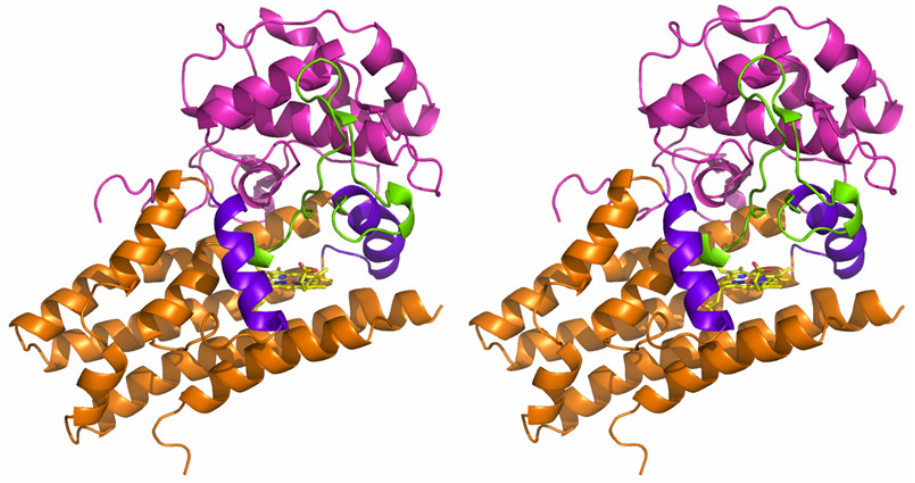

**Figure 4.17**

PrnB homologues multiple alignment (Clustal W). First row *P.fluorescens* BL915, second row *P.fluorescens* Pf-5, third row *Burkholdeia pyrrocinia*, fourth row *Burkholderia cepacia*, fifth row *Burkholderia cepacia* 383, sixth row *Burkholderia pseudomallei, seventh* row *Burkholderia ambifaria, eight row Myxococcus fulvus*. The yellow box highlights the gap present in PrnB structure. The green box highloghts the residues belonging to the loop above the heme cavity. Green star highlights main residues involved in heme cofactor binding. Black stars highlights residues defining the heme distal pocket shown in Fig. 4.14, while orange ones the one identified by using CASTp Fig. 4.15. Shaded in red identical residues in blue similar ones (threshold 85%). PrnB secondary structures elements in black.

### 5.4.5 PrnB and IDO comparison

Despite a protein sequence identity lower than 10% PrnB structure shows striking similarity with the recently solved indoleamine 2,3-dioxygenase (IDO) structure (PDB accession code 2D0T) [98]. Interestingly the IDO structure was solved from $P2_12_12_1$ crystals containing two monomers into the asymmetric unit. The two monomers are bonded together through a disulfide bond involving cysteine 308. Native human IDO has been reported to be a monomeric protein [119], but oligomerization studies have not been reported for the recombinant product. The presence of the disulfide linkage is not discussed in IDO structure paper [98]. By analogy with the Indolemine-2,3-dioxygenase structure, the PrnB model can be divided into two major domains, see Fig. 4.18. The large domain, in green is formed by the heme binding helices. It is separated by the small domain, in blue by the two helices perpendicular to the bundle, in cyan and the loop closing the heme proximal site, in red. From this graphical representation the two proteins structural homology and overall conserved topology appears obvious. PrnB consists of 12 major $\alpha$-helices one more than IDO. The extra PrnB $\alpha$-helix is at the protein N-terminal. The presence of this ordered $\alpha$1 helix could be explained by the presence in PrnB crystal of the 41aa long pCIB-HIS-TEV tag that could help stabilizing the first secondary structure. In IDO crystal there are only 3 extra aa before the first methionine (during the purification the N-terminal hexahistidine tag is cleaved by thrombin) and the first 21 aa in the structure form a flexible loop.

A



B

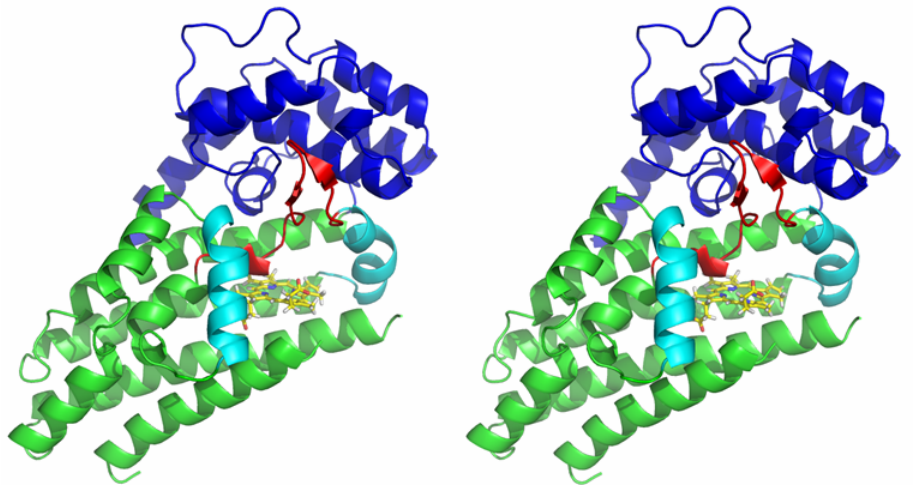


**Figure 4.18**
Cartoon representation of (A) Indoleamine 2,3-dioxygenase (IDO, PDB accession code 2D0T) and (B) PrnB. The proteins are composed of two major domains in accordance with IDO description by Sugimoto *et al*. [98]. The small domain is colored in magenta for IDO and in blue for PrnB, the large domain in orange for IDO and in green for PrnB. In violet for IDO and in cyan for PrnB are the connection helices between the two domain and in yellow for IDO and in red for PrnB the loop that mask the heme distal site.

Structural superposition of PrnB on IDO model using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50] resulted in a Z-score of 5.2 with a RMSD of 2.4Å for 273 aligned residues. In Fig. 4.19 is depicted the ribbon model of the two superimposed structure. Despite small differences in the position and length of the two protein helices the heme cofactor results to be perfectly superimposed in the two models. In Fig. 4.20 is shown the structure based alignment between PrnB and IDO.

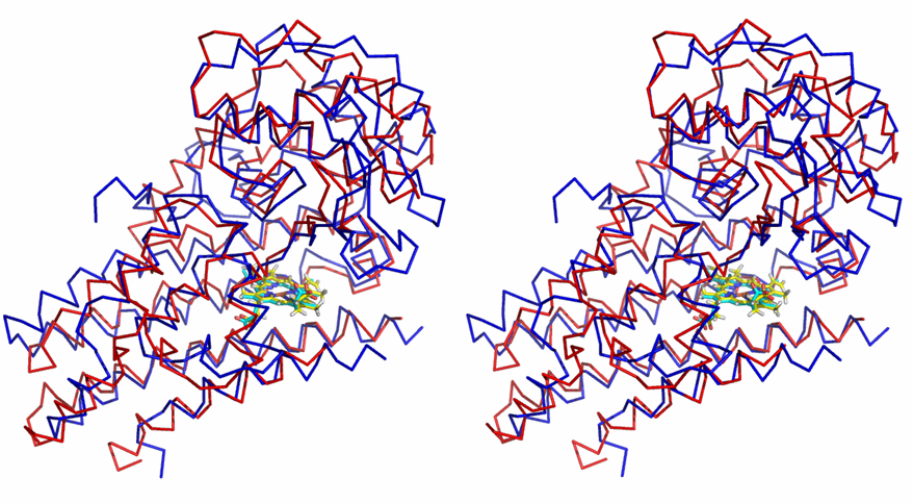


**Figure 4.19**
Ribbon representation of PrnB (red , yellow heme) superposition to IDO (blue, cyan heme), using the SSM algorithm of WinCOOT [51]. Despite differences in helices length and numbers, Fig. 4.20, the overall protein topology is clearly conserved. Note the conserved position of the heme prosthetic group position inside the main four helix bundle.

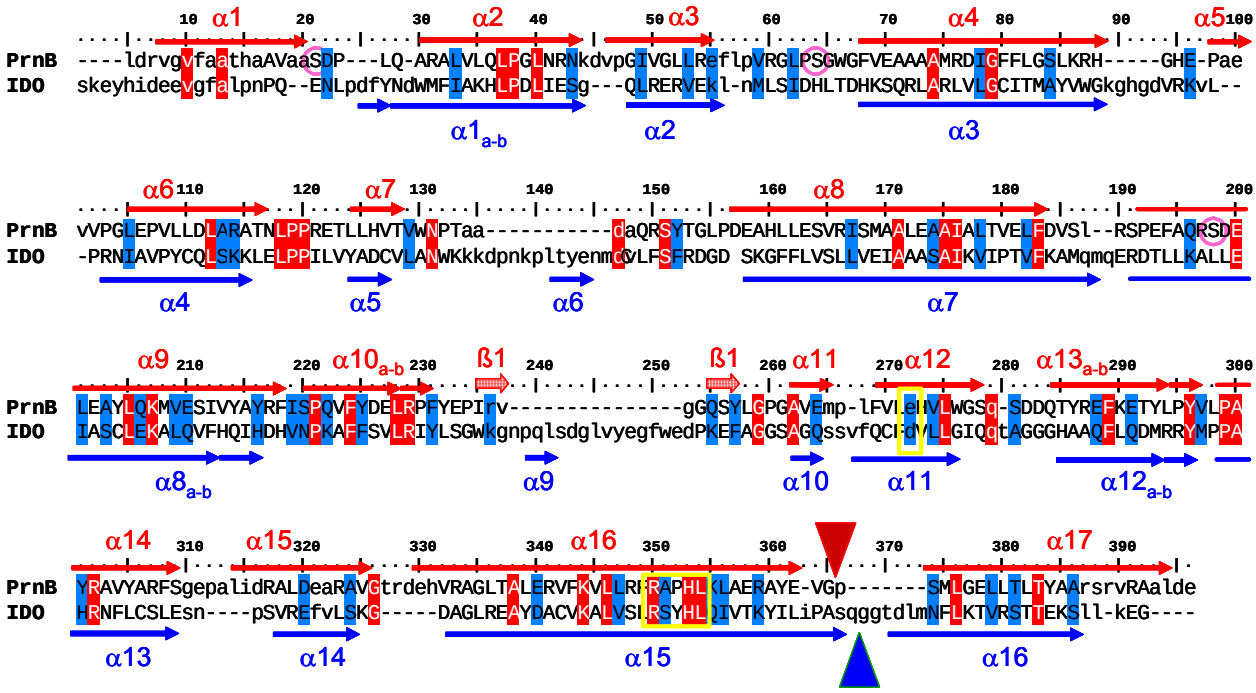



**Figure 4.20**
Sequence alignment based on structural superposition of PrnB on IDO using the *Protein structure comparison service SSM at European Bioinformatics Institute* (http://www.ebi.ac.uk/msd-srv/ssm) [50]. In capital superposed residues; shaded in red identical residues, in blue similar ones. The yellow boxes indicate the RxxH motif involved in heme proximal binding toghether with the arginine interacting E273/D274. Secondary structure elements, calculated with STRIDE [96] (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py) and visually inspected, are numbered from the N-terminal onward (α = helix, β = sheet). The two triangles indicated the position of the unique gap (disordered loop) present in both the protein models. Circled in purple the three Cysteine residues mutated to Serine for crystallization purpose.

A close examination of IDO heme binding site compared with PrnB shows other conserved features between the two proteins, see Fig. 4.21. The E273, R350, H353 triad in the PrnB heme distal site is equivalent to the D274, R343, H346 triad of IDO. D274 has been reported to be essential for proper heme loading on recombinant human IDO probably for its ability to stabilize the proximal histidine [120]. In PrnB S183 is close to the heme 3 methyl group suggesting a heme stabilizing interaction between the two groups. S183 is conserved in IDO where we found S167 in the same position. Y293 and Y361 which interacts through waters with the heme 7-propionate in the D-tryptophan bonded PrnB model are replaced by M295 (Y293) and I354 (Y361) in IDO structure. While PrnB amino group of V265 and A264 interacts with D and L tryptophan, S293 in IDO structure is directly involved in hydrogen bonding the heme 7-propionate.

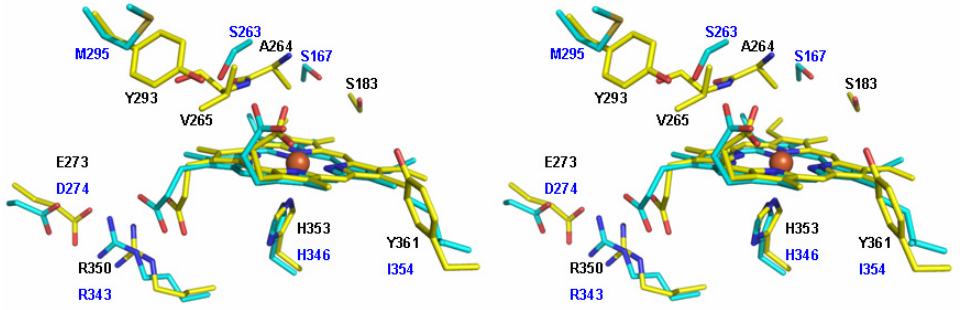



**Figure 4.21**
Comparison of IDO (in cyan) and PrnB (in yellow) heme binding site. The E273, R350, H353 triad on the PrnB heme distal site results to be equivalent to the D274, R343, H346 triad of IDO. S183 is replaced by S167 in IDO. While the 7-propionate interacts in with Y293 in PrnB through a water molecule, in IDO structure it interacts directly with S293. The S293 analog V265 in PrnB structure interacts with the tryptophan ligand carboxyl group or in the native structure with the Fe atom through two water molecules.

Other major differences between IDO and PrnB structure arise from differences in the length of various helix connecting loops. The more evident are from the structure sequence based alignment in Fig. 4.20 are highlighted in Fig. 4.22. The first one connects $\alpha$5 to $\alpha$7 and contains the three residues long $\alpha$6. The second one is the loop involved in covering the heme distal side, it extends from $\alpha$8 to $\alpha$10 and contains the three residue long $\alpha$9. In

PrnB this second loop connects α10 to α11 and contains the small antiparallel beta-sheet β1.

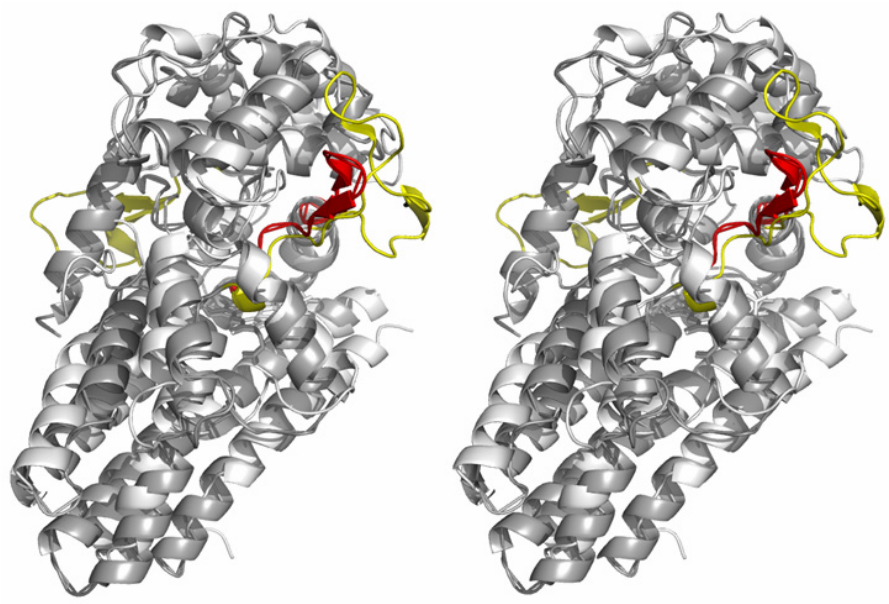


**Figure 4.22**
Major loops differences between IDO (light gray, yellow loops) and PrnB (dark gray, red loop). IDO: the loop connecting α5 to α7 on the back of the protein heme binding region and the one connecting α8 to α10 on top of the heme cofactor. In PrnB (B) this second loop result 15 aminoacids shorter.

The bigger extension of this second loop in IDO produces a larger heme distal cavity. CASTp calculation on PrnB (See 5.4.3) gives a 520Å$^3$ pocket dimension compared to IDO 1831Å$^3$. Fig. 4.23 shows a superposition of the the cavity forming residues in the two proteins.

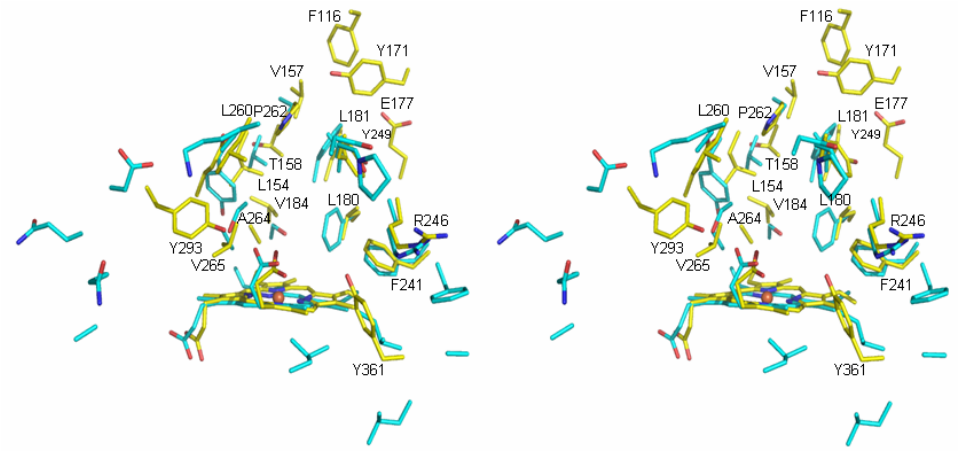



**Figure 4.23**
Superposition of IDO (in cyan) and PrnB (in yellow) heme distal cavity residues as calculated by CASTp. For clarity only PrnB residues are labeled.

### 4.4.6 PrnB Structure and Enzymatic Mechanism

PrnB structure elucidation confirmed heme b as the protein cofactor and resulted to display the same overall fold of Indoleamine 2,3-dioxygenase. The enzyme active site has been located on the distal side of the heme cofactor from co-crystallization experiment with D and L tryptophan reported to be enzyme substrates. In our structures tryptophan is bonded to the ferric iron through its $\alpha$ nitrogen. However PrnB needs probably to be reduced to its ferrous form to catalyze the tryptophan $\rightarrow$ aminophenylpyrrole reaction. So we expect tryptophan or 7-Cl-tryptophan to sit on the cavity above the reduced heme cofactor.

Solution of IDO structure and comparison of wild type and mutants activity lead Sugimoto *et al*. [98] to propose a mechanism in which the tryptophan indole decyclization is driven by iron-bound dioxygen. See Fig. 4.24.

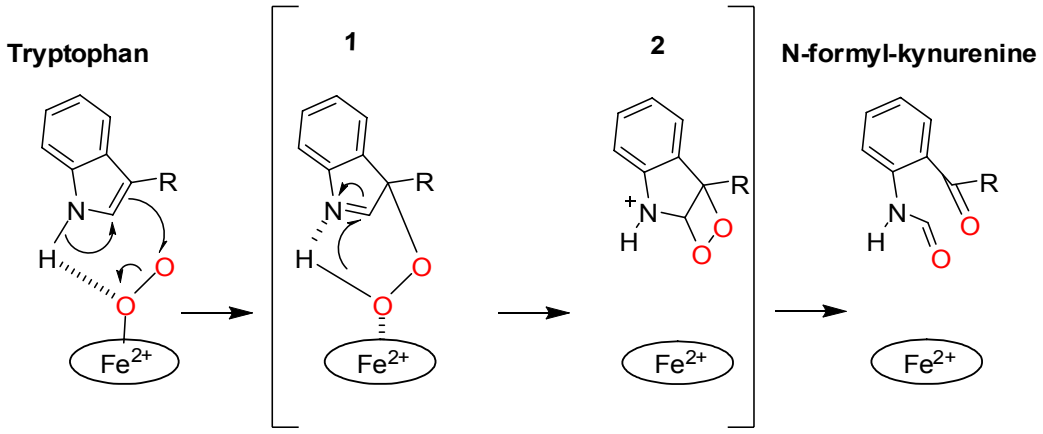



**Figure 4.24**
Indoleamine 2,3-dioxygenase proposed mechanism [98]. IDO catalyzes the cleavage of the bond between carbon 2 and carbon 3 of the tryptophan indole ring. The trigger of the reaction is abstraction of a proton from nitrogen 1. Binding of oxygen and tryptophan enables the interaction between the heme bonded dioxygen and the NH group of indole. Dioxygen then abstracts the proton. Subsequent rearrangement of the indole ring electronic structure leads to formation of a bond between the proximal dioxygen oxygen and indole carbon 3. Cleavage of the Fe-O bond results in the formation of intermediate 1 (3-hydroperoxyindolenine). Intermediate 1 rearranges into 2 (dioxyetane) that decomposes into N-formylkynuirenine.

This mechanism has been proposed due to the absence of relevant polar/charged residues protein side chain that could act as catalytic bases in IDO active pocket.

Computer assisted modeling of PrnB substrate in the enzyme active site , using ArgusLab 4.0 software [97] (http://www.planaria-software.com), predicted 7-Cl-L-tryptophan to bind with its 7 position facing the heme cavity ceiling. Thus presenting its carbon 2-3 double bond to the heme iron in the opposite way respect to IDO, see Fig. 4. 25.

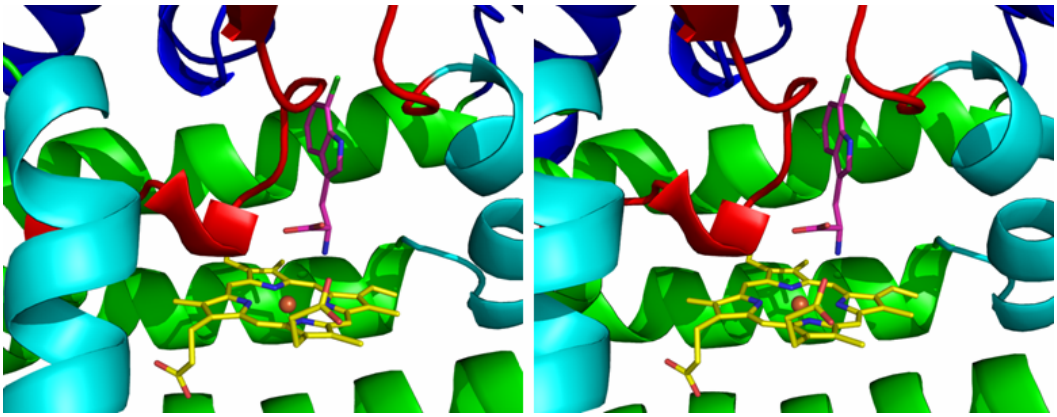


**Figure 4.24**
One representative result obtained modeling 7-Cl-L-tryptophan into PrnB heme cavity using the ArgusLab software [97]. 7-Cl-tryptophan in purple, chlorine atom green. Heme in yellow. PrnB cartoon color as in Fig. 4.17. Green large domain, blue small domain, cyan interconnecting helices and in red the above heme cavity loop. All the simulation resulted in the chlorine pointing at the cavity ceiling. The carbon 2-3 double bond is facing the heme in the opposite way compared to proposed tryptophan binding to IDO.

By analogy with the IDO mechanism and taking into account the modelling results we can propose more than one way in which PrnB catalyzes a reaction using heme bound dioxygen as the protein catalysts, see examples in Fig. 4.26. These are of course only speculative models and rely on the different way the substrate should bind into the enzyme active pocket. Central to the understanding of the PrnB reaction mechanism is the solution of the crystal structure of the reduced protein together with is physiological substrate 7-Cl-tryptophan. A difficult technical task made more difficult by the difficulty to obtain enantiomeric pure D and L chloro-tryptophan.

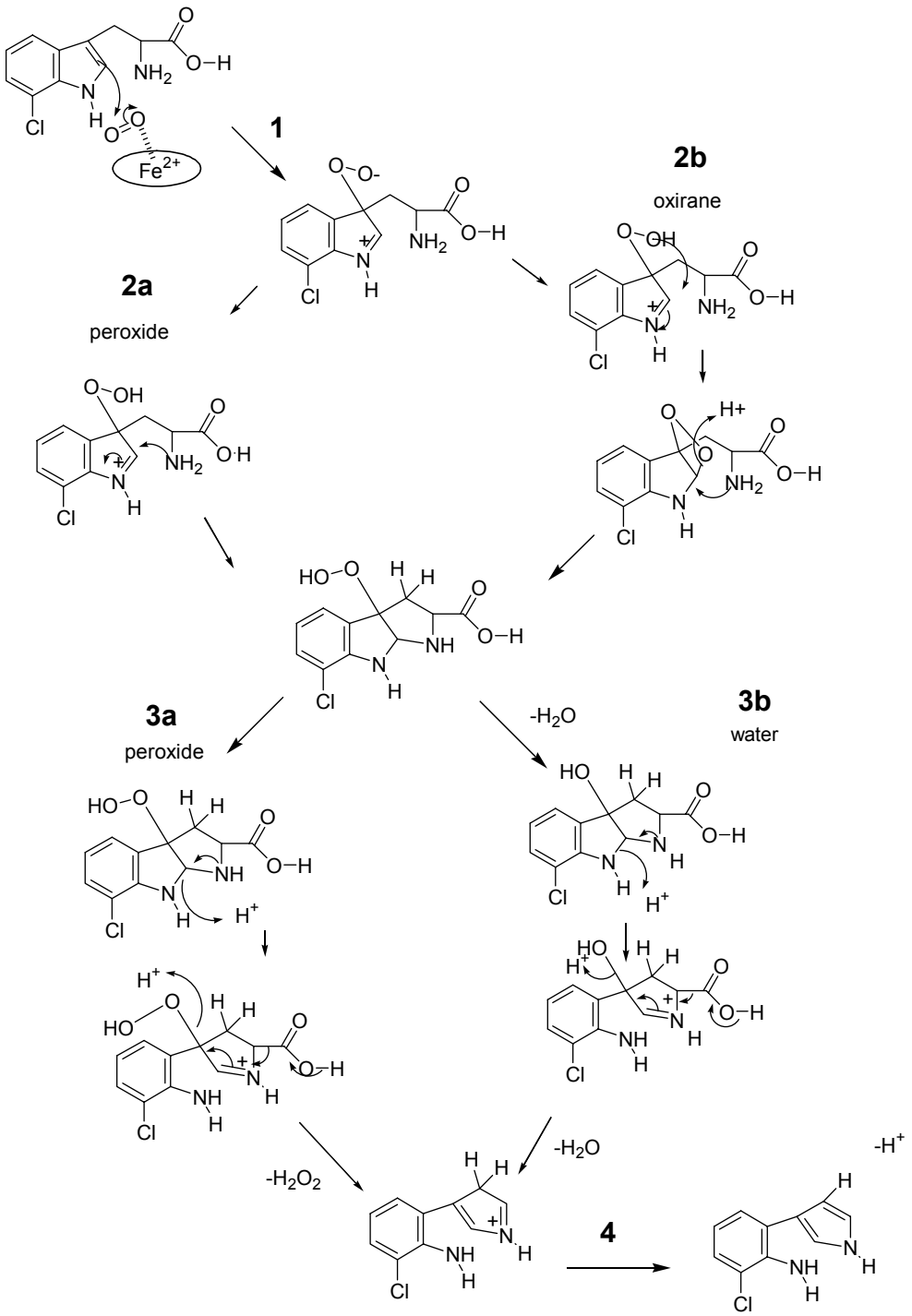



**Figure 4.26**
Proposed reaction mechanism for PrnB. Oxidation of the 2-3 carbon double bond, by iron bonded dioxygen, activates the indole ring for nucleophilic attack. The reaction could then proceed either via peroxide (2a) mediated ring rearrangement or through an oxyrane intermediate (2b) and ring rearrangement. The ring rearrangement and the formation of a three ring intermediated is mediated by the $\alpha$ nitrogen nucleophilic attack on carbon 2. The three ring compound could then degrade in two ways by peroxide elimination (3a) or via two water elimination (3b) driven in both cases by the decarboxylation occurring at the carbon $\alpha$.

A second possibility is that 7-Cl-tryptophan will bind to the enzyme in the same way as D- and L-tryptophan and that the true active form will be the PrnB oxidized one. Given these assumptions we could expect a different enzymatic mechanism which would work in the same way with both L-tryptophan and 7-Cl-tryptophan. In this case too a speculative reaction mechanism can be depicted, see Fig. 4.27.

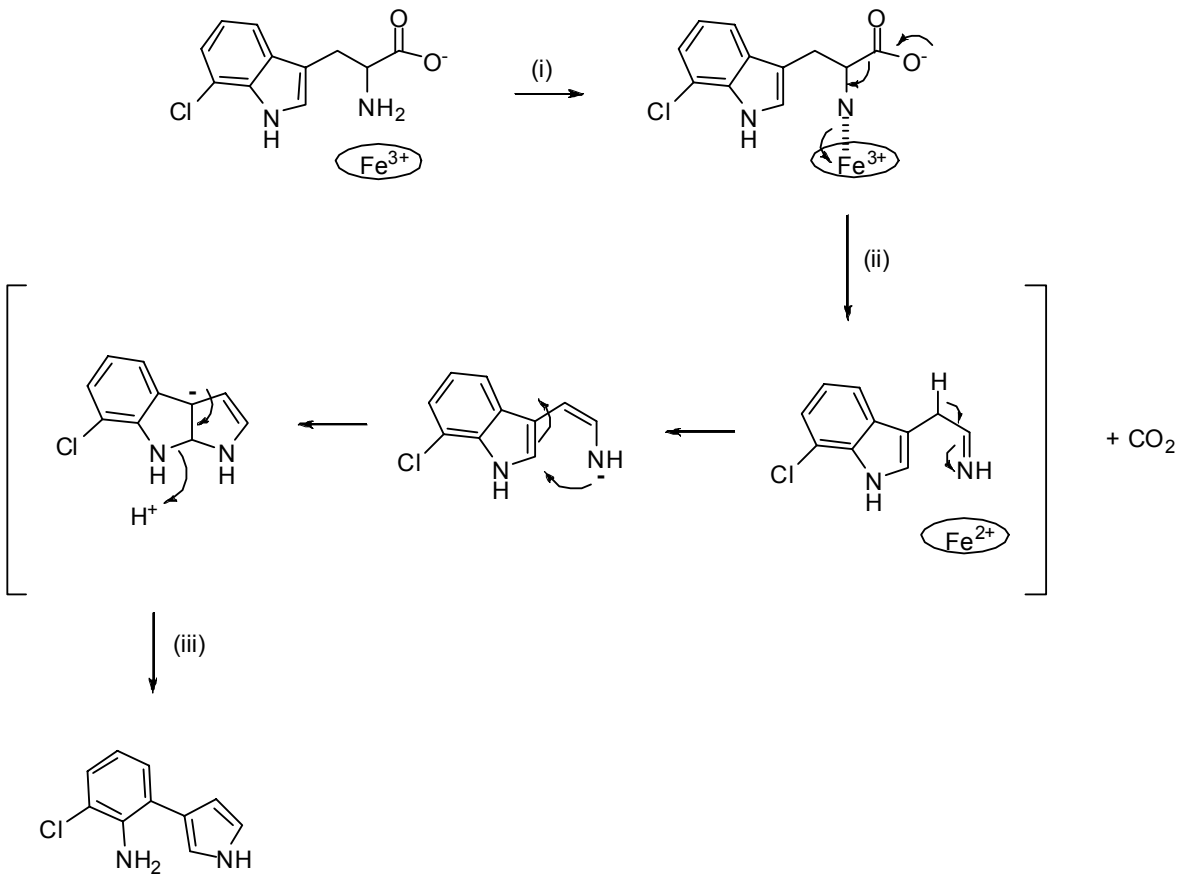



**Figure 4.27**
Proposed reaction mechanism for PrnB. (i): from the crystallographyc data of the D- and L-tryptophan bond PrnB the ferric iron may oxidize the α-amino nitrogen providing the driving force for the decarboxylation of the carboxylate group and subsequent formation of an imine. Tautomerization of the imine to an electrophilic amine group, in bracket, will provide the driving force for the indole ring rearrangment and formation of the phenylpyrrole ring of aminopyrrolnitrin (iii).

## 4.5 Future work

Apart from trying to obtain a crystal structure of reduced PrnB in the presence of one of its substrates (both tryptophan and chloro tryptophan), we have now a working hypothesis regarding the enzyme reaction mechanism. A priority for a better understanding of this unique enzyme is to eastablish a reproducible robust enzymatic assay. A starting point could be using IDO assay conditions. Moreover availability of a robust PrnB expression system with protein yields of about 10mg/liter represent an ideal system to generate site directed mutants to further investigate the function of conserved residues among PrnB homologues or residues that will be identified next to the tryptophan in the reduced form protein structure.

# REFERENCES


1. Murphy, C. D. (2003). New frontiers in biological halogenation. *J Appl Microbiol* 94, 539-48.

2. Gribble, G. W. (2004). Amazing Organohalogens. *American Scientist* 92, 342-349.

3. Gribble, G. W. (2003). The diversity of naturally produced organohalogens. *Chemosphere* 52, 289-297.

4. van Peé, K. H. & Patallo, E. P. (2006). Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl Microbiol Biotechnol* 70, 631-641.

5. Harris, C. M., Kannan, R., Kopecka, H. & Harris, T. M. (1985). The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J. Am. Chem. Soc.* 107, 6652-6658.

6. Gerhard, U., Mackay, J. P., Maplestone, R. A. & Williams, D. H. (1993). The role of the sugar and chlorine substituents in the dimerization of vancomycin antibiotics. *J. Am. Chem. Soc.* 115, 245-252.

7. Piraee, M., White, R. L. & Vining, L. C. (2004). Biosynthesis of the dichloroacetyl component of chloramphenicol in *Streptomyces venezuelae* ISP5230: genes required for halogenation. *Microbiology* 150, 85-94.

8. Pereira, E. R., Belin, L., Sancelme, M., Prudhomme, M., Ollier, M., Rapp, M., Severe, D., Riou, J. F., Fabbro, D. & Meyer, T. (1996). Structure-activity relationships in a series of substituted indolocarbazoles: topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J Med Chem* 39, 4471-7.

9. Kerksiek, K., Mejillano, M. R., Schwartz, R. E., Georg, G. I. & Himes, R. H. (1995). Interaction of cryptophycin 1 with tubulin and microtubules. *FEBS Letters* 377, 59-61.

10. Golakoti, T., Ogino, J., Heltzel, C. E., Husebo, T. L., Jensen, C. M., Larsen, L. K., Patterson, G. M. L., Moore, R. E., Mooberry, S. L. & al., a. e. (1995). Structure determination, conformational analysis, chemical stability studies, and antitumor evaluation of the cryptophycins. Isolation of 18 new analogs from *Nostoc* sp. strain GSV 224. *J. Am. Chem. Soc.* 117, 12030-12049.

11. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. (2004). Halogen bonds in biological molecules. *Proc Natl Acad Sci U S A* 101, 16789-94.

12. Grégori Gerebtzoff, X. L.-B., Holger Fischer, Adrian Frentzel, Anna Seelig,. (2004). Halogenation of Drugs Enhances Membrane Binding and Permeation. *ChemBioChem* 5, 676-684.

13. Paterson, I. & Anderson, E. A. (2005). Chemistry. The renaissance of natural products as drug candidates. *Science* 310, 451-3.

14. Floss, H. G. (2006). Combinatorial biosynthesis--potential and problems. *J Biotechnol* 124, 242-57.

15. Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Peé, K. H. & Naismith, J. H. (2005). Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* 309, 2216-9.

16. Shaw, P. D. & Hager, L. P. (1959). Biological Chlorination. IV . Peroxidative nature of enzymatic chlorination. *J. Am. Chem. Soc.* 81, 6527-6528.

17. Hager, L. P., Morris, D. R., Brown, F. S. & Eberwein, H. (1966). Chloroperoxidase. II. Utilization of halogen anions. *The Journal Of Biological Chemistry* 241, 1769-1777.

18. Murphy, C. D. (2006). Recent developments in enzymatic chlorination. *Nat Prod Rep* 23, 147-152.

19. Dairi, T., Nakano, T., Aisaka, K., Katsumata, R. & Hasegawa, M. (1995). Cloning and nucleotide sequence of the gene responsible for chlorination of tetracycline. *Biosci Biotechnol Biochem* 59, 1099-106.

20. Hammer, P. E., Hill, D. S., Lam, S. T., van Peé, K. H. & Ligon, J. M. (1997). Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl Environ Microbiol* 63, 2147-54.

21. Kirner, S., Hammer, P. E., Hill, D. S., Altmann, A., Fischer, I., Weislo, L. J., Lanahan, M., van Peé, K. H. & Ligon, J. M. (1998). Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J Bacteriol* 180, 1939-43.

22. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. (2006). Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 440, 368-71.

23. Morris, D. R. & Hager, L. P. (1966). Chloroperoxidase. I. Isolation and properties of the crystalline glycoprotein. *J Biol Chem* 241, 1763-8.

24. Hofrichter, M. & Ullrich, R. (2006). Heme-thiolate haloperoxidases: versatile biocatalysts with biotechnological and environmental significance. *Appl Microbiol Biotechnol* 71, 276-88.

25. van Peé, K.-H. & Unversucht, S. (2003). Biological dehalogenation and halogenation reactions. *Chemosphere* 52, 299-312.

26. Conesa, A., van De Velde, F., van Rantwijk, F., Sheldon, R. A., van Den Hondel, C. A. & Punt, P. J. (2001). Expression of the *Caldariomyces fumago* chloroperoxidase in Aspergillus niger and characterization of the recombinant enzyme. *J Biol Chem* 276, 17635-40.

27. Sundaramoorthy, M., Terner, J. & Poulos, T. L. (1998). Stereochemistry of the chloroperoxidase active site: crystallographic and molecular-modeling studies. *Chem Biol* 5, 461-73.

28. Wagenknecht, H.-A. & Woggon, W.-D. (1997). Identification of intermediates in the catalytic cycle of chloroperoxidase. *Chemistry & Biology* 4, 367-372.

29. Almeida, M., Filipe, S., Humanes, M., Maia, M. F., Melo, R., Severino, N., da Silva, J. A. L., Frausto da Silva, J. J. R. & Wever, R. (2001). Vanadium haloperoxidases from brown algae of the Laminariaceae family. *Phytochemistry* 57, 633-642.

30. Messerschmidt, A., Prade, L. & Wever, R. (1997). Implications for the catalytic mechanism of the vanadium-containing enzyme chloroperoxidase from the fungus *Curvularia inaequalis* by X-ray structures of the native and peroxide form. *Biol Chem* 378, 309-15.

31. Hasan, Z., Renirie, R., Kerkman, R., Ruijssenaars, H. J., Hartog, A. F. & Wever, R. (2006). Laboratory-evolved vanadium chloroperoxidase exhibits 100-fold higher halogenating activity at alkaline pH: catalytic effects from first and second coordination sphere mutations. *J Biol Chem* 281, 9738-44.

32. Littlechild, J. (1999). Haloperoxidases and their role in biotransformation reactions. *Curr Opin Chem Biol* 3, 28-34.

33. Renirie, R., Hemrika, W. & Wever, R. (2000). Peroxidase and phosphatase activity of active-site mutants of vanadium chloroperoxidase from the fungus *Curvularia inaequalis*. Implications for the catalytic mechanisms. *J Biol Chem* 275, 11650-7.

34. Hecht, H. J., Sobek, H., Haag, T., Pfeifer, O. & van Peé, K. H. (1994). The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an alpha/beta hydrolase fold. *Nat Struct Biol* 1, 532-7.

35. Hofmann, B., Tolzer, S., Pelletier, I., Altenbuchner, J., van Peé, K. H. & Hecht, H. J. (1998). Structural investigation of the cofactor-free chloroperoxidases. *Journal of Molecular Biology* 279, 889-900.

36. Honda, K., Kataoka, M., Sakuradani, E. & Shimizu, S. (2003). Role of *Acinetobacter calcoaceticus* 3,4-dihydrocoumarin hydrolase in oxidative stress defence against peroxoacids. *Eur J Biochem* 270, 486-94.

37. Kirner, S., Krauss, S., Sury, G., Lam, S. T., Ligon, J. M. & van Peé, K. H. (1996). The non-haem chloroperoxidase from *Pseudomonas fluorescens* and its relationship to pyrrolnitrin biosynthesis. *Microbiology* 142 ( Pt 8), 2129-35.

38. Hawkins, C. L., Pattison, D. I. & Davies, M. J. (2003). Hypochlorite-induced oxidation of amino acids, peptides and proteins. *Amino Acids* 25, 259-274.

39. van Peé, K. H. (2001). Microbial biosynthesis of halometabolites. *Arch Microbiol* 175, 250-8.

40. Hohaus, K. A.; Burd, W.; Fischer, I.; Hammer, P. E.; Hill, Dwight S.; Ligon, J. M.; van Peé, K.H.. (1997). NADH-dependent halogenases are more likely to be involved in halometabolite biosynthesis than haloperoxidases. *Angew Chem Int Ed Engl* 36, 2012-2013.

41. Keller, S., Wage, T., Hohaus, K., Holzer, M., Eichhorn, E. & van Peé, K. H. (2000). Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew Chem Int Ed Engl* 39, 2300-2302.

42. Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R. J., Durkin, A. S., Brinkac, L. M., Daugherty, S. C., Sullivan, S. A., Rosovitz, M. J., Gwinn, M. L., Zhou, L., Schneider, D. J., Cartinhour, S. W., Nelson, W. C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E. A., Pierson, L. S., 3rd, Thomashow, L. S. & Loper, J. E. (2005). Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 23, 873-8.

43. Unversucht, S., Hollmann, F., Schmid, A. & van Pée, K. H. (2005). $FADH_2$-Dependence of Tryptophan 7-Halogenase. *Advanced Synthesis & Catalysis* 347, 1163-1167.

44. Yeh, E., Garneau, S. & Walsh, C. T. (2005). Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc Natl Acad Sci U S A* 102, 3960-5.

45. Sanchez, C., Butovich, I. A., Brana, A. F., Rohr, J., Mendez, C. & Salas, J. A. (2002). The biosynthetic gene cluster for the antitumor rebeccamycin: characterization and generation of indolocarbazole derivatives. *Chem Biol* 9, 519-31.

46. Louie, T. M., Xie, X. S. & Xun, L. (2003). Coordinated production and utilization of FADH2 by NAD(P)H-flavin oxidoreductase and 4-hydroxyphenylacetate 3-monooxygenase. *Biochemistry* 42, 7509-17.

47. Otto, K., Hofstetter, K., Rothlisberger, M., Witholt, B. & Schmid, A. (2004). Biochemical characterization of StyAB from *Pseudomonas* sp. strain VLB120 as a two-component flavin-diffusible monooxygenase. *J Bacteriol* 186, 5292-302.

48. Low, J. C. & Tu, S. C. (2003). Energy transfer evidence for in vitro and in vivo complexes of *Vibrio harveyi* flavin reductase P and luciferase. *Photochem Photobiol* 77, 446-52.

49. Dong, C., Kotzsch, A., Dorward, M., van Peé, K. H. & Naismith, J. H. (2004). Crystallization and X-ray diffraction of a halogenating enzyme, tryptophan 7-halogenase, from *Pseudomonas fluorescens*. *Acta Crystallogr D Biol Crystallogr* 60, 1438-40.

50. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60, 2256-68.

51. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-32.

52. Sheng, D., Ballou, D. P. & Massey, V. (2001). Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis. *Biochemistry* 40, 11156-67.

53. Arunachalam, U., Massey, V. & Miller, S. M. (1994). Mechanism of p-hydroxyphenylacetate-3-hydroxylase. A two-protein enzyme. *J Biol Chem* 269, 150-5.

54. Yeh, E., Cole, L. J., Barr, E. W., Bollinger, J. M., Jr., Ballou, D. P. & Walsh, C. T. (2006). Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH. *Biochemistry* 45, 7904-7912.

55. Galonic, D. P., Vaillancourt, F. H. & Walsh, C. T. (2006). Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *J Am Chem Soc* 128, 3900-1.

56. Vaillancourt, F. H., Yin, J. & Walsh, C. T. (2005). SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc Natl Acad Sci U S A* 102, 10111-6.

57. Zechel, D. L., Reid, S. P., Nashiru, O., Mayer, C., Stoll, D., Jakeman, D. L., Warren, R. A. & Withers, S. G. (2001). Enzymatic synthesis of carbon-fluorine bonds. *J Am Chem Soc* 123, 4350-1.

58. Schaffrath, C., Deng, H. & O'Hagan, D. (2003). Isolation and characterisation of 5'-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Letters* 547, 111-114.

59. Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D. & Naismith, J. H. (2004). Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* 427, 561-5.

60. Deng, H., Cobb, S. L., McEwan, A. R., McGlinchey, R. P., Naismith, J. H., O'Hagan, D., Robinson, D. A. & Spencer, J. B. (2006). The fluorinase from *Streptomyces cattleya* is also a chlorinase. *Angew Chem Int Ed Engl* 45, 759-62.

61. van Peé, K. H. & Ligon, J. M. (2000). Biosynthesis of pyrrolnitrin and other phenylpyrrole derivatives by bacteria. *Nat Prod Rep* 17, 157-64.

62. Arima, K., H. Imanaka, M. Kousaka, A. Fukuda and G. Tamura,. (1964). Pyrrolnitrin, a new antibiotic substance, produced by *Pseudomonas*. *Agric. Biol. Chem.* 28, 575-576.

63. Homma, Y., Z. Sato, F. Hirayama, K. Konno, H. Shirahama and T. Suzui. (1989). Production of antibiotics by *Pseudomonas cepacia* as an agent for biological control of soilborne plant pathogens. *Soil Biol. Biochem.* 21, 723-728.

64. Howell, C. R., and R.D. Stipanovic. (1979). Control of *Rhizoctonia solani* on cotton seedlings with *Pseudomonas fluorescens* and with an antibiotic produced by the bacterium. *Phytopathology* 77, 480-428.

65. Janisiewicz, W. J., and J. Roitman. (1988). Biological control of blue mold and grey mold on apple and pear with *Pseudomonas cepacia*. *Phytopathology* 78, 1697-1700.

66. Chernin, L., A. Brandis, Z. Ismailov and I. Chet. (1996). Pyrrolnitrin production by an *Enterobacter agglomerans* strain with a broad spectrum of antagonistic activity towards fungal and bacterial phytopathogens. *Curr. Microbiol* 32, 208-212.

67. Gerth, K., Trowitzsch, W., Wray, V., Hofle, G., Irschik, H. & Reichenbach, H. (1982). Pyrrolnitrin from *Myxococcus fulvus* (Myxobacterales). *J Antibiot (Tokyo)* 35, 1101-3.

68. Kalbe, C., P. Marten and G. Berg. (1996). Strains of the genus *Serratia* as beneficial rhizobacteria of oilseed rape with antifungal properties. *Microbiol. Res.* 151, 433-439.

69. Floss, H. G., Manni, P. E., Hamill, R. L. & Mabe, J. A. (1971). Further studies on the biosynthesis of pyrrolnitrin from tryptophan by *Pseudomonas*. *Biochem Biophys Res Commun* 45, 781-7.

70. Hill, J. S., NR Torkewitz, AM Morse, CR Howell, JP Pachlatko, JO Becker and JM Ligon. (1994). Cloning of Genes Involved in the Synthesis of Pyrrolnitrin from *Pseudomonas fluorescens* and Role of Pyrrolnitrin Synthesis in Biological Control of Plant Disease. *Appl. Environ. Microbiol* 60, 78-85.

71. Hammer, P. E., Burd, W., Hill, D. S., Ligon, J. M. & van Peé, K. (1999). Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. *FEMS Microbiol Lett* 180, 39-44.

72. Lee, J., Simurdiak, M. & Zhao, H. (2005). Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation. *J Biol Chem* 280, 36719-27.

73. Lee, J. & Zhao, H. (2006). Mechanistic Studies on the Conversion of Arylamines into Arylnitro Compounds by Aminopyrrolnitrin Oxygenase: Identification of Intermediates and Kinetic Studies. *Angew Chem Int Ed Engl* 45, 622-625.

74. Zehner, S., Kotzsch, A., Bister, B., Sussmuth, R. D., Mendez, C., Salas, J. A. & van Peé, K. H. (2005). A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem Biol* 12, 445-52.

75. Ding, W., Williams, D. R., Northcote, P., Siegel, M. M., Tsao, R., Ashcroft, J., Morton, G. O., Alluri, M., Abbanat, D., Maiese, W. M. & et al. (1994). Pyrroindomycins, novel antibiotics produced by *Streptomyces rugosporus* sp. LL-42D005. I. Isolation and structure determination. *J Antibiot (Tokyo)* 47, 1250-7.

76. Abbanat, D., Maiese, W. & Greenstein, M. (1999). Biosynthesis of the pyrroindomycins by *Streptomyces rugosporus* LL-42D005; characterization of nutrient requirements. *J Antibiot (Tokyo)* 52, 117-26.

77. de Boer, H. A., Comstock, L. J. & Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc Natl Acad Sci U S A* 80, 21-5.

78. Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. R. (1980). Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc Natl Acad Sci U S A* 77, 7347-51.

79. Bertani, I., Devescovi, G. & Venturi, V. (1999). Controlled specific expression and purification of 6 x His-tagged proteins in *Pseudomonas*. *FEMS Microbiol Lett* 179, 101-6.

80. Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J. & Ausubel, F. M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium mutants*. *Gene* 18, 289-96.

81. Staskawicz, B., Dahlbeck, D., Keen, N. & Napoli, C. (1987). Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. glycinea. *J Bacteriol* 169, 5789-94.

82. (March 2003). QIAprep Miniprep Handbook. *QIAGEN*.

83. (August 2003). QIAGEN Plasmid Purification Handbook. *QIAGEN*.

84. (July 2002). QIAquick spin Handbook. *QIAGEN*.

85. (October 2003). NuPAGE Technical Guide. *INVITROGEN* Version E.

86. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *EAMCB Newsletter on Protein Crystallography* 26.

87. Evans, P. R. (1997). Joint CCP4 and ESF-EACBM. *Newsletter on Protein Crystallography* 33, 22-24.

88. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-3.

89. Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr* 57, 1373-82.

90. Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* 12, 1865-71.

91. Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-7.

92. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* 57, 122-33.

93. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-55.

94. Wynands, I. & van Peé, K. H. (2004). A novel halogenase gene from the pentachloropseudilin producer *Actinoplanes* sp. ATCC 33002 and detection of in vitro halogenase activity. *FEMS Microbiol Lett* 237, 363-7.

95. Krissinel E. & Henrick K. (2005). Detection of Protein Assemblies in Crystals. *Lecture Notes in Computer Science* 3695, 163-174.

96. Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* 23, 566-79.

97. Thompson, M. A. (2004). ArgusLab 4.0.1. *Planaria Software LLC Seattle, WA*.

98. Sugimoto, H., Oda, S. I., Otsuki, T., Hino, T., Yoshida, T. & Shiro, Y. (2006). Crystal structure of human indoleamine 2,3-dioxygenase: Catalytic mechanism of O2 incorporation by a heme-containing dioxygenase. *Proc Natl Acad Sci U S A* 103, 2611-2616.

99. Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A., Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R. T., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J. A., Timmis, K. N., Dusterhoft, A., Tummler, B. & Fraser, C. M. (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* 4, 799-808.

100. Regenhardt, D., Heuer, H., Heim, S., Fernandez, D. U., Strompl, C., Moore, E. R. & Timmis, K. N. (2002). Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environ Microbiol* 4, 912-5.

101. (September 2003). GATEWAY TECHNOLOGY. *INVITROGEN* Version E.

102. (October 2003). *E. coli* EXPRESSION SYSTEM WITH GATEWAY TECHNOLOGY. *INVITROGEN* Version D.

103. (August 2001). QIAGEN Genomic DNA Handbook. *QIAGEN*.

104. (July 2002). pET SYSTEM MANUAL. *NOVAGEN* 10th Edition.

105. (July 2003). The QIAexpressionist™. *QIAGEN* 5th Edition.

106. Van Peé, K. H. S., Olga; Lingens, Franz. (1980). Formation of pyrrolnitrin and 3-(2-amino-3-chlorophenyl) pyrrole from 7-chlorotryptophan. *Angewandte Chemie, International Edition in English* 92, 855-856.

107. Pesce, A., Dewilde, S., Nardini, M., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. & Bolognesi, M. (2003). Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* 11, 1087-95.

108. Vallone, B., Nienhaus, K., Matthes, A., Brunori, M. & Nienhaus, G. U. (2004). The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc Natl Acad Sci U S A* 101, 17351-6.

109. Takikawa, O. (2005). Biochemical and medical aspects of the indoleamine 2,3-dioxygenase-initiated L-tryptophan metabolism. *Biochem Biophys Res Commun* 338, 12-9.

110. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55, 849-61.

111. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56, 965-72.

112. Terwilliger, T. C. (2002). Automated structure solution, density modification and model building. *Acta Crystallogr D Biol Crystallogr* 58, 1937-40.

113. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6, 458-63.

114. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98, 10037-41.

115. DeLano, W. L. D. S., San Carlos, CA, USA. (2002). The PyMOL Molecular Graphics System. *DeLano Scientific, San Carlos, CA, USA*.

116. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8, 127-34.

117. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7, 1884-97.

118. Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res* 31, 3352-5.

119. Terentis, A. C., Thomas, S. R., Takikawa, O., Littlejohn, T. K., Truscott, R. J., Armstrong, R. S., Yeh, S. R. & Stocker, R. (2002). The heme environment of recombinant human indoleamine 2,3-dioxygenase. Structural properties and substrate-ligand interactions. *J Biol Chem* 277, 15788-94.

120. Littlejohn, T. K., Takikawa, O., Truscott, R. J. & Walker, M. J. (2003). Asp274 and his346 are essential for heme binding and catalytic function of human indoleamine 2,3-dioxygenase. *J Biol Chem* 278, 29525-31.

# ACKNOWLEGMENTS

I would like to extend a warm thank you to my supervisor Professor J.H. Naismith for his gentle guidance, support and always sage advice throughout this thesis project.

Thanks also go to the members of my examining comitee, Professor C.J. Schofield and Professor D. O'Hagan, for reviewing my work and providing helpful commentary.

Acknowledgment is made to my friends and colleagues in the Biomolecular Sciences Department, University of St Andrews; to our collaborators in Professor K.H. van Pee laboratory at the Institut für Biochemie, Technische Universität Dresden, Germany and in Professor S.K. Chapman laboratory at the School of Chemistry, University of Edinburgh.

Finally, I wish to express my deepest gratitude to the School of Chemistry, University of St. Andrews and to the The Biotechnology and Biological Sciences Research Council for giving me the opportunity and the financial support to carry out my thesis project.