University of Kent

Kent Academic Repository

Palmer, David James (2014) *The Alternative Haem Biosynthesis Pathway: Structure, Function and Properties of Sirohaem Decarboxylase.* Doctor of Philosophy (PhD) thesis, University of Kent,.

Downloaded from

https://kar.kent.ac.uk/47619/ The University of Kent's Academic Repository KAR

The version of record is available from

This document version

UNSPECIFIED

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

The Alternative Haem Biosynthesis Pathway:

Structure, Function and Properties of Sirohaem Decarboxylase

A thesis submitted to the University of Kent for the degree of PhD in the Faculty of Sciences

2014

David James Palmer

Declaration

Name: David J Palmer

Degree: PhD-Biochemistry

Title: The Alternative Haem Biosynthesis Pathway: Structure, Function and Properties of Sirohaem Decarboxylase

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Abstract

Haem, a cyclic tetrapyrrole, is found in organisms from all three domains of life. Haem is a prosthetic group for many proteins involved in essential biological processes such as respiration and oxygen transport. Synthesis of haem in eukaryotes and most bacteria follows a well defined route with highly conserved intermediates. However, an alternative haem biosynthesis pathway in Archaea and some bacteria was recently elucidated. This newly discovered pathway utilises sirohaem as a metabolic intermediate rather than a prosthetic group.

The alternative haem biosynthesis pathway is catalysed by the Ahb enzymes A, B, C and D. Initial decarboxylation of sirohaem occurs at the two acetic acid side-chains attached to carbons C12 and C18 to give didecarboxysirohaem, a process catalysed by AhbA and AhbB. Subsequently, the radical SAM enzyme AhbC converts didecarboxysirohaem to Fe-coproporphyrin III. Finally, AhbD catalyses the conversion of Fe-coproporphyrin III into haem in another SAM dependent reaction.

This study focused on the AhbA and AhbB proteins from three sources, *Desulfovibrio desulfuricans*, *Desulfovibrio vulgaris* and *Methanosarcina barkeri*. Purifications of individual recombinantly produced proteins revealed that both AhbA and AhbB are highly unstable. However, low concentrations of *D. desulfuricans* and *D. vulgaris* AhbA and AhbB proteins were isolated and were discovered to have decarboxylase activity. Simultaneous overproduction of AhbA and AhbB proteins facilitated the co-purification of stable heteromeric AhbA/B complexes from all three organisms. However, despite their sequence similarities, distinctly different properties were observed between the homologues including differences in oliogmeric state and haem/product binding capabilities. The *D. desulfuricans* enzyme has been crystallised and its structure has been elucidated in both apo- and product-bound forms. Mutagenesis of the *D. desulfuricans* complex has provided further information about active site residues and a mechanism of action has been proposed. Finally, novel functional chimeric complexes were produced using the *Desulfovibrio* proteins.

First Author Papers Published During PhD

Palmer DJ, Schroeder S, Lawrence AD, Deery E, Lobo SA, Saraiva LM, McLean KJ, Munro AW, Ferguson SJ, Pickersgill RW, Brown DG, Warren MJ, (2014) The structure, function and properties of sirohaem decarboxylase--an enzyme with structural homology to a transcription factor family that is part of the alternative haem biosynthesis pathway. *Mol Microbiol*. **93**:247-61

Bali, S.[*], D.J. Palmer[*], S. Schroeder[*], S.J. Ferguson & M.J. Warren, (2014) Recent advances in the biosynthesis of modified tetrapyrroles: the discovery of an alternative pathway for the formation of heme and heme d_1. *Cell Mol Life Sci*. **71**:2837-63. (*Co-authors)

Contents

Figures

Tables

Abbreviations

Å	Angstrom
ALA	5-Aminolevulinic acid
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
CX	Carbon X
CoA	Coenzyme A
Copro'gen III	Coproporphyrinogen III
Da	Dalton
DEAE	Diethylaminoethyl
dH_2O	Distilled water
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
EPR	Electron paramagnetic resonance
ESI	Electrospray ionisation
FAD	Flavin adenine dinucleotide (oxidised form)
Fe-S	Iron-sulphur
FMN	Flavin adenine mononucleotide
FPLC	Fast protein liquid chromatography
HCl	Hydrochloric acid
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulphonic acid
HPLC	High performance liquid chromatography
HtH	Helix turn helix
IPTG	Isopropyl-β-D-thiogalactopyronoside
kb	Kilobase pair
LB	Luria-Bertani
MES	2-(N-morpholino)ethanesulphonic acid
MHz	Megahertz
MPD	2-Methyl-2,4-pentanediol
MS	Mass spectrometry
mV	Millivolt
NAD^+	Nicotinamide adenine dinucleotide (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
nm	Nanometers
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBG	Porphobilinogen
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PLP	Pyridoxal phosphate
PMP	Pyridoxamine phosphate
ppm	parts-per million
Proto'gen IX	Protoporphyrinogen IX
RAM	Regulation of amino acid
RNA	Ribonucleic acid

SAM	*S*-Adenosyl-L-methionine
SDS	Sodium dodecyl sulphate
TAE	Tris acetate EDTA
TCA	Trichloroacetic acid
TFA	Trifluoroacetic acid
TEMED	Tetramethylenediamine
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
tRNA	Transfer ribonucleic acid
Uro'gen III	Uroporphyrinogen III
UV	Ultra violet

Acknowledgements

Firstly, I would like to thank Prof. Martin Warren for giving me the opportunity to do this PhD and for all of the help and support he has given me. I am also very grateful for the funding provided by the BBSRC which allowed me to complete this work.

I would like to thank all of the people who have been in the Warren Lab throughout these four years; your help and friendship has made it an exceptionally enjoyable experience. I would especially like to thank Dr. Evelyne Deery who provided me with plasmids and schooled me in the art of cloning, and Dr. Andrew Lawrence who helped me with a wide array of things from mass spectrometry, to NMR and kinetics.

I would also like to say a massive thank you to Dr. Susanne Schroeder who took me under her wing from an early stage and gave me so much support and advice, from simple chats about experiments to hours and hours of proof reading. You have taught me so much, I couldn't have done it without you and I am eternally grateful.

Huge thanks go to Dr. Susana Lobo for providing me with the original DNA that got this project started and for continued advice and ideas.

I would like to express my gratitude toward Prof. Dave Brown. He provided me with an inordinate amount of his time, from the sleepless nights of data collection at the synchrotron to the hours spent trying to solve the structure of the AhbA/B complex. I would also like to thank him and his team at Argenta for all of the crystallography advice they gave me and for always being helpful and kind no matter how many times I came down to bother them. Special thanks go to Dr. Stephen Irving for taking the time to teach me how to do the thermal shift assay.

From Manchester University I would like to thank Prof. Andrew Munro and Dr. Kirsty McLean for performing the redox potentiometry experiments and Dr. Steve Rigby for running EPR experiments.

I would like to thank Dr. Mark Howard for running the NMR experiments. I would also like to thank Prof. Richard Pickersgill for help with crystallography, especially for the emergency data collection of product-bound crystals.

Finally I would like to thank my family and friends for all their support throughout these past years. I am especially grateful to Sinead for putting up with all the late nights and the madness that has occurred throughout this experience and providing so much moral support, you have been amazing.

Chapter 1

Introduction

1.1 Tetrapyrroles - The pigments of life

Tetrapyrroles are a class of molecules which, as the name suggests, consist of four polymerised pyrrole rings arranged either in a linear fashion or as a macrocycle, usually connected via 4 methine bridges. The rings are labelled A-D in a clockwise direction (Fig. 1.1). Linear tetrapyrroles are formed by the cleavage of cyclic tetrapyrroles and therefore contain 3 methine bridges, as do the cyclic corrinoids where the carbon 20 has been extruded. Modified tetrapyrroles are found in almost all organisms across the three domains of life (Bacteria, Eukaryota and Archaea) and participate in a diverse array of indispensable biological reactions. These reactions include electron and oxygen transport, and probably the most fundamental reaction required for life on Earth, photosynthesis. Naturally occurring tetrapyrroles include chlorophylls, haems, cobalamins, sirohaem, coenzyme F_{430} and haem d_1 (Banerjee & Ragsdale, 2003, Battersby, 2000, Thauer & Bonacker, 1994, Warren *et al.*, 2009).

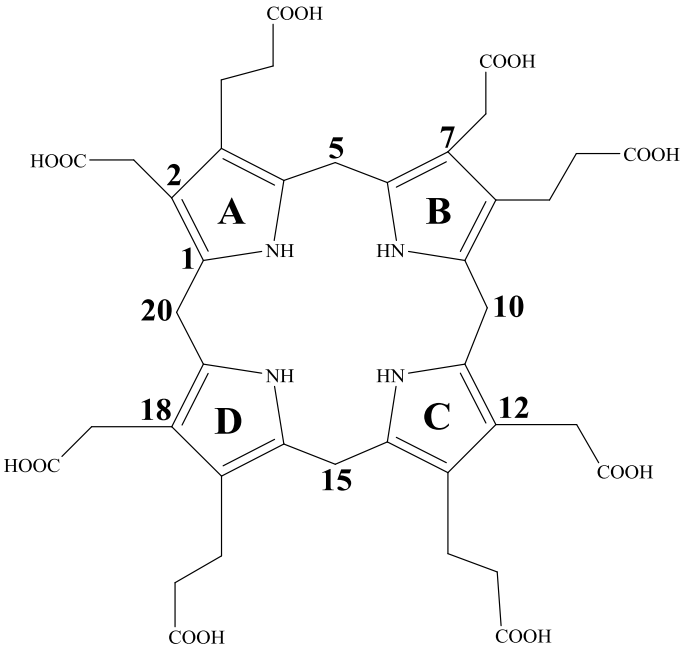


Figure 1.1. The structure of uroporphyrinogen III (uro'gen III). Carbon atoms that are frequently referred to in this thesis are labelled, as are the four pyrrole rings A-D.

Modified cyclic tetrapyrroles are able to chelate a metal ion in the centre of the ring system, including Fe, Ni, Co, and Mg. The centrally chelated metal ion varies among the members of the family along with the oxidation state of the macrocycle and ring substituents all of which allow for the unique functional variation found within this class of molecules (Frankenberg *et al.*, 2003). The combination of the conjugated ring system and bound metal ion gives

modified tetrapyrroles their distinctive colourations, from the green of chlorophyll to the deep purple of sirohaem. Due to their essential role in nature and characteristic colours they have been referred to as the pigments of life (Battersby, 2000). The similarities between cyclic tetrapyrroles are a consequence of their synthesis. All cyclic tetrapyrroles are derived from a central 'blueprint' molecule uroporphyrinogen (uro'gen III). From this molecule, following an array of modifications including metal insertion, side chain modification and ring oxidation an assortment of modified tetrapyrroles can be formed (Fig. 1.2).

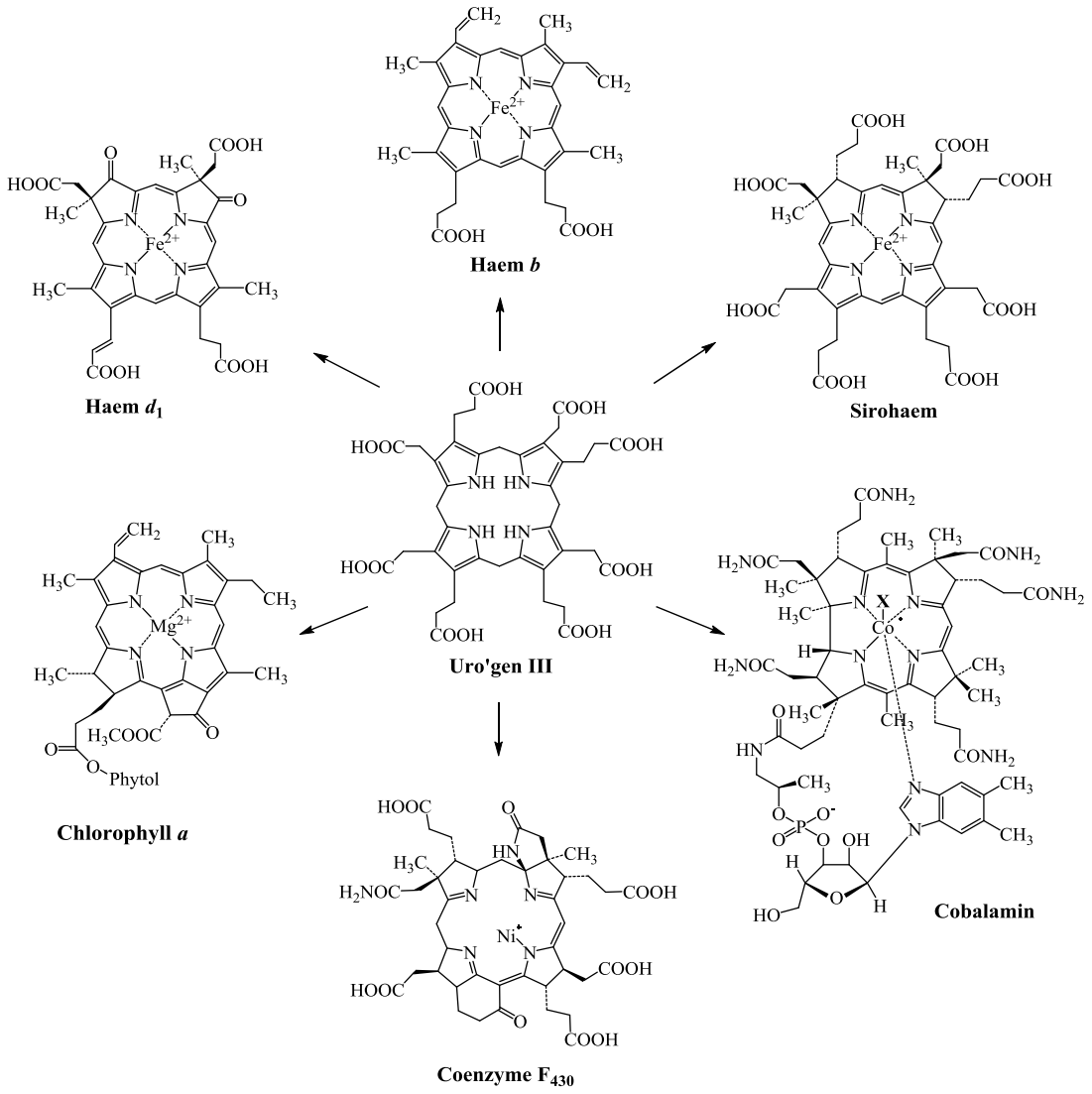


Figure 1.2. Important members of the cyclic tetrapyrrole family stemming from the common precursor uro'gen III.

1.1.1 Haems

Haems, iron containing porphyrins with variants *a*, *b*, *c* and *o*, have a plethora of functionalities, from oxygen and carbon dioxide transport in the cardiovascular system, to being cofactors to enzymes such as catalases, peroxidases and cytochrome P450s (Matsunaga & Shiro, 2004). As haem *b* is the subject of this thesis, more details of the structure and function of haems are given in a later section.

1.1.2 Chlorophylls

Perhaps the best known tetrapyrroles are the chlorophylls, which is reflected in the fact that they are the most abundant of all tetrapyrroles in nature with an estimated 10^9 tons being produced each year in the changing seasons (Rudiger, 1997). Chlorophylls and bacteriochlorophylls are magnesium chelating tetrapyrroles and are indispensable components of the photosynthetic apparatus utilised by plants and bacteria to harness energy from sunlight and convert it to chemical energy for the organism (Bollivar, 2006). The production of oxygen as a by-product of photosynthesis is responsible for the formation of an aerobic atmosphere and the development of advanced eukaryotic life on Earth (Blankenship & Hartman, 1998). Bacteriochlorophyll is the bacterial equivalent of chlorophyll and may represent the evolutionary ancestor of chlorophyll (Lockhart *et al.*, 1996, Blankenship & Hartman, 1998).

1.1.3 Cobalamin

Cobalamins and other cobalt containing corrinoids are unique in the sense that they lack the methine bridge between rings A and D (Friedmann & Thauer, 1992). Cobalamin is the most complex of all the tetrapyrroles. It contains a lower nucleotide loop formed of dimethylbenzimidazole, as well as an upper axial ligand, which can be water (aquacobalamin), a methyl group (methylcobalamin), adenosine (adenosylcobalamin), or cyanide (vitamin B_{12} or cyanocobalamin). Methyl- and adenosyl-cobalamin represent the two active forms of cobalamin, utilising the properties of the cobalt carbon bond to augment a range of enzymatic processes including, rearrangement, methylation and reductive dehalogenation reactions (Banerjee & Ragsdale, 2003).

1.1.4 Sirohaem

Sirohaem is an iron containing isobacteriochlorin involved in assimilatory nitrite and sulphite reduction (Fritz *et al.*, 2005, Richardson & Watmough, 1999). Sirohaem is considered to be one of the 'simpler' tetrapyrroles, requiring relatively few modifications for its synthesis beyond uro'gen III. Synthesis of sirohaem is discussed in greater detail later in this chapter.

1.1.5 Coenzyme F_{430} & haem d_1

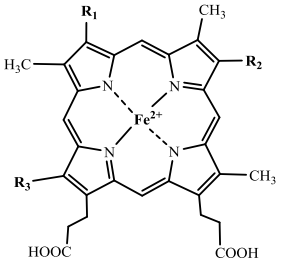
Coenzyme F_{430} and haem d_1 remain some of the more elusive tetrapyrroles with relatively little known about their syntheses. Coenzyme F_{430} is a yellow nickel-containing tetrapyrrole and is involved in methanogenesis as the cofactor of methyl-coenzyme M reductase (Thauer & Bonacker, 1994). Despite its name, haem d_1 is not actually a member of the haem family as it is not a true porphyrin. In fact haem d_1 is an iron containing isobacteriochlorin. Haem d_1 is a cofactor to a single enzyme, cytochrome cd_1 nitrate reductase, involved in the conversion of nitrite to nitric oxide in denitrifying bacteria (Zajicek *et al.*, 2009). The current understanding of the synthesis of haem d_1 is discussed in more detail later in this chapter.

1.2 Structure and function of haems

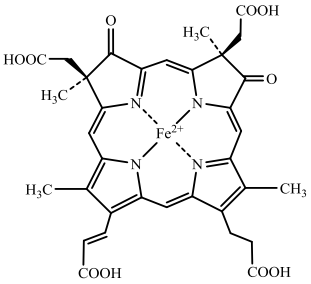
Haems are categorised according to their side chain modifications and belong to one of the four types: *a*, *b*, *c* and *o* (Fig. 1.3). All haems are derived from haem *b* which is often referred to as protohaem (Kranz *et al.*, 2009). Haems *a* and *o* have a farnesyl group in place of the vinyl side chain at carbon C3. Haem *a* also has a formyl group replacing the methyl at carbon C18.

Haem *a* is the prosthetic group of cytochrome *a*, which forms part of the terminal enzyme complex of the respiratory chain responsible for reduction of dioxygen to water, causing release of energy in the form of a transmembrane electrochemical gradient (Hederstedt, 2012). Haem *a* is found in bacteria, archaea, plants and animals. Haem *o* was first discovered as the prosthetic group of cytochrome *o* from *Escherichia coli* (Puustinen & Wikstrom, 1991), but it was subsequently found that some archaea contain novel haem *o* variants as cofactors of cytochrome oxidases (Lubben & Morand, 1994).

Haem	R_1	R_2	R_3
a	$-C_{17}H_{29}O$	$-CH=CH_2$	$-CH=O$
b	$-CH=CH_2$	$-CH=CH_2$	$-CH_3$
c	$-CH(S\text{-protein})-CH_2$	$-CH(S\text{-protein})-CH_2$	$-CH_3$
o	$-C_{17}H_{29}O$	$-CH=CH_2$	$-CH_3$



Haem framework



Haem d_1

Figure 1.3. Structures of biologically important variants of haem with the dioxo-isobacteriochlorin haem d_1 for comparison. Ligands at different residues on the haem macrocycle are donated R_1, R_2, and R_3 and are listed on the left. In archaea, the side chains at R_1 of haem *a* and *o* can have longer groups (Lubben & Morand, 1994).

Haem *b* has a diverse array of functions and is found in almost all forms of life. Haem *b* is responsible for binding and transport of oxygen and carbon dioxide as a prosthetic group of haemoglobin and myoglobin (O'Brian & Thony-Meyer, 2002). Cytochrome *b* utilises haem *b* molecules for electron transfer in the protonmotive Q cycle mechanism in the electron transport chain; it is therefore an indispensable component for energy generation in many organisms (Trumpower, 2002). Some enzymes, including peroxidases and catalases, employ a haem *b* cofactor to produce a highly reactive radical, compound I (oxo-ferryl porphyrin π-cation radical), to catalyse a variety of reactions including oxidation, peroxidation, oxygenation and peroxygenation (Matsunaga & Shiro, 2004). The cytochrome P450 enzymes are another class of proteins which utilise haem *b* as a cofactor, allowing them to catalyse complex reactions involved with the detoxification of endogenous biochemicals and xenobiotics (Munro *et al.*, 2007). As well as having an extensive role as an enzyme cofactor, haem *b* has numerous other applications including being a source of iron, a transcriptional regulator, a component in diatomic gas sensing and having a role in pathogenicity control (Wilks & Burkhard, 2007, Rodgers, 1999).

In contrast to the other haems, haem *c* is covalently attached to proteins via thioether bonds between two cysteine residues and the Cα of the vinyl groups at carbons C3 and C8 (Kranz *et al.*, 2009). The so-called cytochrome *c* proteins contain a conserved CXXCH binding motif which provides the cysteines required for covalent modification and axial histidine ligand for the haem iron (Barker & Ferguson, 1999). Multiple copies of this motif can be present in a single polypeptide chain leading to a single protein with several covalently bound haem groups. Cytochrome *c* proteins are involved in the aerobic respiratory chain as

well as photosynthesis and reduction of terminal electron acceptors such as nitrates, dimethyl sulfoxide and metals (Richardson, 2000). As this thesis is restricted to the discussion of the synthesis of *b*-type haem further references to 'haem' should be considered to be referring to haem *b*.

1.3 Haem biosynthesis

1.3.1 An overview of the pathway

Haem biosynthesis can be separated into three sections: (1) synthesis of 5-aminolevulinic acid (ALA), (2) formation of uro'gen III, (3) decoration of the ring system. The first two steps are common in the biosynthesis pathways of all cyclic tetrapyrroles. ALA, a small 5 carbon aminoketone, is the progenitor of all cyclic tetrapyrroles and can be synthesised by one of two routes. The C-4 or 'Shemin' pathway is established in animals, fungi and the α-proteobacteria (Shemin & Rittenberg, 1946, Kikuchi *et al.*, 1958) and uses succinyl coenzyme A (succinyl-CoA) and glycine as substrates in a single step reaction to form ALA. The C-5 pathway, found in plants, archaea and most bacteria, uses glutamic acid to form ALA via glutamyl tRNA and glutamate-1-semialdehyde (Beale & Castelfranco, 1973, Jahn, 2009). In general organisms employ one of the two pathways, although it has been shown that the photosynthetic organism *Euglena gracilis* contains functional enzymes from both pathways (Weinstein & Beale, 1983, Iida *et al.*, 2002). A summary of ALA synthesis is shown in Figure 1.4.

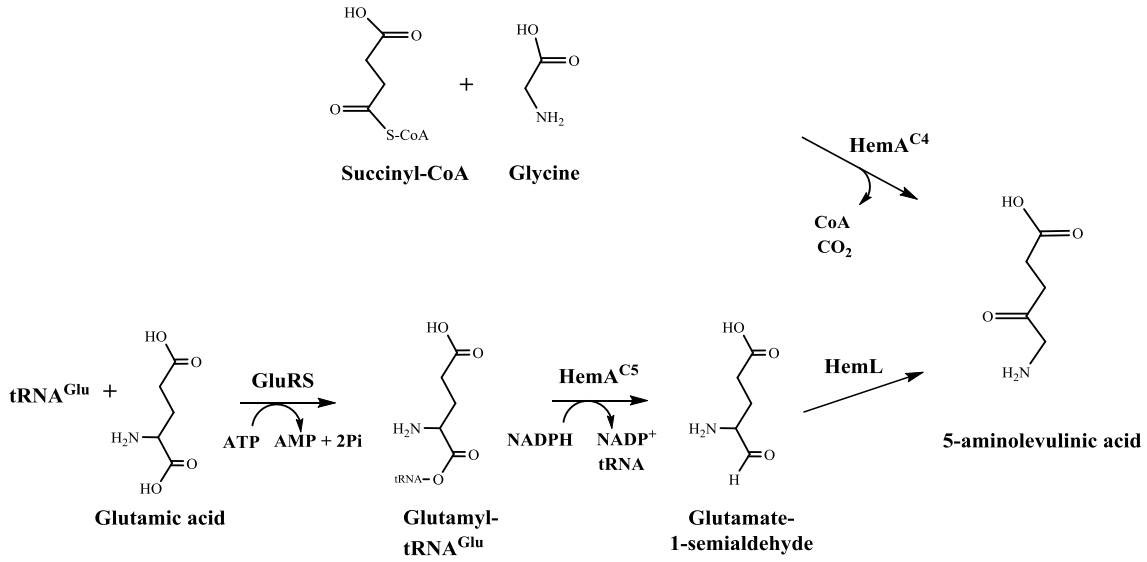


Figure 1.4. A summary of ALA synthesis via the Shemin and C-5 pathways. ALA can be synthesised by two distinct routes. The C-4 or Shemin pathway (top) uses succinyl-CoA and glycine in a single step reaction. The C-5 pathway (bottom) proceeds from glutamyl-tRNA and glutamic acid to ALA in a three step process.

After the formation of ALA, all organisms use a highly conserved pathway for the formation of the common cyclic tetrapyrrole intermediate uro'gen III. Porphobilinogen (PBG) is formed by the condensation of two molecules of ALA. Four molecules of PBG are then oligomerised to form pre-uroporphyrinogen (often referred to as hydroxymethylbilane). Subsequent ring closure and inversion of ring D produces the type III isomer of uro'gen (Fig. 1.5). Therefore a total of eight molecules of ALA make up a single molecule of uro'gen III. The fact that uro'gen III is asymmetric may allow recognition of the molecule by enzymes facilitating selective modifications to the ring system. From the scaffold of uro'gen III a diverse array of chemical modifications can occur to form one of the many naturally occurring cyclic tetrapyrroles. The transformation of uro'gen III to haem requires decaboxylation of acetate side chains at carbons C2, C7, C12 and C18, conversion of propionate groups at carbons C3 and C8 to vinyl groups, oxidation of the ring system, and iron insertion.

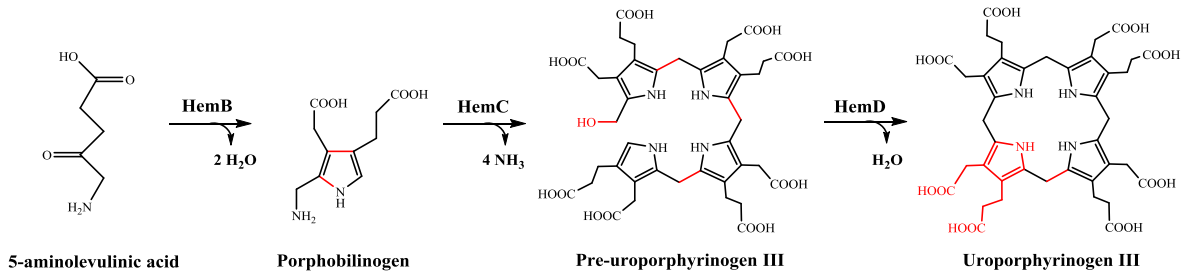


Figure 1.5. Synthesis of uro'gen III from ALA. Synthesis of uro'gen III occurs via a highly conserved route in all organisms that synthesise cyclic tetrapyrroles. Chemical modifications are labelled in red.

Enzymes involved in the biosynthesis of haem in eukaryotes are highly conserved and have been well characterised (Panek & O'Brian, 2002). Equally many prokaryotes follow the same conserved pathway with homologous enzymes, some of which are adapted for anaerobic haem synthesis (Panek & O'Brian, 2002). However, it has been noted that some bacteria and archaea lack the genes necessary for completion of haem synthesis via this route (Cavallaro *et al.*, 2008). Indeed it has been demonstrated that they follow a distinctly different route from uro'gen III, via the formation of sirohaem and novel intermediates (Bali *et al.*, 2011). Therefore, to distinguish between the two pathways they have been termed the 'classical' and 'alternative' pathways and will be referred to as such throughout. A summary of the two haem biosynthesis pathways is shown in Fig. 1.6.

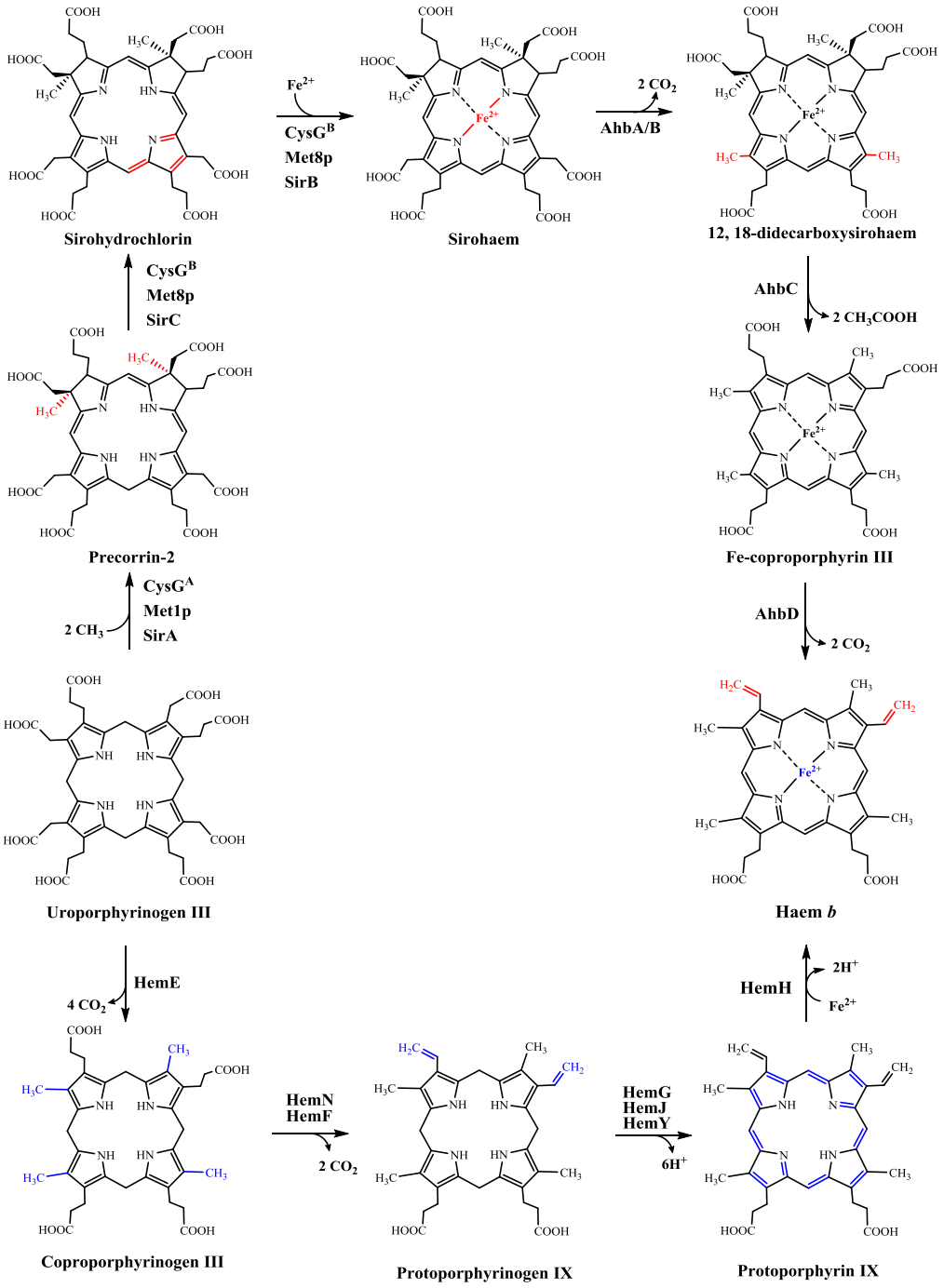


Figure 1.6. Classic and alternative haem biosynthesis pathways. Classical haem synthesis route (bottom) proceeds from uro'gen III to haem in 4 steps with the final stage being iron insertion. For alternative haem biosynthesis (top) uro'gen III must first be converted to sirohaem. Modifications are labelled in blue for classical and red for sirohaem/alternative haem biosynthesis pathways.

1.3.1.1 The classical haem biosynthesis pathway

The classical biosynthesis pathway proceeds from uro'gen III by the decarboxylation of all four acetate side chains yielding coproporphyrinogen. This is followed by the oxidative decarboxylation of the northern propionate side chains (rings A and B) forming protoporphyrinogen IX. Subsequent removal of six protons and six electrons and finally chelation of iron into the macrocycle yields the completed haem *b* molecule.

1.3.1.2 The alternative haem biosynthesis pathway

To start the alternative haem biosynthesis pathway uro'gen III must first be converted to sirohaem. This is initiated by the S-adenosyl-L-methionine (SAM) dependent methylation of carbons C2 and C7 to form precorrin-2. The ring system is then oxidised to produce sirohydrochlorin and subsequent iron chelation converts it to sirohaem. From sirohaem a double decarboxylation of acetate side chains occurs to form 12,18-didecarboxysirohaem. Removal of the northern acetate side chains (rings A and B) produces iron coproporphyrin III. Finally, oxidative decarboxylation of the northern propionate side chains (rings A and B) to vinyl groups concludes haem *b* synthesis.

1.3.2 Synthesis of 5-aminolevulinic acid

1.3.2.1 The C-4 or Shemin pathway

The C-4 pathway is the simpler of the two routes to ALA formation as it is a single step process through the condensation of glycine and succinyl-CoA. To demonstrate the incorporation of glycine into tetrapyrroles Shemin ingested ^{15}N labelled glycine resulting in the isotopic labelling of his own haem (Shemin & Rittenberg, 1946). The requirement for succinyl-CoA was later demonstrated using radioactively labelled acetate (Radin *et al.*, 1950, Shemin & Kumin, 1952, Muir & Neuberger, 1950).

ALA synthase (ALAS or $HemA^{C4}$) catalyses the condensation of glycine and succinyl-CoA in a pyridoxal-5'-phostphate (PLP) dependent manner (Nandi, 1978, Alexander *et al.*, 1994). ALAS has been purified from many bacterial and eukaryotic sources and has been biochemically characterised (Ferreira & Gong, 1995). Interestingly, bacterial sources usually contain a single isoform of ALAS, whereas mammals have two, the 'housekeeping' isoform (ALAS1) and the secondary isoform (ALAS2) which maintains high levels of haem

synthesis in erythrocytes (Bishop *et al.*, 1990). The full length ALAS structure from *Rhodobacter capsulatus* is the only one to have been solved to date, allowing further insight to the mechanism of action of the enzyme (Astner *et al.*, 2005). ALAS acts as a homodimer with two active sites, each binding a PLP molecule via a conserved lysine residue when in the substrate-free state. Addition of glycine causes transaldimination allowing the PLP to bind directly to glycine instead of the lysine. PLP-glycine then nucleophilically attacks succinyl-CoA, initiating addition of succinic acid to glycine and the loss of CoA. Finally, decarboxylation of the glycine moiety of this intermediate yields ALA bound to PLP ready for release.

1.3.2.2 The C-5 pathway

Surprisingly, the ALA synthesis pathway in plants was not discovered for a further 25 years, long after the C-4 pathway had been established and characterised. It was clear that using labelled glycine in plant tissues did not produced the same incorporation pattern in ALA that had been observed in organisms using the C-4 pathway. The use of ^{14}C labelled glutamate, α-ketoglutarate and glutamine was employed and it was discovered that these yielded ^{14}C labelled ALA (Beale & Castelfranco, 1973). Later it was revealed that the entire carbon skeleton from glutamate was incorporated into ALA (Beale *et al.*, 1975). This work lead to the unearthing of the C-5 pathway; highly distinct from the C-4 pathway it is a multi-enzyme process that converts glutamic acid to ALA.

The enzyme GluRS catalyses the addition of glutamic acid to tRNA, producing glutamyl-tRNA ($tRNA^{Glu}$), which is often considered the 'initial' substrate of the C-5 pathway (Jahn *et al.*, 1992). Both tetrapyrrole and protein biosyntheses utilise $tRNA^{Glu}$. Subsequently, the NADPH-dependent reduction of $tRNA^{Glu}$ by the enzyme glutamyl-tRNA reductase (GluTR or $HemA^{C5}$) occurs, converting it to glutamate-1-semialdehyde [GSA; (Moser *et al.*, 1999, Chen *et al.*, 1990)]. Finally GSA is converted to ALA in a PLP-dependent manner by glutamate-1-semialdehyde-2,1-aminomutase [GSAM or HemL; (Hoober *et al.*, 1988)].

GluTR functions as a homodimer and recognises the tRNA substrate by its size and shape (Randau *et al.*, 2004). Use of NADPH analogues also showed that GluTR is highly specific for NADPH (Moser *et al.*, 1999). The structure of *Methanopyrus kandleri* GluTR has been solved which revealed it to be an unusual extended V-shaped dimer linked at the C-terminal domain of each monomer. During the reaction a conserved cysteine acts as a nucleophile attacking the α-carbonyl group of tRNA-bound glutamate resulting in an enzyme-bound

thioester intermediate. Hydride transfer from NADPH to the thioester-bound glutamate produces GSA. Without NADPH the reactive thioester intermediate is hydrolysed releasing free glutamate (Schauer *et al.*, 2002).

There are two possible reaction mechanisms for the transamination of GSA to ALA and all GSAM proteins characterised to date have been shown to be able to catalyse both reactions (Smith *et al.*, 1991, Ilag & Jahn, 1992). The first mechanism involves GSAM-bound pyridoxamine 5'-phosphate (PMP). In this instance the PMP reacts with the aldehyde of GSA creating the intermediates 4,5-diaminovalerate (DAVA) and PLP. The C4 amino group of DAVA is then abstracted, regenerating PMP and yielding ALA. The second mechanism starts with PLP as a cofactor and the reaction proceeds via the intermediates 4,5-dioxovalerate (DOVA) and PMP, with the final step regenerating PLP and producing ALA.

Crystal structures have been determined for two GSAM proteins, one from *Synechococcus* (Hennig *et al.*, 1997) the other from *Thermosynechococcus elongates* (Schulze *et al.*, 2006). Interestingly, both demonstrated differing dimer conformations and bound cofactors. *Synechococcus* revealed an asymmetric dimer containing a covalently bound PLP in one monomer and a non-covalent PMP in the other. The *T. elongates* enzyme was crystallised as a symmetric homodimer containing one PLP in each monomer. These results leave the true mechanism of action for GSAM still to be determined.

Given the unstable nature of GSA it was proposed that the intermediate could be protected from the surrounding environment by the formation of a GluTR-GSAM complex (Layer *et al.*, 2010). Based on surface complementarity between the two proteins it was suggested that direct substrate channelling occurred preventing exposure of GSA to the cellular environment. Indeed it has been shown that GluTR and GSAM form a complex in both *Chlamydomonas reinhardtii* and *E. coli* (Nogaj & Beale, 2005, Luer *et al.*, 2005).

1.3.3 Synthesis of uroporphyrinogen III

From ALA, uro'gen III synthesis is completed in 3 further steps that are highly conserved in all organisms. The first step is the asymmetric condensation of two molecules of ALA to form the pyrrole PBG. This reaction is catalysed by the enzyme porphobilinogen synthase (PBGS) also known as ALA dehydratase (ALAD) or HemB (Gibson *et al.*, 1954). Despite a high level of sequence similarity between PBGS proteins metal dependency of these enzymes varies between organisms (Jaffe, 2000). For example, PBGS enzymes from yeast

and mammals contain two zinc ions (Senior *et al.*, 1996), whereas the enzyme from *Psuedomonas aeruginosa* contains magnesium instead (Shoolingin-Jordan *et al.*, 2002). Dual zinc and magnesium proteins also exist, in *E. coli* for example (Jaffe, 2000). However, there is also evidence of PBGS enzymes that are functional independent of metals (Bollivar *et al.*, 2004).

Structural studies showed that the fundamental subunit of PBGS is formed from an asymmetric homodimer that arranges into a final octameric structure. Substrate binding is achieved by two conserved lysine residues with the initial ALA molecule binding the 'P-site' that will form the propionate group on PBG, and the secondary ALA binding the 'A-site', to form the acetate side chain (Jordan & Gibbs, 1985). After formation of a Schiff base between the lysine residues and ALA molecules the reaction continues with C-C bond formation between carbon C3 of the A-site ALA and carbon C4 of the P-site ALA. Consequently, the C-N bond forms between the ALA molecules and final deprotonation of the P-site ALA yields the completed PBG molecule (Frere *et al.*, 2002, Jaffe *et al.*, 2002, Kervinen *et al.*, 2001).

The subsequent stage in uro'gen synthesis is the oligomerisation of four molecules of PBG to form the linear tetrapyrrole pre-uroporphyrinogen [hydroxymethylbilane (Warren & Jordan, 1988)]. This reaction is catalysed by the enzyme porphobilinogen deaminase (PBGD) also known as hydroxymethylbilane synthase or HemC. PBGD contains a novel dipyrromethane cofactor that is in fact produced from two molecules of PBG (Hart *et al.*, 1987, Jordan & Warren, 1987). Dipyrromethane is covalently linked to the enzyme and is utilised as a primer for the ligation of the PBG molecules. This occurs sequentially starting with ring A and continuing with rings B, C and D in order releasing four molecules of ammonia and resulting in a linear protein-bound hexapyrrole (Warren & Scott, 1990, Battersby *et al.*, 1979, Jordan & Seehra, 1979). Finally, the tetrapyrrole is released by the cleavage of the product-cofactor bond, regenerating the enzyme. Catalysis is achieved using a conserved aspartate residue that carries out the deamination of PBG (Woodcock & Jordan, 1994). Structural studies revealed that this monomeric enzyme utilises a conserved cysteine residue in the active site to retain the cofactor through a thioether bond (Louie *et al.*, 1992).

In the final step of uro'gen III synthesis ring D is inverted, before closure of the ring system by linkage of rings A and D. Uro'gen III synthase (UROS or HemD) carries out these reactions via the formation of a spirocyclic pyrrolenine intermediate (Layer *et al.*, 2004). Use of inhibitory spirocyclic intermediates and ^{13}C NMR as well as discovery of an azafulvene

reaction intermediate has contributed to the current mechanistic understanding of these reactions (Shoolingin-Jordan, 1995, Silva & Ramos, 2008, Spivey *et al.*, 1996).

Completion of uro'gen III synthesis is initiated with the rearrangement of ring A producing a positive charge on the pyrrole nitrogen and removal of the ring A hydroxyl group as water. This represents the formation of the first azafulvene intermediate. After bond rearrangement of ring D, the ring A azafulvene then attacks the substituted α position of ring D to form the spirocyclic pyrrolenine intermediate. Bond breakage between rings C and D produces another azafulvene intermediate on ring C. After the orientation of ring D moves the free α position in the vicinity of the ring C azafulvene they react forming a closed ring system. Subsequent deprotonation and bond rearrangements yield the final reaction product uro'gen III. Formation of uro'gen III represents the final branch point in the biosynthesis of all cyclic tetrapyrroles.

1.3.4 Completion of haem synthesis via the classical pathway

The classical pathway of haem synthesis has been well characterised for many years and proceeds in a four step process from uro'gen III to haem through the modification of side chains, oxidation of the ring system and finally iron insertion. Crystal structures of enzymes involved in each stage of this process have been solved from various sources. Interestingly, the first two steps in this pathway are shared with the chlorophyll biosynthesis pathway which then diverges at the point of metal insertion.

Uro'gen III is fated down the haem/chlorophyll pathway by the sequential decarboxylation of the four acetate side chains at carbons C2, C7, C12 and C18 (Dailey, 2002). These reactions are catalysed by the enzyme uro'gen III decarboxylase (UROD or HemE) which acts in a clockwise direction starting at ring D and successively decarboxylating to form the final product coproporphyrinogen III (copro'gen III) with the concomitant release of four molecules of CO_2 (Luo & Lim, 1993). The decarboxylations occur in a random order at high concentrations of uro'gen III. UROD is capable of using uro'gen I and uro'gen III as substrates as well as all of the semi-decarboxylated intermediates between uro'gen III and copro'gen III (Smith & Francis, 1979). Interestingly, unlike many other decarboxylases UROD does not require any cofactor or metal ion to catalyse these reactions.

The crystal structure of human HemE showed that it forms a homodimer with active sites from each monomer facing each other. Use of a single chain HemE along with the crystal

structure of HemE containing copro'gen III lead to the current understanding of the mechanism where-by after each decarboxylation the tetrapyrrole re-orientates 90° in the active site ready for the next decarboxylation, as opposed to the use of two active sites or the release and flipping of the tetrapyrrole by 180° (Phillips *et al.*, 2009, Phillips *et al.*, 2003). The exact reaction mechanism of HemE has not been determined; however, a reaction mechanism has been proposed involving two conserved residues, Arg37 and Asp86 (Lewis & Wolfenden, 2008). Initially, Arg37 is expected to form an ion pair with the scissile acetate carboxyl group, allowing that portion of the substrate to enter the nonpolar cavity within the active site where the decarboxylation occurs. Subsequently, Asp86 protonates the pyrrole ring at the α-position adjacent to the position of the acetate side chain as CO_2 is removed, producing a carbanionic intermediate. The product is then formed by proton abstraction from Arg37. Each decarboxylation reaction is suspected to proceed using an identical reaction mechanism after reorientation of the decarboxylated intermediates by 90º. It has been estimated that if the decarboxylation of uro'gen III was to occur spontaneously the half-life would be in excess of two billion years, which has led to HemE being referred to as a 'benchmark for the catalytic proficiency of enzymes' (Lewis & Wolfenden, 2008).

Conversion of copro'gen III to protoporphyrinogen IX (proto'gen IX) requires the oxidative decarboxylation of the two propionate side chains on rings A and B to vinyl groups. These reactions are catalysed by one of two structurally and mechanistically distinct enzymes, the oxygen-dependent copro'gen oxidase (CPO) HemF and oxygen-independent CPO HemN, which is often called copro'gen dehydrogenase (CPDH) to avoid confusion (Layer *et al.*, 2010). HemF is found in eukaryotes and only a few bacteria, which normally contain HemN (Cavallaro *et al.*, 2008).

HemF is a dimeric metal-independent membrane associated enzyme which utilises molecular oxygen as a terminal electron acceptor during catalysis (Macieira *et al.*, 2003, Medlock & Dailey, 1996, Breckau *et al.*, 2003). The exact mechanism of action is yet to be determined but it has been demonstrated that the propionate group on ring A is decarboxylated before that of ring B forming a monovinyl intermediate harderoporphyrinogen (Elder *et al.*, 1978). Mutagenesis of the human CPOs has suggested the involvement of two conserved arginine residues and a conserved aspartate residue in substrate binding and catalysis (Stephenson *et al.*, 2007). Structures of *Saccharomyces cerevisiae* and human HemF proteins have been solved and confirmed the presence of the dimeric structure (Phillips *et al.*, 2004, Lee *et al.*, 2005), but the lack of a substrate bound complex has restricted the understanding of the reaction mechanism of HemF. However, putative mechanisms have been proposed (Silva & Ramos, 2008, Lash, 2005).

Unlike HemF, HemN belongs to the radical SAM family of enzymes (Sofia *et al.*, 2001). Radical SAM enzymes are known to catalyse a wide range of reactions using SAM and a reductant to activate and functionalise typically unreactive C-H bonds (Atta *et al.*, 2010, Booker, 2009). This is achieved through the use of a [4Fe-4S] centre which is coordinated by three cysteines usually found in the conserved motif CX_3CX_2C (Layer *et al.*, 2004). Each cysteine coordinates one iron, with the final iron coordinated to the methionine moiety of SAM. The reaction cycle is initiated by the reduction of the iron sulphur cluster from the resting $[4Fe\text{-}4S]^{2+}$ to the active $[4Fe\text{-}4S]^{+}$ state. An electron is transferred onto SAM causing homolytic cleavage and the formation of a $[4Fe\text{-}4S]^{2+}$-methionine intermediate and a 5'-deoxyadenosyl radical which is capable of abstracting a hydrogen atom from the substrate or a glycine on a partner protein forming a substrate or protein localised radical. After this the following steps are unique to each radical SAM enzyme but generally result in the irreversible consumption of SAM which therefore acts as a co-substrate.

The initial electron donor and final electron acceptor for the oxidative decarboxylation reactions of the northern propionate groups of copro'gen III are yet to be discovered, but the reaction mechanism has been determined using isotopic labelling (Layer *et al.*, 2006). The 5'-deoxyadenosyl radical abstracts the pro-S-hydrogen at the β-position of the propionate side chain of the substrate, which forms an allylic coproporphyrinogenyl radical and 5'-deoxyadenosine. After the formation of the substrate radical the first of the two decarboxylations can occur using an unknown electron acceptor. The crystal structure of HemN was the first radical SAM enzyme to have its crystal structure solved and was shown to be a monomeric enzyme which contained two SAM binding sites (Layer *et al.*, 2003). It has also been shown that turnover of copro'gen III to proto'gen IX requires two molecules of SAM for completion, *i.e.* one for the oxidative decarboxylation of each propionate group (Layer *et al.*, 2005).

Some proteins from bacteria, plants and animals have been annotated as HemN due to high sequence similarity although they lack CPO/CPDH function and have since been annotated as HemW (Abicht *et al.*, 2012). Indeed, HemW from *Lactococcus lactis* has ~50% similarity to *E. coli* HemN but is not active (Abicht *et al.*, 2012). However, it does bind haem and is suspected to play a role in haem transport.

Prior to iron insertion the ring system is oxidised to form a fully conjugated, aromatic, and planar macrocycle. Proto'gen IX is converted to protoprophyrin IX by the action of one of three proto'gen IX oxidases, HemY, HemG or HemJ, which remove six electrons and six

protons (two from the imino groups and one from each of the four bridging carbon atoms) from the substrate (Dailey & Dailey, 1996, Sasarman *et al.*, 1993, Boynton *et al.*, 2011).

HemY is an oxygen- and flavo-dinucleotide (FAD) dependent enzyme found in most organisms as a membrane bound protein [inner mitochondrial membrane in eukaryotes, and cytoplasmic membrane in prokaryotes; (Dailey & Dailey, 1996, Dailey & Dailey, 1997)], although a soluble form has been described in *Bacillus subtilis* (Corrigall *et al.*, 1998). HemY is present in monomeric or dimeric forms depending on the organism and has been crystallised from various sources revealing a non-covalently bound FAD cofactor in each of them (Koch *et al.*, 2004, Corradi *et al.*, 2006, Heinemann *et al.*, 2007). Little is known about the reaction mechanism of HemY other than the involvement of FAD, formation of H_2O_2 and oxygen being the terminal electron acceptor (Layer *et al.*, 2010).

The second proto'gen IX oxidase, HemG, is unique to *γ*-proteobacteria and allows synthesis of haem in both oxic and anoxic environments (Boynton *et al.*, 2009). HemG is able to utilise a range of compounds as terminal electron acceptors including oxygen, fumarate or nitrate and is coupled to electron transport chains leading to the generation of ATP instead of H_2O_2 as a by-product (Jacobs & Jacobs, 1976, Mobius *et al.*, 2010).

The third proto'gen oxidase, HemJ, has only recently been identified and very little is currently known about it (Kato *et al.*, 2010). However, it has been shown through comparative genomics that bacteria that contain the classical haem synthesis pathway but lack both *hemY* and *hemG* tend to contain a *hemJ* gene and that deletion of *hemJ* in *Actinobacter baylyi* causes haem auxotrophy (Boynton *et al.*, 2011).

The conclusion of classical haem synthesis is the chelation of iron into the macrocycle of protoprophyrin IX to form protohaem (haem *b*). The ferrochelatase HemH catalyses iron chelation independent of additional cofactors or metal ions. Some HemH proteins have been found to contain [2Fe-2S] centres, the role of which is yet to be determined (Wu *et al.*, 2001). Ferrochelatase enzymes can be found as monomeric or dimeric enzymes, both membrane-associated and cytoplasmic depending on the organism (Dailey *et al.*, 2000). Several crystal structures of HemH have been solved including those from human, yeast and *B. subtilis* (Wu *et al.*, 2001, Al-Karadaghi *et al.*, 1997, Karlberg *et al.*, 2002). The exact mechanism for iron chelation is unknown but it has been shown that HemH can accept a wide range of porphyrin substrates with differing substituents on rings A and B; however, the propionate groups on rings C and D appear essential for activity (Dailey *et al.*, 1989). HemH shares structural homology to other chelatase enzymes including those involved in

sirohaem synthesis and cobaltochelatases from the anaerobic cobalamin synthesis pathway, suggesting that they share a common ancestral source (Romao *et al.*, 2011).

An additional protein (HemQ) has recently been discovered that is essential for haem synthesis in Gram-positive bacteria (Dailey *et al.*, 2010). The study showed that HemQ is required along with HemY and HemH to form haem both *in vitro* and *in vivo*. Exactly what HemQ does is currently unclear, although it has been proposed to be involved in removal of H_2O_2 generated by HemY.

1.3.5 Biosynthesis of sirohaem

To introduce the alternative haem biosynthesis pathway the route to the precursor sirohaem must first be discussed. Sirohaem follows a comparatively simple biosynthetic pathway from uro'gen III containing three stages: (1) addition of SAM derived methyl groups to carbons C2 and C7 of uro'gen III to form precorrin-2, (2) oxidation of precorrin-2 to sirohydrochlorin, (3) chelation of iron into the macrocycle (Fig. 1.6). A variety of enzymes are employed by different organisms to catalyse these reactions. For example, *Bacillus megaterium* contains individual enzymes for each stage of the process, a SAM-dependent uro'gen III methyltransferase (SUMT) SirA, a precorrin-2 dehydrogenase SirC and a ferrochelatase SirB (Raux *et al.*, 2003). *S. cerevisiae*, however, contains two enzymes, a SUMT (Met1p) and a bi-functional precorrin-2 dehydrogenase and ferrochelatase Met8p (Raux *et al.*, 1999a). *E. coli* contains a single multifunctional enzyme (CysG) that is capable of converting uro'gen III to sirohaem alone (Warren *et al.*, 1994). CysG contains two domains, $CysG^A$ which is homologous to SUMT enzymes SirA and Met1p, and $CysG^B$ which is homologous to the precorrin-2 dehydrogenases and ferrochelatases.

As with haem and chlorophyll, the sirohaem pathway shares intermediates with other tetrapyrrole biosynthesis pathways. Precorrin-2 is converted to precorrin-3 by CobI, to initiate aerobic cobalamin biosynthesis (Warren *et al.*, 2002), as well as being an intermediate in the as yet uncharacterised biosynthesis of coenzyme F_{430} (Warren *et al.*, 2009). The anaerobic pathway of cobalamin synthesis diverges at the point of sirohydrochlorin with the chelation of cobalt into the macrocycle by the enzymes CbiK or CbiX (Raux *et al.*, 2003). A summary of the branch points of tetrapyrrole biosynthesis is displayed in Figure 1.7.

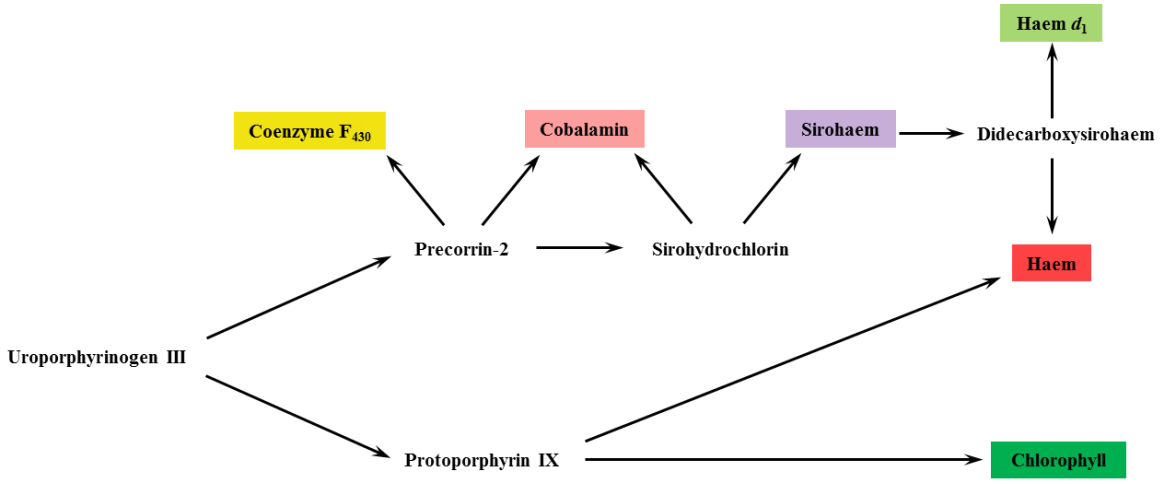


Figure 1.7. An overview of branch points in tetrapyrrole biosynthesis. The known branch points for the synthesis of modified tetrapyrroles starting from the common precursor uro'gen III. Functional tetrapyrroles are highlighted with their natural colouration.

The synthesis of sirohaem in many of the bacteria and archaea that use the alternative haem biosynthesis pathway is yet to be fully characterised. Many archaeal genomes appear to have a SUMT, although annotations are varied between *cobA*, *cysG-1*, *cysG-2*, *cysG*, *uroM*, or *hemX*. Similarly, the gene encoding precorrin-2 dehydrogenase is referred to as *sirC*, *hemX*, *cysG*, or *cysG1* (Storbeck *et al.*, 2010). Archaeal genomes also lack a dedicated sirohaem ferrochelatase, although it has been demonstrated that the short cobaltochelatase $CbiX_S$ from *Methanosarcina barkeri* and *Methanobacter thermoautotrophicum* is capable of chelating iron into sirohydrochlorin, although at a lower rate to that of cobalt insertion (Brindley *et al.*, 2003). It is therefore possible that this protein has a dual role in cobalt and iron chelation. The sulphate reducing bacterium *Desulfovibrio vulgaris* possesses a bi-functional uro'gen III synthase/methyltransferase (HemD-CobA) and a precorrin-2 dehydrogenase and is therefore capable of synthesising sirohydrochlorin (Lobo *et al.*, 2009). A unique sirohaem ferrochelatase is absent in the *D. vulgaris* genome. However, *D. vulgaris* does possess two putative CbiK cobaltochelatases ($CbiK^C$ and $CbiK^P$), which have been shown to chelate both cobalt and iron into sirohydrochlorin (Lobo *et al.*, 2008). Genomes of other δ-proteobacteria also contain genes encoding these proteins including members of the groups *Desulfovibrio*, *Desulfobulbus*, *Desulfatibacillum*, and *Desulfobacterium* (Lobo *et al.*, 2012).

1.3.6 Biosynthesis of haem d_1

Again, a brief introduction to the biosynthesis of haem d_1 is required in order for a fuller understanding of the alternative haem biosynthesis pathway. As mentioned previously haem d_1 is only required as a cofactor for one enzyme, cytochrome cd_1 nitrite reductase, which is encoded by the gene *nirS* (Bali *et al.*, 2014). The essential genes for haem d_1 synthesis belong to the *nir* family and can be found in the region downstream of *nirS* and include the genes *nirFDLGHJE* (Suzuki *et al.*, 2006, Palmedo *et al.*, 1995). NirE represents the most well characterised Nir protein and shares high sequence homology with known SUMT enzymes including the C-terminal region of CysG, CobA (methyltransferase of the cobalamin pathway), SirA, and Met1p (Warren *et al.*, 2009).

NirD, NirL, NirG and NirH are all cytoplasmic proteins which share a high degree of sequence homology (Bali *et al.*, 2014). In some cases *nirD* and *nirL* are found fused resulting in the production of a single protein. Despite their homology all four enzymes are not functionally redundant, and have been shown to be essential for nitrite reductase function (Palmedo *et al.*, 1995, Kawasaki *et al.*, 1997). The roles of NirDL-G-H were elucidated at the same time as the alternative haem biosynthesis pathway. NirDL-G-H from *Paracoccus pantotrophus* were found to catalyse the double decarboxylation of sirohaem to didecarboxysirohaem (Bali *et al.*, 2011). NirDL was found to produce a monodecarboxylated intermediate; however no structural analysis of this product was undertaken.

The roles of NirJ and NirF are currently unknown. NirJ is a radical SAM protein containing two cysteine motifs for iron sulphur clusters. It has been postulated that NirJ catalyses the elimination of the propionate side chains of didecarboxysirohaem at carbons C3 and C8 by activating the alkyl leaving groups using radical chemistry. However, NirJ may also catalyse the two electron oxidation of the propionate side chain at carbon C17 yielding an acrylate functionality. NirF has been shown to form a complex with NirS and another protein NirN, and also binds haem d_1 *in vitro* (Nicke *et al.*, 2013). This lead to the hypothesis that NirF is involved in the terminal stages of haem d_1 synthesis and formation of holo-NirS, although it is currently unknown which reaction it would carry out.

1.3.7 The alternative haem biosynthesis pathway

D. vulgaris was the first organism to be found to synthesise haem via an alternative route some 20 years ago, although the pathway was not unravelled until many years later. The discovery of the alternative pathway was by serendipity when labelled methionine was unexpectedly incorporated into the haem prosthetic group of cytochrome c_3 (Akutsu *et al.*, 1993). To measure redox potentials of cytochrome c_3 selective deuteration of methionine residues was performed by culturing *D. vulgaris* in minimal media supplemented with L-methionine-methyl-d_3. During NMR assignment of the haem groups in the labelled cytochrome c_3 it was noted that two methyl groups had a diminished signal in comparison to the others, signifying that they had acquired labelled methyl groups derived from deuterated methionine. All methyl groups of haem synthesised via the classical pathway are derived from ALA, suggesting an alternative haem biosynthesis pathway exists in this organism utilising methionine for the formation of some methyl groups. It was discovered that the diminished signal belonged to methyl groups at carbons C2 and C7 and it was proposed that the pathway proceeds via precorrin-2, and possibly sirohydrochlorin, where these methyl groups are derived from SAM (Scott, 1990). The same process has also been confirmed for the biosynthesis of haem d_1 (Yapbondoc *et al.*, 1990).

Extraction of labelled tetrapyrroles from cultures of *D. vulgaris* supplemented with deuterated L-methionine-methyl-d_3 further confirmed the presence of an alternative pathway (Ishida *et al.*, 1998). Several potential pathway intermediates were isolated, including sirohydrochlorin, coproporphyrin III, and protoporphyrin IX as well as a previously unknown hexacarboxylic acid termed 12,18-didecarboxysirohydrochlorin. From these findings the pathway was proposed to proceed from uro'gen III to precorrin-2 and then re-join the classical pathway after the formation of copro'gen III. Following this pathway would require 6 steps from uro'gen III as opposed to the 4 present in the classical pathway and would proceed as follows: (1) methylation of carbons C2 and C7 of uro'gen III to form precorrin-2, (2) decarboxylation of carbons C12 and C18, to form 12,18-didecarboxyprecorrin-2, (3) elimination of acetate groups from carbons C2 and C7 to form copro'gen III, (4) decarboxylation of propionate side chains on carbons C3 and C8 to vinyl groups forming proto'gen IX, (5) oxidation of ring system to form protoporphyrin IX, (6) insertion of ferrous iron to form haem. As with the classical pathway insertion of iron was proposed to be the final stage in the process as metal-chelated intermediates may be cytotoxic.

The search for the enzymes catalysing the alternative haem biosynthesis pathway was greatly aided by the growing number of sequenced bacterial and archaeal genomes, which were also able to reveal that the alternative pathway is not restricted to sulphate-reducing bacteria as originally thought (Panek & O'Brian, 2002, Cavallaro *et al.*, 2008, Storbeck *et al.*, 2010). Indeed, it was shown using similar labelling studies as in *D. vulgaris*, that haem in *M. barkeri* also contains two SAM-derived methyl groups, suggesting the use of the same pathway in both organisms (Buchenau *et al.*, 2006). Through genome wide searches for homologs of the classical haem synthesis pathway, it was discovered that some bacteria and archaea only contain the genes required for early haem biosynthesis up to uro'gen III (*i.e. hemA, hemL, hemB, hemC,* and *hemD*; Fig. 1.5) and lack the necessary genes to convert uro'gen III to haem via the classical pathway [*hemE, hemF*/*N*, *hemY*/*G*/*J*, and *hemH*; Fig. 1.6; (Cavallaro *et al.*, 2008, Panek & O'Brian, 2002, Storbeck *et al.*, 2010)]. Therefore the theory that the alternative pathway re-entered the classical pathway with copro'gen III could not be upheld and the alternative pathway was proposed to have a unique set of enzymes that catalyse the conversion of uro'gen III to haem.

D. vulgaris was among the bacteria found to only contain genes required for uro'gen III synthesis. Further genomic analysis revealed it also contained homologs of *nirD* and *nirJ* (Lobo *et al.*, 2009), which had been shown to be involved with haem d_1 synthesis (Kawasaki *et al.*, 1997), a somewhat surprising discovery given that *D. vulgaris* does not contain cytochrome cd_1 nitrite reductase. Analysis of archaeal genomes revealed clusters containing early *hem* genes with genes encoding SUMT and precorrin-2 dehydrogenase, further supporting the use of sirohydrochlorin as an intermediate in alternative haem synthesis (Storbeck *et al.*, 2010). Similarly, many of these gene clusters also contained homologs of *nirD, nirH* and *nirJ*, despite the fact only 3 of 27 of the species containing these genes also contained a potential cytochrome cd_1 nitrite reductase. It was therefore suspected that these genes may be involved in alternative haem biosynthesis.

Given their different role these genes were re-annotated as the *ahb* (**a**lternative **h**aem **b**iosynthesis) genes and will be referred to as such throughout. The genes *ahbA* and *ahbB* are homologous to *nirD* and *nirH*, respectively, and *ahbC* and *ahbD* are homologous to *nirJ*. Following these discoveries a new pathway was proposed based on previously suggested roles for the Nir proteins (Kawasaki *et al.*, 1997, Storbeck *et al.*, 2010): (1) decarboxylation of the acetate side chains at carbons C12 and C18 of sirohydrochlorin by AhbA and AhbB, (2) removal of the acetate side chains at carbons C2 and C7 by AhbC, (3) the oxidative decarboxylation of propionate side chains on carbons C3 and C8 to vinyl groups by AhbD, (4) chelation of iron into the macrocycle by an unidentified ferrochelatase.

The lack of a dedicated ferrochelatase was somewhat confusing until the final piece of puzzle was discovered some 20 years after the initial revelation of the alternative pathway. The iron chelated form of sirohydrochlorin, sirohaem, was found to be the true initial substrate of the alternative haem biosynthesis pathway (Bali *et al.*, 2011). Once regarded as the end product of a biosynthesis pathway, the use of sirohaem was unexpected, but equally so was the link between the alternative haem and haem d_1 pathways. Indeed, both NirDL-G-H and AhbA/B were found to convert sirohaem to the intermediate didecarboxysirohaem, at which point the two pathways diverge. Total conversion of sirohaem to haem was achieved using cell lysates of *D. vulgaris* which resulted in the isolation of haem as well as the intermediates monodecarboxysirohaem, didecarboxysirohaem and a monovinyl iron coproporphyrin.

A dedicated pathway for the formation of sirohaem for the alternative pathway is unlikely to exist given that many of the organisms that utilise the alternative pathway also have a requirement for sirohaem as a cofactor in nitrite and sulphite reductases (Susanti & Mukhopadhyay, 2012), or are able to produce cobalamin. Indeed, in a previous study of archaeal genomes only four organisms (*Aeropyrum pernix* K1, *Ignicoccus hospitalis* KIN4/I, *Methanosphaerula palustris* E1-9c, and *Hyperthermus butylicus* DSM 5456) were shown to contain *ahbA, B, C,* and *D* homologues but did not have evidence of a sirohaem-containing sulphite or nitrite reductase or lacked cobalamin biosynthesis genes (Storbeck *et al.*, 2010). The use of iron chelated sirohaem may explain why a ferrochelatase dedicated to the alternative haem biosynthesis pathway was not discovered. The discovery that sirohaem is an intermediate as well as a functional tetrapyrrole has raised questions about the potentially complex regulatory pathways required by organisms that utilise sirohaem and/or cobalamin/coenzyme F_{430} and also contain the alternative haem or haem d_1 synthesis pathways. Use of precorrin-2, sirohydrochlorin and sirohaem must be tightly regulated to allow flux along all of the pathways.

1.3.8 Completion of haem synthesis via the alternative pathway

The alternative haem biosynthesis pathway is now known to proceed in three stages and can be assigned to specific enzymes although mechanistic studies are yet to be completed. (1) Acetate side chains at carbons C12 and C18 of sirohaem are decarboxylated by AhbA and AhbB to form didecarboxysirohaem, although the specificity and order of decarboxylation is unknown. (2) AhbC catalyses the oxidative loss of the northern acetic acid side chains at

carbons C2 and C7 in a SAM dependent manner (Bali *et al.*, 2011). AhbC is a member of the radical SAM family mentioned previously and contains two motifs for iron sulphur clusters although it is unknown whether both are required for functionality. (3) The final step requires the oxidative decarboxylation of the propionate groups at carbons C3 and C8 of iron coporporphyrin to vinyl groups producing haem, which is catalysed by the enzyme AhbD. This reaction is suspected to occur in a homologous fashion to the conversion of coproprophyrinogen III to protoporphyrinogen IX seen in the classical pathway, which is catalysed by HemN. Like HemN, AhbD is also a radical SAM enzyme and catalyses the conversion of iron coproporphyrin III to haem in a SAM dependent manner (Bali *et al.*, 2011).

1.4 Organisms of interest

This study focuses on the characterisation enzymes derived from two sulphate reducing bacteria, *Desulfovibrio desulfuricans* and *D. vulgaris*, and the methanogenic archaeon *M. barkeri*.

1.4.1 Sulphate reducing bacteria – *Desulfovibrio* species

Desulfovibrio belongs to the *δ*-proteobacteria and is one of the best studied genera of sulphate reducing bacteria (SRB). SRB are a large, heterogeneous group of prokaryotes capable of dissimilatory sulphate reduction, *i.e.* they obtain energy by coupling the anaerobic oxidation of organic compounds or molecular hydrogen to the reduction of sulphate to hydrogen sulphide (Muyzer & Stams, 2008). It is likely that some of the oldest forms of life on Earth belong to this group of microorganisms, given that the process of dissimilatory sulphate reduction is a very ancient process (Shen *et al.*, 2001). SRB are found in a variety of habitats, usually in sulphate-rich anoxic environments including soil, marine and fresh waters, and the gut of animals including humans. Despite their name sulphate reducers are also able to use other terminal electron acceptors for anaerobic respiration, such as sulphur, fumarate, dimethylsulphoxide, Mn (IV), and Fe (III) (Thauer *et al.*, 2007).

As well as their natural habitat SRB also reside in industrial niches including water systems and oil pipelines. The most prevalent in oil and gas industries are strains belonging to the *Desulfovibrio* genus (Voordouw, 2008). High concentrations of hydrogen sulphide are produced by *Desulfovibrio* reducing the quality of oil in a process known as souring. The sulphide reacts with Fe^{2+} forming corrosive iron sulphides which can damage pipelines and

industrial equipment. Control of these species is therefore of industrial interest to increase oil quality and reduce costs associated with metal corrosion. In contrast to the negative influence in the oil and gas industry bacteria from the *Desulfovibrio* genus also play a major role in bioremediation of toxic metals (Hockin & Gadd, 2007). SRB are able to reduce a range of toxic metals including uranium(VI), copper(II), chromium(VI) and manganese(II) (Cabrera *et al.*, 2006). Reducing these heavy metals to their insoluble form greatly facilitates their removal from contaminated water. Therefore the further understanding of sulphate reducers and their metabolic pathways can lead to their control in industrial settings or exploitation in the case of bioremediation.

1.4.2 Methanogenic archaea – *Methanosarcina barkeri*

M. barkeri belongs to the methanogenic archaea, a diverse group of strictly anaerobic Euryarchaeota. Methanogenic archaea are characterised by their ability to form methane as a metabolic end product from a restricted range of substrates that include CO_2 and H_2, formate, methanol, methylamines and/or acetate (Thauer *et al.*, 2008). These methanogens inhabit a range of anoxic environments including freshwater sediments, swamps, paddy fields, land-fills and the intestinal tracts of ruminants and termites and contribute to the global carbon cycle through the production of an estimated 1 billion tons of methane annually (Thauer, 1998).

The *Methanosarcina* species have been viewed as key organisms for the production of methane during anaerobic digestion (De Vrieze *et al.*, 2012) due to their high growth rates and relative robustness in comparison to other methanogens. The genome of *M. barkeri* has been published (Maeder *et al.*, 2006) which has aided the use of *M. barkeri* as a model organism for methanogenesis.

1.5 Aims and objectives

The research reported in this thesis aimed to expand upon the knowledge of the enzymes involved in the alternative haem biosynthesis pathway. Of specific interest was the first step in the pathway that is catalysed by the two proteins AhbA and AhbB. A wide range of molecular and biophysical techniques have been employed including molecular biology, recombinant DNA technology, enzymology, crystallography and spectroscopy in order to achieve the following goals:

- investigate biochemical characteristics of AhbA and AhbB from a range of sources;
- ascertain the activity of individual AhbA and AhbB enzymes and AhbA/B complexes using *in vitro* assays;
- confirm the sites of decarboxylation of sirohaem;
- determine kinetic properties of AhbA/B;
- perform structural analysis of apo- and product bound complexes of AhbA/B using X-ray crystallography;
- assess product binding capabilities of AhbA/B;
- use site-directed mutagenesis to analyse influence of specific amino acids on substrate binding and catalysis;
- analyse AhbA/B properties and interactions by generation of chimeric complexes.

The results of this investigation have enhanced the overall understanding of the early stages of the alternative haem biosynthesis pathway by providing structural and functional information about the AhbA and AhbB enzymes and the cross-species variation within this family of enzymes

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals

Most chemicals, DEAE-sephacel and antibiotics were purchased from Sigma-Aldrich Ltd. Other materials were purchased from manufacturers as follows: IPTG and ampicillin from Melford Laboratories Ltd; chelating fast flow sepharose, disposable and empty PD10 columns from GE Healthcare; MiniElute™ Gel Extraction Kit, QIAprep® Spin Miniprep Kit, and QIAquick® Gel Extraction Kit from Qiagen; restriction enzymes were purchased from Promega and from New England Biolabs Inc; cloning vectors from Novagen; tryptone, yeast and bacterial agar from Oxoid Ltd; QuickChange® II Site-Directed Mutagenesis Kit from Agilent Technologies; Roche FastStart High Fidelity PCR System from Roche Diagnostics GmbH. Crystallisation screens were purchased from Molecular Dimensions Ltd. and Hampton Research.

2.1.2 Bacterial strains

Bacterial strains were purchased from Novagen or Promega.

Table 2.1. Bacterial strains.

***E. coli* strain**	**Genotype and/or phenotype**	**Reference**
JM109	e14$^-$(McrA$^-$) *recA1 endA1 gyrA96 thi$^-$1 hsdR17* (r_K^- m_K^+) *supE44 relA1* Δ(*lac-proAB*) [F´ *traD36 proAB lacI^{q}ZΔM15*]	Promega
BL21 (DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3)	Novagen
BL21 (DE3) pLysS	F$^-$ *ompT hsdSB* (r_B^-, m_B^-) *gal dcm* (DE3) pLysS (Cm^R)	Novagen
NovaF$^-$	F$^-$ *endA1 hsdR17* (r_{K12}^- m_{K12}^+) *supE44 thi$^-$1 recA1 gyrA96 relA1 lac*	Novagen
KRX	[F$'$, *traD36*, Δ*ompP, proA$^+$B$^+$,lacI^q*, Δ(*lac*Z)M15] Δ*ompT, endA1, recA1, gyrA96* (*Nal^r*), *thi$^-$1, hsdR17* (r_K^-, m_K^+), e14$^-$ (McrA$^-$), *relA1, supE44,* Δ(*lac-proAB*), Δ(*rhaBAD*)::T7 RNA polymerase	Promega
XL1-Blue	*recA1 endA1 gyrA96 thi$^-$1 hsdR17 supE44 relA1 lac* [F*' proAB lacI^{q}ZΔM15* Tn*10* (Tet^r)]	Agilent
DH5-α	F$^-$ ϕ80 *lacZΔM15* Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (r_K-,m_K+) *phoA supE44 λ$^-$ thi$^-$1 gyrA96 relA1*	Invitrogen

2.1.3 Plasmids

Table 2.2. Plasmids. Mb: *M. barkeri*, Bm: *B. megaterium*, Mth: *Methanothermobacter thermautotrophicus*, Dv: *D. vulgaris*, Dd: *D. desulfuricans*.

Name	Plasmid	Description	Reference or source
pET3a	pET3a	Overproduction vector with T7 promoter, Amp^R	Novagen
pET14b	pET14b	Overproduction N-terminal His_6tag fusion protein vector with T7 promoter, Amp^R	Novagen
pETcoco-2	pETcoco-2	Protein overproduction vector, regulation of the copy number with glucose and arabinose: dula origins of replication control: single-copy (*oriS*) and medium copy (*oriV*) states, Amp^R)	Novagen
pED_ABCDC	pETcoco-2-*cobA-hemBCD-sirC*	pETcoco-2 vector carrying *Mb cobA*, *Bm hemC-D*, and *Mth sirC* and *hemB*, for sirohydrochlorin production	Dr. Evelyne Deery
pSL201	pET23b-$ahbA^{Dv}$	*Dv ahbA* cloned into pET23b using NdeI and XhoI	Dr. Susana Lobo
pSL202	pET23b-$ahbB^{Dv}$	*Dv ahbB* cloned into pET23b using NdeI and XhoI	Dr. Susana Lobo
pSL203	pET23b-$ahbA^{Dd}$	*Dd ahbA* cloned into pET23b using NdeI and XhoI	Dr. Susana Lobo
pSL204	pET23b-$ahbB^{Dd}$	*Dd ahbB* cloned into pET23b using NdeI and XhoI	Dr. Susana Lobo
pDP001	pET3a-*cysG*	*E.coli cysG* cloned into NdeI and SpeI sites of pET3a	This study
pDP002	pET3a-$ahbA^{Dv}$	*Dv ahbA* cloned into NdeI and SpeI sites of pET3a	This study
pDP003	pET3a-$ahbB^{Dv}$	*Dv ahbB* cloned into NdeI and SpeI sites of	This study

		pET3a	
pDP011	pET14b-*ahbA*^{Dv}	*Dv ahbA* cloned into NdeI and SpeI sites of pET14b	This study
pDP012	pET14b-*ahbB*^{Dv}	*Dv ahbB* cloned into NdeI and SpeI sites of pET14b	This study
pDP017	pET3a-*cysG-ahbA*$^{Dv\text{-}His}$	Link and Lock of *E.coli cysG* and *Dv ahbA* with His_6-tag	This study
pDP018	pET3a-*cysG-ahbA*$^{Dv\text{-}His}$*B*^{Dv}	Link and Lock of *E.coli cysG*, *Dv ahbA* with His_6-tag and *Dv ahbB*	This study
pDP019	pET3a-*cysG-ahbA*^{Dv}*B*$^{Dv\text{-}His}$	Link and Lock of *E.coli cysG*, *Dv ahbA* and *ahbB* with His_6-tag	This study
pDP026	pET3a-*ahbA*^{Dd}	*Dd ahbA* cloned into NdeI and SpeI sites of pET3a	This study
pDP027	pET3a-*ahbB*^{Dd}	*Dd ahbB* cloned into NdeI and SpeI sites of pET3a	This study
pDP028	pET14b-*ahbA*^{Dd}	*Dd ahbA* cloned into NdeI and SpeI sites of pET14b	This study
pDP029	pET14b-*ahbB*^{Dd}	*Dd ahbB* cloned into NdeI and SpeI sites of pET14b	This study
pDP030	pET3a-*ahbA*^{Dv}*B*$^{Dv\text{-}His}$	Link and Lock of *Dv ahbA* and *Dv ahbB* with His-tag	This study
pDP031	pET3a-*ahbA*^{Dd}*B*$^{Dd\text{-}His}$	Link and Lock of *Dd ahbA* and *Dd ahbB* with His_6-tag	This study
pDP032	pET3a-*cysG-ahbA*^{Dd}*B*$^{Dd\text{-}His}$	Link and Lock of *E. coli cysG*, *Dd ahbA* and *Dd ahbB* with His_6-tag	This study
pDP033	pET3a-*ahbA*^{Mb}	*Mb ahbA* cloned into NdeI and SpeI sites of pET3a	This study
pDP034	pET3a-*ahbB*^{Mb}	*Mb ahbB* cloned into NdeI and SpeI sites of pET3a	This study
pDP035	pET14b-*ahbA*^{Mb}	*Mb ahbA* cloned into NdeI and SpeI sites of pET14b	This study
pDP036	pET14b-*ahbB*^{Mb}	*Mb ahbB* cloned into NdeI and SpeI sites of pET14b	This study
pDP037	pET3a-*ahbA*^{Mb}*B*^{Mb}	Link and Lock of *Mb ahbA* and *Mb ahbB*	This study

		with His_6-tag	
pDP038	pET3a-*cysG-ahbA*^{Mb}*B*$^{Mb\text{-}His}$	Link and Lock of *E. coli cysG, Mb ahbA* and *Mb ahbB* with His_6-tag	This study
pTH001	pET3a-*ahbA*^{Dd} R73A	*Dd ahbA* R73A cloned into NdeI and SpeI sites of pET3a	This study
pTH002	pET3a-*ahbA*^{Dd} R74A	*Dd ahbA* R74A cloned into NdeI and SpeI sites of pET3a	This study
pTH003	pET3a-*ahbA*^{Dd} H115A	*Dd ahbA* H115A cloned into NdeI and SpeI sites of pET3a	This study
pTH004	pET3a-*ahbA*^{Dd} Y117A	*Dd ahbA* Y117A cloned into NdeI and SpeI sites of pET3a	This study
pTH005	pET3a-*ahbA*^{Dd} R119A	*Dd ahbA* R119A cloned into NdeI and SpeI sites of pET3a	This study
pTH006	pET14b-*ahbB*$^{Dd\text{-}His}$ R55A	*Dd ahbB* R55A cloned into NdeI and SpeI sites of pET14b	This study
pTH007	pET14b-*ahbB*$^{Dd\text{-}His}$ R56A	*Dd ahbB* R56A cloned into NdeI and SpeI sites of pET14b	This study
pTH008	pET14b-*ahbB*$^{Dd\text{-}His}$ H63A	*Dd ahbB* H63A cloned into NdeI and SpeI sites of pET14b	This study
pTH009	pET14b-*ahbB*$^{Dd\text{-}His}$ H98A	*Dd ahbB* H98A cloned into NdeI and SpeI sites of pET14b	This study
pTH010	pET14b-*ahbB*$^{Dd\text{-}His}$ Y100A	*Dd ahbB* Y100A cloned into NdeI and SpeI sites of pET14b	This study
pTH011	pET14b-*ahbB*$^{Dd\text{-}His}$ R102A	*Dd ahbB* R102A cloned into NdeI and SpeI sites of pET14b	This study
pTH012	pET14b-*ahbB*$^{Dd\text{-}His}$ H118A	*Dd ahbB* H118A cloned into NdeI and SpeI sites of pET14b	This study
pTH013	pET3a-*ahbA*^{Dd} R73A *ahbB*$^{Dd\text{-}His}$	Link and Lock of *Dd ahbA* R73A and *Dd ahbB* with His_6-tag	This study
pTH014	pET3a-*ahbA*^{Dd} R74A *ahbB*$^{Dd\text{-}His}$	Link and Lock of *Dd ahbA* R74A and *Dd ahbB* with His_6-tag	This study
pTH015	pET3a-*ahbA*^{Dd} H115A *ahbB*$^{Dd\text{-}His}$	Link and Lock of *Dd ahbA* H115A and *Dd ahbB* with His_6-tag	This study
pTH016	pET3a-*ahbA*^{Dd} Y117A *ahbB*$^{Dd\text{-}His}$	Link and Lock of *Dd ahbA* Y117A and *Dd ahbB* with His_6-tag	This study
pTH017	pET3a-*ahbA*^{Dd}	Link and Lock of *Dd*	This study

	R119A *ahbB*$^{Dd\text{-His}}$	*ahbA* R119A and *Dd ahbB* with His_6-tag	
pTH018	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ R55A	Link and Lock of *Dd ahbA* and *Dd ahbB* R55A with His_6-tag	This study
pTH019	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\ \text{-His}}$ R56A	Link and Lock of *Dd ahbA* and *Dd ahbB* R56A with His_6-tag	This study
pTH020	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H63A	Link and Lock of *Dd ahbA* and *Dd ahbB* H63A with His_6-tag	This study
pTH021	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H98A	Link and Lock of *Dd ahbA* and *Dd ahbB* H98A with His_6-tag	This study
pTH022	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ Y100A	Link and Lock of *Dd ahbA* and *Dd ahbB* Y100A with His_6-tag	This study
pTH023	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ R102A	Link and Lock of *Dd ahbA* and *Dd ahbB* R102A with His_6-tag	This study
pTH024	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H118A	Link and Lock of *Dd ahbA* and *Dd ahbB* H118A with His_6-tag	This study
pTH025	pET3a-*ahbA*^{Dd} R119 *ahbB*$^{Dd\text{-His}}$ R102A	Link and Lock of *Dd ahbA* R119 and *Dd ahbB* R102A with His_6-tag	This study
pTH026	pET3a-*ahbA*^{Dd} R73A *ahbB*$^{Dd\text{-His}}$ *cysG*	Link and Lock of *Dd ahbA* R73A, *Dd ahbB* with His_6-tag and *E. coli cysG*	This study
pTH027	pET3a-*ahbA*^{Dd} R74A *ahbB*^{Dd} *cysG*	Link and Lock of *Dd ahbA* R74A, *Dd ahbB* with His_6-tag and *E. coli cysG*	This study
pTH028	pET3a-*ahbA*^{Dd} H115A *ahbB*$^{Dd\text{-His}}$ *cysG*	Link and Lock of *Dd ahbA* H115A, *Dd ahbB* with His_6-tag and *E. coli cysG*	This study
pTH029	pET3a-*ahbA*^{Dd} Y117A *ahbB*$^{Dd\text{-His}}$ *cysG*	Link and Lock of *Dd ahbA* Y117A, *Dd ahbB* with His_6-tag and *E. coli cysG*	This study
pTH030	pET3a-*ahbA*^{Dd} R119A *ahbB*$^{Dd\text{-His}}$ *cysG*	Link and Lock of *Dd ahbA* R119A, *Dd ahbB* with His_6-tag and *E. coli cysG*	This study
pTH031	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ R55A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* R55A with His_6-tag and *E. coli cysG*	This study
pTH032	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ R56A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* R56A with His_6-tag and *E.*	This study

		coli cysG	
pTH033	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H63A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* H63A with His_6-tag and *E. coli cysG*	This study
pTH034	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H98A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* H98A with His_6-tag and *E. coli cysG*	This study
pTH035	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ Y100A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* Y100A with His_6-tag and *E. coli cysG*	This study
pTH036	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ R102A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* R102A with His_6-tag and *E. coli cysG*	This study
pTH037	pET3a-*ahbA*^{Dd} *ahbB*$^{Dd\text{-His}}$ H118A *cysG*	Link and Lock of *Dd ahbA*, *Dd ahbB* H118A with His_6-tag and *E. coli cysG*	This study
pTH038	pET3a-*ahbA*^{Dd} R119 *ahbB*$^{Dd\text{-His}}$ R102A *cysG*	Link and Lock of *Dd ahbA* R119, *Dd ahbB* R102A with His_6-tag and *E. coli cysG*	This study

2.1.4 Primers

A list of all primers used for amplification in this study can be found in Table 2.3. All primers were obtained from Invitrogen Life Technologies. Nucleotide bases in bold letters refer to restriction enzyme site and nucleotides in lower case and red denote mutations to genomic sequence.

Table 2.3. List of primers. Ec: *E. coli*, Dv: *D vulgaris*, Dd: *D. desulfuricans*, Mb: *M. barkeri*.

Primer name	**Sequence**	**Restriction site**
Ec cysG for	5' CAG**CATATG**GATCATTTGCCTATATTTTGCC 3'	NdeI
Ec cysG rev	3' CAT**ACTAGT**TTAATGGTTGGAGAACCAG 5'	SpeI
Dd ahbA for	5' CAG**CATATG**ACCACGCAAACATCTG 3'	NdeI
Dd ahbA rev	3' CAT**ACTAGT**CTAGCTGTCATTGTCCATG 5'	SpeI
Dd ahbB for	5' CAG**CATATG**AGCCATCAATTCAGCC 3'	NdeI
Dd ahbB rev	3' CGG**ACTAGT**CTAGGTAAAATACGTCATG 5'	SpeI
Dv ahbA for	5' CAG**CATATG**ACGGAAGCGCACAAC 3'	NdeI
Dv ahbA rev	3' CAT**ACTAGT**CTAATCCATCCTGAAGTC 5'	SpeI
Dv ahbB for	5' CAG**CATATG**AGCCGTTACGATGAC 3'	NdeI
Dv ahbB rev	3' CGG**ACTAGT**tTAGAAATAGGTCATGGATG 5'	SpeI

Mb ahbA for	5’ CGG**CATATG**ATTGACATTGATAATTTAAAAGA TCAGCTC 3’	NdeI
Mb ahbA rev	3’ CGG**ACTAGT**TCACCTGATATCAAATTTTACC 5’	SpeI
Mb ahbB for	5’ CGG**CATATG**GATAAAACGGATGTAAAACTGC 3’	NdeI
Mb ahbB rev	3’CGT**ACTAGT**TTACAGTCTGACCCCAGTTTTTTT G 5’	SpeI
Dd ahbA R73A for	5’ CAAGATCATCgcGCGACTCGGGGC 3’	-
Dd ahbA R73A rev	3’ GCCCCGAGTCGcgCGATGATCTTG 5’	-
Dd ahbA R74A for	5’ GATCATCAGGgcACTCGGGGCCAAC 3’	-
Dd ahbA R74A rev	3’ GTTGGCCCCGAGTGCCCTGATGATC 5’	-
Dd ahbA H115A for	5’ CGGGCGTTACCgcCAACTACCTGC 3’	-
Dd ahbA H115A rev	3’ GCAGGTAGTTGgcGGTAACGCCCG 5’	-
Dd ahbA Y117A for	5’ GTTACCCACAACgcCCTGCGCGAGC 3’	-
Dd ahbA Y117A rev	3’ GCTCGCGCAGGgcGTTGTGGGTAAC 5’	-
Dd ahbA R119A for	5’ CACAACTACCTGgcCGAGCACGACTAC 3’	-
Dd ahbA R119A rev	3’ GTAGTCGTGCTCGgcCAGGTAGTTGTG 5’	-
Dd ahbB R55A for	5’ GTCCGGGGCCATCgcTCGCTTCGGGGC 3’	-
Dd ahbB R55A rev	3’ CGCCCCGAAGCGAgcGATGGCCCCGGAC 5’	-
Dd ahbB R56A for	5’ GGCCATCCGTgcCTTCGGGGCGAG 3’	-
Dd ahbB R56A rev	3’ CTCGCCCCGAAGgcACGGATGGCC 5’	-
Dd ahbB H63A for	5’ GCGAGCATCAAGgcCCAGAAAACAGG 3’	-
Dd ahbB H63A rev	3’ CCTGTTTTCTGGgcCTTGATGCTCGC 5’	-
Dd ahbB H98A for	5’ GCCATATTTCGgcCGTCTATTATCG 3’	-
Dd ahbB H98A rev	3’ CGATAATAGACGgcCGAAATATGGC 5’	-
Dd ahbB Y100A for	5’ CCATATTTCGCACGTCgcTTATCGCCCCAGC 3’	-
Dd ahbB Y100A rev	3’ GCTGGGGCGATAAgcGACGTGCGAAATATGG 5’	-
Dd ahbB R102A for	5’ GCACGTCTATTATgcCCCCAGCTCCGC 3’	-
Dd ahbB R102A rev	3‘ GCGGAGCTGGGGgcATAATAGACGTGC 5’	-
Dd ahbB H118A for	5’ GTACACCATGATA**gcCGGC**CGCAGCG 3’	NaeI
Dd ahbB H118A rev	3’ CGCTGCGGCCGgcTATCATGGTGTAC 5’	-

2.1.5 Media and solutions used for bacterial work

Luria-Bertani (LB) broth:

Tryptone	10	g
Yeast extract	5	g
NaCl	5	g

Made up to 1 L with dH_2O and autoclaved.

2YT broth:

Tryptone	16	g
Yeast extract	10	g
NaCl	5	g

Made up to 1 L with dH_2O and autoclaved.

Super optimal broth + catabolite repression (SOC):	Tryptone	20	g
	Yeast extract	5	g
	1 M NaCl	10	mL
	2 M Mg^{2+} stock (see below)	10	mL
	20% (w/v) glucose	10	mL
	1 M KCl	2.5	mL
Mg^{2+} stock:	$MgCl_2(H_2O)_6$	2	g
	$MgSO_4(H_2O)_7$	2.5	g

Made up to 10 mL with dH_2O.

For SOC media tryptone and yeast extract were dissolved in 967.5 mL dH_2O and autoclaved prior to the addition of the other components. All components that were not autoclaved were filter sterilised (0.2 µm pore size) prior to use.

Luria-Bertani agar:

15 g of Bacterial agar were added to 1 L LB broth before autoclaving.

10x M9 salts:	Na_2HPO_4	60	g
	KH_2PO_4	30	g
	NH_4Cl	10	g
	NaCl	5	g

Made up to 1 L with dH_2O and autoclaved.

Trace elements solution:

EDTA	5	g
$FeCl_3.6H_2O$	0.5	g
$ZnSO_4.7H_2O$	177	mg
$CuCl_2.2H_2O$	10	mg
$Co(NO_3)_2.6H_2O$	10	mg
$(NH_4)_6Mo_7O_{24}.4H_2O$	10	mg

All metals were dissolved separately in a small quantity of dH_2O before being mixed, adjusted to pH 7 and made up to 1 litre. The solution was then autoclaved before use.

Selenomethionine labelling media:	10x M9 salts	100	mL
	Trace elements	10	mL
	50% (v/v) Glycerol	20	mL
	1 M $MgSO_4$	1	mL
	1 M $CaCl_2$	0.4	mL
	Biotin (1 mg mL^{-1})	1	mL
	Thiamine (1 mg mL^{-1})	1	mL
	Ampicillin (100 mg mL^{-1})	1	mL
	Selenomethionine	125	mg
	dH_2O	865.6	mL

All solutions were made up separately and autoclaved (M9 salts, trace elements, glycerol) or filter sterilised ($MgSO_4$, $CaCl_2$, biotin, thiamine, ampicillin; 0.2 µm pore size) prior to mixing.

Table 2.4. Additives and antibiotics.

Additive	Stock concentration	Final concentration
IPTG	1 M in dH_2O	100 mM
Ampicillin	100 mg mL^{-1} in dH_2O	100 µg mL^{-1}
Chloramphenicol	34 mg mL^{-1} in dH_2O	34 µg mL^{-1}

2.1.6 Solutions used for DNA work

TE buffer:	Tris-HCl, pH 8.0	10 mM
	EDTA, pH 8.0	1 mM

6 x DNA loading buffer	Bromophenol blue (w/v)	0.25%
	Glycerol (v/v)	50%
	TE buffer (v/v)	50%

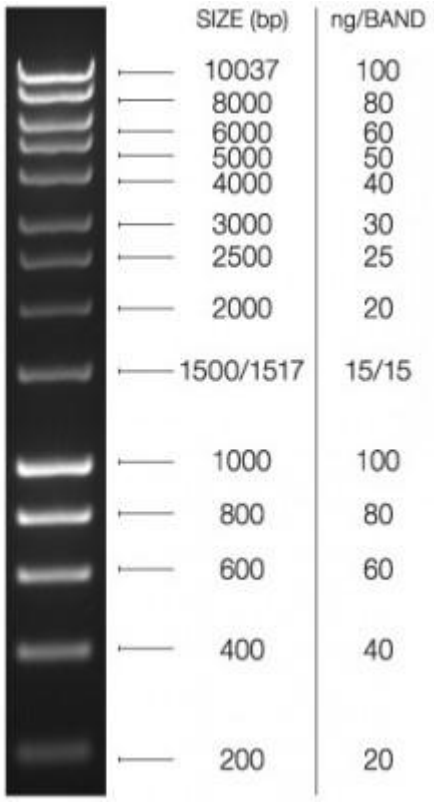


Figure 2.1. 5 µl Hyperladder 1 kb (Bioline) applied to a 1% (w/v) TBE/agarose gel visualised using ethidium bromide.

2.1.7 Solutions used for protein work

2.1.7.1 Solutions for nickel affinity chromatography

Charge buffer:	$NiSO_4(H_2O)_6$	100	mM
Binding buffer:	Tris-HCl, pH 8.0	20	mM
	NaCl	100	mM
	Imidazole	5	mM
Wash buffer:	Tris-HCl, pH 8.0	20	mM
	NaCl	100	mM
	Imidazole	50	mM
Elution buffer:	Tris-HCl, pH 8.0	20	mM
	NaCl	100	mM
	Imidazole	400	mM
Strip buffer:	Tris-HCl, pH 8.0	20	mM
	NaCl	500	mM
	EDTA	100	mM

Buffers used for purification of *D. desulfuricans* AhbA/B for crystallography contained HEPES in place of Tris-HCl at identical pH and concentrations.

2.1.7.2 Solutions for protein polyacrylamide gel electrophoresis.

10x running buffer:	Tris-HCl	30	$g\ L^{-1}$
	Glycine	144	$g\ L^{-1}$

$10\ mL\ L^{-1}$ 10% (w/v) SDS added to 1x running buffer for SDS gels.

2x SDS sample buffer:	0.5 M Tris-HCl, pH 6.8	6.0	mL
	Glycerol	4.8	mL
	SDS 10% (w/v)	9.6	mL
	Bromophenol blue 0.05% (w/v)	1.2	mL
	dH_2O	24.0	mL

14 μL β-mercaptoethanol added per mL 2x SDS sample buffer.

2x Native sample buffer:	0.5 M Tris-HCl, pH 6.8	6.0	mL
	Glycerol	4.8	mL
	SDS 10% (w/v)	9.6	mL
	Bromophenol blue 0.05% (w/v)	1.2	mL
	dH_2O	24.0	mL

Denaturing molecular weight marker - Dalton VII (Sigma):

Protein	Molecular weight (Da)
Bovine serum albumin	66,000
Ovalbumin	45,000
G-3-P dehydrogenase	36,000
Carbonic anhydrase	29,000
Trypsin inhibitor	20,100
A-Lactalbumin	14,200

Denaturing molecular weight marker – Prestained protein marker, broad range (New England BioLabs):

Protein	Molecular weight (Da)
MBP-β-galactosidase	175,000
MBP-parayosin	80,000
MBP-CBD	58,000
CBD-*Mxe* Intein-2CBD	46,000
CBD-*Mxe* Intein	30,000
CBD-BmFKBP13	25,000
Lysozyme	17,000
Aprotinin	7,000

MBP: maltose-binding protein, CBD: chitin binding domain. BmFKBP13: *Bombyx mori* FK506-binding protein.

Denaturing molecular weight marker – Untained protein molecular weight marker (Fisher Scientific):

Protein	Molecular weight (Da)
β-galactosidase	116,000
Bovine serum albumin	66,200
Ovalbumin	45,000
Lactate dehydrogenase	35,000
REase Bsp981	25,000
β-lactoglobulin	18,400
Lysozyme	14,400

Molecular weight marker for native electrophoresis (GE Healthcare):

Protein	Molecular weight (Da)
Thyroglobulin	669,000
Ferritin	440,000
Catalase	232,000
Lactate dehydrogenase	140,000
Bovine serum albumin	66,000

Coomassie blue stain:	Trichloroacetic acid (100%)	250.00	mL
	Coomassie blue R250	0.60	g
	SDS	0.10	g
	Tris-HCl	0.25	g
	Glycine	0.15	g

Made up to 500 mL with dH_2O.

Table 2.5. SDS and native gel compositions.

SDS gels				
Running gels	12.5%	15%	Stacking gel	5%
dH_2O (mL)	3.4	2.2	dH_2O (mL)	3.4
30% Acrylamide (mL) Acryl/Bis™ 29:1	6.3	7.5	30% Acrylamide (mL) Acryl/Bis™ 29:1	1.5
1.5 M Tris-HCl pH 8.8 (mL)	3.8	3.8	1.5 M Tris-HCl pH 6.8 (mL)	1.9
10% (w/v) SDS (mL)	1.5	1.5	10% (w/v) SDS (mL)	0.75
10% (w/v) APS (mL)	0.15	0.15	10% (w/v) APS (mL)	0.075
TEMED (mL)	0.01	0.01	TEMED (mL)	0.01
Native gels				
Running gels	7%	9%	Stacking gel	5%
dH_2O (mL)	7.8	6.7	dH_2O (mL)	4
30% Acrylamide (mL) Acryl/Bis™ 29:1	3.45	4.56	30% Acrylamide (mL) Acryl/Bis™ 29:1	1.5
1.5 M Tris-HCl pH 8.8 (mL)	3.8	3.8	1.5 M Tris-HCl pH 6.8 (mL)	1.9
10% (w/v) APS (mL)	0.15	0.15	10% (w/v) APS (mL)	0.075
TEMED (mL)	0.01	0.01	TEMED (mL)	0.01

2.1.7.3 Solutions for in-gel haem staining

Fixing solution:	12.5% (w/v) Trichloroacetic acid		
Staining solution:	*o*-dianisidine	0.2	g
	Acetic acid	20.0	mL
	dH_2O	160.0	mL
	0.5M sodium citrate, pH 4.4	20.0	mL
	30% (w/w) H_2O_2	0.4	mL

2.1.7.4 Solutions for gel filtration chromatography (FPLC)

Sample buffer: AhbA/B complex size estimation	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
Sample buffer: AhbA/B purification - crystallisation	HEPES, pH 8.0	20 mM
	NaCl	100 mM
Wash buffers:	NaOH	0.2 M
	Ethanol	20.0%

All solutions were filtered prior to use.

Table 2.6. Molecular weight markers for gel filtration chromatography (Sigma).

Protein	Approximate molecular weight (Da)	Concentration
Carbonic anhydrase, bovine erythrocytes	29,000	3 mg mL^{-1}
Albumin, bovine serum	66,000	10 mg mL^{-1}
Alcohol dehydrogenase, yeast	15,0000	5 mg mL^{-1}
β-amylase, sweet potato	200,000	4 mg mL^{-1}
Apoferritin, horse spleen	443,000	10 mg mL^{-1}
Thyroglobulin, bovine	669,000	8 mg mL^{-1}
Blue dextran	2,000,000	2 mg mL^{-1}

2.1.7.5 Solutions for HPLC-MS

Tetrapyrrole HPLC-MS analysis:

Solution A:	Trifluoroacetic acid (v/v)	0.1%
Solution B:	Acetonitrile	100%

Protein HPLC-MS analysis:

Solution A:	Trifluoroacetic acid (v/v)	0.05%
Solution B:	Acetonitrile	80.000%
	Trifluoroacetic acid (v/v)	0.045%

All solutions were filtered prior to use.

2.1.7.6 Solutions for crystallisation

Ammonium acetate	2 M
Sodium acetate, pH 4.8-5.6	1 M
Tri-sodium citrate, pH 4.8-5.6	1 M
Polyethylene glycol MW 4000	50% (w/v)
Polyethylene glycol MW 3350	50% (w/v)
Barium chloride	0.1 M

Solutions other than polyethylene glycol (PEG) were filtered (0.2 µm pore size) prior to use. Buffers were pH corrected using glacial acetic acid and diluted appropriately before screening. Original crystal screen compositions are listed in the Appendix.

2.1.7.7 Haemin solution

Haemin	0.1	g
Tween 80	2.5	mL
dH_2O	22.5	mL
NaOH (10N)	2	drops

2.1.7.8 Pyridine haemochrome solutions

Pyridine solution:	NaOH	200	mM
	Pyridine (v/v)	40%	
Oxidising solution:	Potassium ferricyanide	0.1	M

2.2 Microbiological methods

2.2.1 Sterilisation

Unless stated otherwise media and buffers were sterilised for 15 min at 121 °C and 1 bar pressure in an autoclave. Temperature sensitive substances were filter sterilised (0.2 µm pore size).

2.2.2 Storage of bacteria

For long-term storage of bacteria, glycerol stocks were prepared. Glycerol was added to an overnight bacterial culture to a final concentration of 25% (v/v). The culture was then incubated on ice for 30 min before storage at -80 °C.

2.2.3 Plate cultures

Bacteria were usually streaked directly from a glycerol stock onto agar plates and antibiotics, ALA was added where required. The agar plates were incubated overnight at 37 °C.

2.2.4 Liquid cultures

Liquid cultures were inoculated with a single colony from an agar plate culture. The medium was supplemented with antibiotics and ALA where required. Liquid cultures were shaken at ≈ 160 rpm in baffled flasks overnight at 37 °C.

2.2.5 Preparation of *E. coli* competent cells

Competent *E. coli* cells were prepared based on a method previously described (Sambrook *et al.*, 1989). A bacterial overnight starter culture derived from a single colony was inoculated into 10 mL fresh LB broth and grown to an OD_{600} of 0.3. The cells were cooled on ice for 15 min and centrifuged at 850 x *g* at 4 °C. The pellets were gently resusupended in 25 mL of ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 min. Cells were collected again by centrifugation and resuspended in 0.25 mL of 0.1 M $CaCl_2$ containing 25% (v/v) glycerol. Aliquots of 30 µL were frozen rapidly and stored at -80 °C.

2.2.6 Transformation of *E. coli* competent cells

Competent cells were defrosted on ice for 10 min before adding 1 µL plasmid DNA. The mixture was incubated on ice for 15 min and then heat-shocked by incubation at 42 °C for 50 sec before rapid transfer to ice and incubation for 2 min. After the addition of 250 µL SOC media, the cells were incubated at 37 °C for 20-60 min to allow antibiotic resistance expression. The mixture was then spread on a LB agar plate containing the required antibiotics and incubated at 37 °C overnight.

2.2.7 Recombinant protein overproduction in *E. coli*

The *E.coli* strains BL21 (DE3) and BL21 (DE3) pLysS were transformed with a vector containing the gene(s) of interest cloned in frame. The recombinant strain was grown in LB with ampicillin (and chloramphenicol for pLysS strain) at 37 °C and shaken at ≈160 rpm until the culture reached an OD_{600} of approximately 0.6. Protein production was induced with 0.4 mM IPTG overnight at 19 °C. The cells were collected by centrifugation at 3,500 x *g* for 20 min at 4 °C. The pellet was resuspended in 15 mL of binding buffer (Section 2.1.7.1) and either directly used for protein purification (Section 2.4.1) or stored at -80 °C.

Cultures containing trapping strain plasmid constructs were grown for 24 hours at 19 °C in 2YT media supplemented with ampicillin. Production of tetrapyrroles was enhanced by the addition of 1 mL of ALA (20 mg mL^{-1}) 6 hours prior to harvest. Cells were harvested and stored as described previously.

2.2.8 Production of selenomethionine labelled protein

Selenomethionine labelled protein was produced using the KRX strain of *E. coli*. 250 mL LB media, containing 0.8% (w/v) glucose and 100 mg/L ampicillin, were inoculated with a 5 mL overnight culture of *E. coli* KRX transformed with plasmid pDP031. This culture was grown overnight at 37 ˚C (<18 hours) before being harvested and resusupended in labelling media (Section 2.1.5). Cells were incubated at 25 ˚C for 30-60 mins before induction with 0.2% (w/v) rhamnose for 1 day. Incorporation of selenomethionine was confirmed by mass spectrometry (Section 2.5.2.2).

2.2.9 Lysis of cells by sonication

Harvested cells were lysed by sonication using a Sonics Vibracell Unltrasonic processor, with an output watt of between 20 and 30 in 30 sec bursts with 30 sec breaks repeated 6 times. The sonicated cells were centrifuged at 35, 000 x *g* for 20 min at 4 ℃ to remove cell debris. For anaerobic protein purification harvested cells were transferred into an anaerobic chamber 1 hour prior to sonication. Sonication was carried out with 10 sec bursts with 30 sec breaks and repeated 18 times to prevent overheating of the sample.

2.3 Molecular biological methods

2.3.1 PCR reactions

All PCR reactions were performed in an Eppendorf Mastercycler 5341 PCR machine using the FastStart High Fidelity PCR System (Roche). The basic PCR reaction and the cycles used are outlined below:

Table 2.7. Basic PCR reaction.

Component	µL
dH_2O	37.5-36.5
10x PCR buffer (containing 18 mM $MgCl_2$)	5
DMSO	0-1
5 mM dNTPs each	2
10 µM 5' primer	2
10 µM 3' primer	2
DNA template	1
Taq polymerase (5 U μL^{-1})	0.5

Table 2.8. Temperature protocol for PCR reactions.

Step	Temp	Time	Cycles	Function
1	96 °C	2 min	1	Initial denaturation of DNA
2	95 °C	30 sec	35	Denaturation of amplified DNA
3	55-58 °C	30 sec		Annealing of primers
4	72 °C	1 min per 1000 bp		Elongation
5	72 °C	5 min	1	Final elongation
Stop	4 °C	Constant		

Once the PCR reaction was completed the product was subjected to agarose gel electrophoresis (Section 2.3.2).

2.3.2 Electrophoresis of DNA

DNA fragments were separated by agarose gel electrophoresis.

2.3.2.1 Agarose gel

The agarose gel percentage was chosen according to the size of the DNA fragments to be separated [0.7-1.5% (w/v) agarose in TAE buffer]. Routinely, a 1% (w/v) agarose gel was used. Agarose gels were set with the appropriate amount of agarose in 1x TAE buffer with the addition of ethidium bromide to a final concentration of 0.5 µg mL^{-1}. The DNA samples containing 20% (v/v) DNA loading buffer were loaded into the agarose gel wells and electrophoresis was carried out at 70 V for approximately 1 hour using a SubCell GT electrophoresis tank (BioRad) connected to a Power PAC 300 power supply (BioRad).

2.3.2.2 Visualisation of DNA

Ethidium bromide is a fluorescent dye that intercalates between the DNA base pairs. UV radiation (312 nm) is absorbed by the ethidium bromide and re-emitted at 590 nm in the red-orange wavelength range. This allows visualisation of the DNA using a UV transilluminator.

2.3.3 Isolation and purification of a DNA fragment

The DNA band of interest was excised from an agarose gel using a scalpel blade. Purification was carried out using a QIAquick® Gel Extraction Kit (Qiagen) according to the protocol for using a microcentrifuge as stated in the handbook.

Digests of PCR products were purified using a QIAquick® Gel Extraction Kit (Qiagen) according to the protocol for QIAquick® PCR purification using a microcentrifuge.

2.3.4 Ligation of DNA

Vectors and inserts were digested with the relevant restriction enzymes as described in Section 2.3.6 and purified (Section 2.3.3). The ligation of DNA fragments into the vector was carried out at room temperature for 2 hours or at 4 °C overnight using the following components:

Insert	2.5 µL
Vector	1.5 µL
2x rapid ligation buffer	5 µL
(Promega)	1 µL
T4 DNA Ligase (3 U μL^{-1}; Promega)	

2.3.5 Isolation of plasmid DNA

A QIAprep® Miniprep Kit (Qiagen) was used for the purification of plasmid DNA as described in the handbook according to the protocol for using a microcentrifuge.

2.3.6 Restriction digests

Plasmid or PCR product DNA was digested using the relevant enzymes (10 U μL^{-1}) and the optimal buffer chosen according to either the Promega or New England Biolabs information provided. The reactions were incubated for 2 hours at the temperature required by the restriction enzyme before being subjected to electrophoresis.

Table 2.9. Typical DNA digest protocol

Component	Single digest	Double digest
dH_2O	4 µL	3.5 µL
Restriction enzyme 1	1 µL	0.5 µL
Restriction enzyme 2	-	0.5 µL
10x buffer	1 µL	1 µL
Plasmid/PCR DNA	4 µL	4 µL

2.3.7 Cloning into vectors

2.3.7.1 Single gene cloning into vectors

The vectors and the genes of interest were digested at their restriction sites, which were introduced into the PCR product (Section 2.3.1) via the 5' and 3' primers, respectively (Section 2.3.6). After isolation and purification of the required DNA fragments (Section 2.3.3.1), vectors and genes were ligated together (Section 2.3.4). Competent *E. coli* (JM109 or DH5α™) were transformed with the ligation mixture and plated onto appropriate media. Single colonies were selected, grown and the recombinant plasmids amplified and sequenced to verify the DNA sequence (GATC Biotech AG). The vector pET14b provides the gene product with an N-terminal short peptide sequence containing a stretch of six consecutive histidine residues, whereas the pET3a vector does not.

2.3.7.2 Multiple gene cloning

Multiple gene constructs were formed by the consecutive cloning of genes into pET3a via the 'Link and Lock' cloning strategy developed by Dr. Evelyne Deery and Dr. Helen McGoldrick (McGoldrick *et al.*, 2005). The 'Link and Lock' method exploits compatible cohesive ends allowing fusion of two fragments digested by different enzymes forming a site which cannot be recleaved by either of the original enzymes. This permits the systematic addition of genes to a plasmid construct using the same restriction sites. Figure 2.2 summarises the 'Link and Lock' process. A PCR was performed and the genes were individually cloned into pET3a introducing a SpeI site at the 3' end of the genes via the 3' primers (Table 2.3). The ribosome binding site and genes to be subcloned into the 'Link and Lock' construct were excised using the restriction enzymes XbaI and EcoRI (Fig. 2.2B). The recipient plasmid containing the initial gene(s) was digested using SpeI and EcoRI. Ligating the two fragments caused the fusion of SpeI and XbaI, as they have compatible cohesive ends, leading to an un-cleavable site, whereas the EcoRI sites fused as normal (Fig. 2.2C). The removal of the SpeI site means the newly formed construct only contained one SpeI site so the process could be repeated again to add more genes.

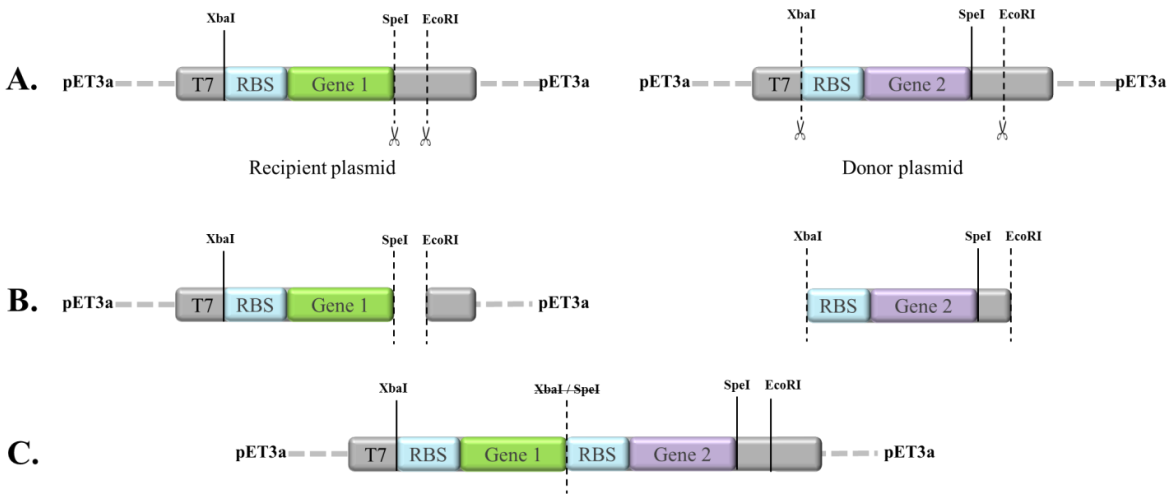


Figure 2.2. Schematic of the 'Link and Lock' cloning technique. A. Target genes are cloned into identical vectors and digested with appropriate restriction enzymes. **B.** Recipient plasmid and donor fragment containing gene and ribosome binding site (RBS) are ligated. **C.** Construct is formed with compatible cohesive ends (XbaI and SpeI) ligating and forming an uncleavable site. The new construct is then capable of undergoing the process again to add a new gene of interest.

2.3.7.3 Generation of mutants

The mutations of *D. desulfuricans* AhbA and AhbB were created by Dr. Susanne Schroeder (University of Kent) using the QuikChange® II Site-Directed Mutagenesis Kit (Aligent). Briefly the method is described as follows. QuikChange® II Site-Directed Mutagenesis Kit can be used to introduce point mutations that switch amino acids as well as delete and insert single or multiple base pairs. The mutagenesis method was performed using *PfuUltra*™ high-fidelity (HF) DNA polymerase. The basic procedure utilises a supercoiled double-stranded DNA vector with an inset of interest and two synthetic oligonucleotide primers (Table 2.3), both of which contain the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, were extended during temperature cycling by *PfuUltra*™ HF DNA polymerase. Extensions of the oligonucleotide primers generated a mutated plasmid containing staggered nicks. Following temperature cycling, the product was treated with DpnI. The DpnI endonuclease (target sequence: 5' –GmATC-3') is specific for methylated and hemimethylated DNA and was used to digest parental DNA template and to select for mutation-containing synthesised DNA. The nicked vector DNA containing the desired mutations was then transformed into XL1-Blue supercompetent cells and the nicks were repaired. Figure 2.3 represents a schematic diagram of the procedure. Clones were

purified and sequenced prior to formation of a mutant pET3a-*ahbA/B* construct using the Link and Lock technique.

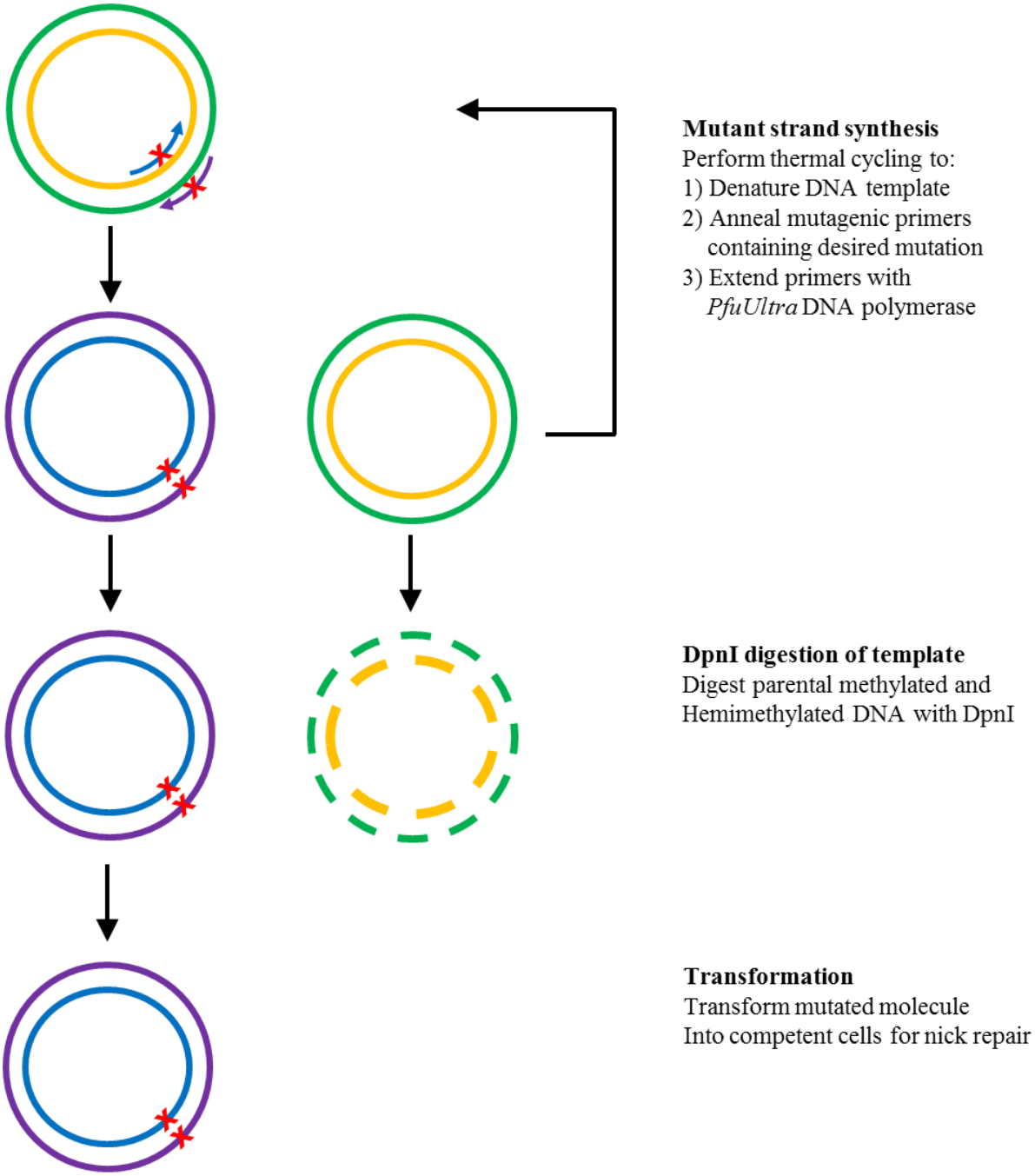


Figure 2.3. Schematic diagram of the QuikChange™ II site directed mutagenesis method.

2.4 Biochemical methods

2.4.1 Protein purification by immobilised metal ion affinity chromatography

The His_6tag sequence of the fused protein has affinity for transition metals such as Co^{2+}, Ni^{2+}, Zn^{2+} and Fe^{3+}. Chelating Sepharose™ Fast Flow resin (GE Healthcare) loaded with the suitable metal ion will therefore selectively retain protein with exposed histidines, which form complexes with the metal ions. The bound protein can be isolated by competitive elution with increasing concentrations of imidazole.

Proteins were purified at room temperature, but were kept on ice as much as possible. Columns containing approximately 3 to 5 mL of the resin were used to purify proteins from a 1 L culture. The resin was initially rinsed with dH_2O, charged with $NiSO_4$ (charge buffer) and equilibrated with 50 mL of binding buffer. Soluble cell extract acquired after sonication and centrifugation (Section 2.2.9) was applied to the resin and allowed to flow through by gravity. Unbound proteins were washed away with 15 mL of binding buffer and 15 mL of wash buffer. Elution buffer was applied to the column to recover the protein which was collected in 2 mL fractions. The resin was regenerated with 15 mL strip buffer and washed with dH_2O. Composition of buffers can be found in Section 2.1.7.1

2.4.2 Buffer exchange

Purified protein was desalted on a pre-packed disposable Sephadex G25 column (PD-10, bed volume 8.3 mL; GE Healthcare). In brief the method consists on the following steps: (1) column equilibration with 25 mL of buffer into which the protein was to be exchanged, (2) application of 2.5 mL protein solution, (3) elution of protein sample in 3.5 mL of desired buffer, (4) restoration of column using dH_2O.

2.4.3 Gel filtration chromatography

Gel filtration chromatography was performed using an ÄKTA P-920 FLPC system (Amersham Biosciences). Samples were resolved on a SuperdexTM 200 10/300 GL column (GE Healthcare) equilibrated in 20mM Tris HCl, pH8, 100mM NaCl. 1 mL sample was loaded onto the column and the protein was eluted using a flow rate of 0.5 mL min^{-1}.

Calibration of the column was achieved using MWGF-1000 gel filtration molecular weight marker kit (Sigma; Table 2.6).

2.4.4 Bradford protein assay

The Bradford assay relies on the dye Coomassie brilliant blue G-250 binding to proteins, most readily to arginyl and aromatic residues, causing a shift in the absorbance maxima of the dye from the red cationic form (470-650 nm) to the blue anionic form (590-620 nm) in the presence of protein. Therefore, the quantity of the protein can be estimated by measuring the amount of dye in the blue form at 595 nm (Bradford, 1976). The protein assay dye from BioRad was added to diluted protein solution, and the reaction carried out as per the instructions. The reaction was left at room temperature for 10-30 min and the OD_{595} was measured. Simultaneously a standard curve with bovine serum albumin (0, 5, 10, 15, 20 and 25 mg mL^{-1}) was generated.

2.4.5 A_{280} protein concentration estimation

Protein concentrations were estimated using readings of the OD_{280} and estimated protein extinction coefficients derived by ProtParam on the ExPASy server (Gasteiger *et al.*, 2005). Protein concentrations were calculated using the following equations:

A_{280} = Concentration x Extinction coefficient

To convert to mg mL^{-1}:

Protein concentration (mg mL^{-1}) = Molecular mass x Concentration

2.4.6 Polyacrylamide gel electrophoresis

2.4.6.1 SDS-PAGE

The composition of the gel, buffers and stain were previously described in Section 2.1.7.2. Polyacrylamide gels were run as previously described (Laemmli, 1970). The sample was denatured by 1:1 addition of SDS sample buffer and boiling for 5 min.

Between 5 and 10 µL of denatured sample was loaded into each well and 7 µL molecular mass marker was run on each gel to estimate the relative molecular mass of the protein of interest. Electrophoresis was performed at a constant voltage of 200 V using an Atto Dual Mini Slab AE6450 electrophoresis apparatus and an Atto Mini Power electrophoresis power supply SJ1082 (GRI Ltd.). The gel was stained with coomassie blue stain and detained using dH_2O.

2.4.6.2 Native PAGE

The composition of the gel and buffers were previously described in Section 2.1.7.2. After the 1:1 addition of native sample buffer 5-10 µL of the sample was directly loaded onto the gel. 7 µL native molecular marker was run on each gel to estimate relative molecular mass of the protein of interest. Electrophoresis was carried out at a constant voltage of 100 V using the same apparatus as described in Section 2.4.6.1.

2.4.6.3 Haem staining protocol

A standard 12.5% (v/v) acrylamide gel was run as described in Section 2.4.6.1. Instead of staining with coomassie blue the gel was fixed and stained using the following protocol. The acrylamide gel was washed in 200 mL 12.5% (w/v) TCA for 30 min and then for a further 30 min in dH_2O. Bands were developed using haem staining solution (Section 2.1.7.3) and appeared within 5-15 min. Cytochrome *c* was used as a positive control.

2.4.7 Anaerobic techniques

Anaerobic work was carried out under the nitrogen environment of the glove box (Belle Technologies) containing less than 2 ppm oxygen. Liquids such as water and buffer stocks were degassed under argon prior to transfer into the glove box. Chemicals were taken into the glove box as powders and were made up with degassed dH_2O. Resins such as PD-10 columns and DEAE columns were brought into the glove box 24 hours prior to use and equilibrated with at least 50 mL of anaerobic buffer.

2.4.8 *In vitro* preparation of sirohaem

Sirohydrochlorin was prepared using a multi-enzyme approach as previously described (Bali *et al.*, 2011). The plasmid pETcoco-2-*cobA-hemBCD-sirC* is a multi-gene construct

engineered by Dr. Evelyne Deery carrying genes for the enzymes ALA dehydratase (HemB), porphobilinogen deaminase (HemC), uro'gen III synthase (HemD), SUMT (CobA) and precorrin-2 dehydrogenase (CobA), all of which are fused to His_6-tags. Lysate of an *E. coli* strain carrying the plasmid was purified (Section 2.4.1) and the most concentrated eluted protein fraction was transferred into the glove box. Following buffer exchange of the protein into anaerobic buffer using a PD-10 column (Section 2.4.2) the protein was pooled with 1 mL cofactor mix. The cofactor mix consisting of 20 mg mL^{-1} SAM, 6.5 mg mL^{-1} NAD and 10 mg mL^{-1} ALA in 20 mM Tris-HCl, pH 8.0, and 100 mM NaCl, was adjusted to pH 8.0 using 2 M NaOH prior to use. The reaction was incubated at room temperature overnight.

Sirohaem was formed by the addition of a 10-fold excess of $FeSO_4$ to the sirohydrochlorin and subsequent incubation at room temperature for 4 hours. Sirohaem was purified from excess iron and porphyrins using a DEAE column (equilibrated in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl) in a stepwise gradient from 100 mM to 800 mM NaCl. Sirohaem was eluted in 20 mM Tris-HCl, pH 8.0 with 1M NaCl.

2.4.9 *In vitro* preparation of didecarboxysirohaem for NMR

Didecarboxysirohaem was prepared enzymatically from sirohaem under anaerobic conditions (<2 ppm O_2). Sirohaem was incubated with a 100-fold excess of purified *D. vulgaris* AhbA/B overnight at 37 °C. Didecarboxysirohaem was purified from the protein using a DEAE column. The column was equilibrated in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl prior to sample loading. The column was then washed with acidified dH_2O with a pH gradient from 8 to 2.5, at which point the tetrapyrrole was eluted. The sample was then freeze-dried, resuspended in 0.5 ml 100% deuterium oxide and sealed in a septum screw-capped 5 mm NMR tube while maintaining anaerobic conditions.

2.4.10 Calculating activity of *D. desulfuricans* AhbA/B

The following method was carried out in an anaerobic glove box (Belle Technology). 23 µM of sirohaem were incubated at 37 °C for 10 min prior to protein addition. Purified *D. desulfuricans* AhbA/B was then added to the reaction mixture to a final concentration of 2.3 µM and incubated at 37 °C for the duration of the reaction. 100 µL aliquots of the reaction mixture were removed at time points and added to an equal volume of acetonitrile to stop the reaction. These samples were then incubated at 4 °C for 10 min before being centrifuged at 2000 x *g* for 10 min to pellet the precipitated protein. The samples were then sealed in a

HPLC vial, removed from the glove box and analysed using HPLC-MS (Section 2.5.2). Peak integration was performed using Compass™ Data Analysis (Bruker). Data were fitted using in-house software developed by Dr. Andrew Lawrence, to the following equations:

$$v_1 = \frac{V_{max1}\, S_1}{K_{m1}\,(1 + S_2/K_{m2} + S_3/K_i) + S_1}$$

$$v_2 = \frac{V_{max2}\, S_2}{K_{m2}\,(1 + S_1/K_{m1} + S_3/K_i) + S_2}$$

where v_1 is rate of sirohaem decarboxylation, S_1 is the concentration of sirohaem, v_2 is the rate of monodecarboxysirohaem decarboxylation, S_2 is the concentration of monodecarboxysirohaem, S_3 is the concentration of didecarboxysirohaem, K_i is the tight-binding constant of didecarboxysirohaem to the enzyme and V_{max1} / V_{max2} and K_{m1} / K_{m2} are the kinetic parameters of each reaction respectively.

2.4.11 Pyridine haemochrome assay

Pyridine haemochrome assays were performed as previously described (Berry & Trumpower, 1987). Spectra were recorded between 520 and 620 nm. 0.5 mL 200 mM NaOH and 40% (v/v) pyridine and 3 µL 0.1 M potassium ferricyanide were placed in a cuvette. 0.5 mL sample (~10 mg mL^{-1}) was then added and mixed thoroughly before an oxidised spectrum was recorded. Solid sodium dithionite was added (2-5 mg) and several reduced spectra were taken. The oxidised spectrum was subtracted from the first stable reduced spectrum to provide the final pyridine haemochrome spectrum.

2.4.12 Haem extraction protocol

Haem extractions were carried out as previously described (Lubben & Morand, 1994). Non-covalently bound haem was extracted from 0.05 ml of sample (membrane suspensions or protein solutions) with 0.45 mL of acetone-HCl (19:1, v:v) at room temperature for 20 min. After centrifuging for 2 min at 14,000 x *g*, 1.0 mL of ice-cold dH_2O and 0.3 mL of ethyl acetate were added to the supernatant, and the sample was vortexed and centrifuged again. The ethyl acetate phase was recovered and the solvent removed by a vacuum centrifuge (miVac DUO Concentrator; Genevac). The residues were dissolved in 100 µL of dH_2O and analysed by mass spectrometry according to the method in Section 2.4.14.1.

2.4.13 UV-visible spectrophotometry

All UV-visible spectra were recorded on a Varian Cary 50 Bio UV/visible spectrophotometer over a range of 200-800 nm. Spectra of anaerobic samples were recorded inside a glove box (Belle Technology Ltd.) using a fibre optic probe.

2.4.14 HPLC-MS

2.4.14.1 HPLC-MS of tetrapyrrole derivatives

Protein was removed prior to analysis by HPLC-MS either by acidification with equal volumes of 0.1% (v/v) TFA and purification on a TELOS C18 column, eluting in 100% acetonitrile, or by diluting with equal volumes of acetonitrile before centrifugation at 2000 x *g*.

Samples were resolved on an ACE 5AQ column (2.1 x 150 mm; Advanced Chromatography Technologies) attached to an Agilent 1100 series HPLC equipped with diode array detector and coupled to a micrOTOF-Q mass spectrometer (Bruker). The column was developed with a binary gradient at a flow rate of 0.2 mL per min. Solvent A was 0.1% (v/v) TFA and solvent B was acetonitrile.

For sirohaem and sirohaem derivatives the column was equilibrated with 5% B. Following sample injection the concentration of B was increased to 20% over 6 min and then to 30% at 25 min and 100% at 35 min where it was held for 5 min before returning to starting conditions. The total length of each run was 50 min. For samples containing haem a linear gradient was used starting at 20% B and reaching 100% B in 30 min. Data was processed using Compass™ Data Analysis (Bruker).

2.4.14.2 HPLC-MS of proteins

Samples were resolved on a Jupiter® 5 µm C4 column (2 x 50 mm; Phenomenex Inc) attached to an Agilent 1100 series HPLC equipped with diode array detector and coupled to a micrOTOF-Q (Bruker) mass spectrometer. The column was developed with a binary gradient at a flow rate of 0.2 mL per min. Solvent A was 0.05% (v/v) TFA and solvent B was 0.045% (v/v) TFA and 80% (v/v) acetonitrile.

The column was equilibrated with 5% solvent B prior to the injection of 15 µg of protein. The concentration of B was increased to 100% over 5 min where it was held for 5 min before returning to starting conditions.

Data was processed using Compass™ Data Analysis (Bruker). Selenomethionine labelling efficiency was determined using the following equation:

$$\% \text{ Labelling Efficiency} = 100 \times \frac{(MW_{labeled\_experimental} - MW_{unlabeled\_experimental})}{(MW_{labeled\_theoretical} - MW_{unlabeled\_theoretical})}$$

2.4.15 Nuclear Magnetic Resonance (NMR) spectroscopy

The NMR sample was prepared as previously described (Section 2.4.9). NMR data were collected by Dr. Mark Howard (University of Kent), and analysed by Dr. Andrew Lawrence (University of Kent). All NMR data were obtained at 298 K using a 14.1 T (600 MHz ^{1}H) Bruker Avance III NMR spectrometer equipped with a TCI cryoprobe. NMR assignments were obtained using ^{1}H homonuclear DQF-COSY, NOESY (500 ms), rotating-frame Overhauser effect spectroscopy [ROESY (200 ms)], and total correlation spectroscopy [TOCSY (80 ms)] and ^{13}C, ^{1}H heteronuclear sequential quantum correlation (HSQC), computed tomography-HSQC (27 ms), heteronuclear multiple bond correlation, and HSQC-TOCSY (80 ms) experiments. Mixing/constant time periods are shown in parentheses. NMR data processing was achieved using TopSpin 3.0 and NMR data analysis using the didecarboxysirohaem analysis suite, version 2.0.

2.4.16 Redox potentiometry

Determination of the midpoint redox potentials was undertaken in collaboration with Prof. Andy Munro and Dr. Kirsty McLean (University of Manchester). Briefly, the method used is as follows and is essentially as described previously (Munro *et al.*, 2001). Redox titrations were performed in a Belle Technology glove box under a nitrogen atmosphere, and all solutions were degassed under vacuum with argon. Oxygen levels were maintained at less than 2 ppm. The protein in 100 mM phosphate buffer, 200 mM KCl, 10% (v/v) glycerol pH 7.0, was titrated electrochemically according to the method of Dutton (Dutton, 1978) using sodium dithionite as the reductant and potassium ferricyanide as the oxidant. Mediators (2 µM phenazine methosulphate, 5 µM 2-hydroxy-1,4-naphthoquinone, 0.5 µM methyl viologen and 1 µM benzyl viologen) were included to mediate the range between +100 to -

480 mV, as described previously (Ost *et al.*, 2001, Munro *et al.*, 2001). At least 10 min was allowed to elapse between each addition of reductant or oxidant to allow stabilisation of the electrode. Spectra (250-800 nm) were recorded using a Cary UV-50 Bio UV-visible spectrophotometer via a fibre optic absorption probe immersed in the enzyme solution. The electrochemical potential of the solution was measured using a Mettler Toldeo FiveEasy meter coupled to a Pt/Arganthan electrode at 25 °C. The electrode was calibrated using the $Fe^{3+/2+}$ EDTA couple (+108 mV). A factor of +207 mV was used to correct relative to the standard hydrogen electrode.

2.4.17 Electron paramagnetic resonance spectroscopy (EPR)

EPR spectra were recorded and analysed in collaboration with Dr. Steve Rigby (University of Manchester). EPR experiments were performed on a Bruker ELEXSYS E500 spectrometer operating at X-band employing a Super High Q cylindrical cavity (Q factor ~16,000) equipped with an Oxford Instruments ESR900 liquid helium cryostat linked to an ITC503 temperature controller. Signals were recorded at 12 K employing a microwave power of 1 mW, modulation frequency of 100 kHz and modulation amplitude of 5 G (unless otherwise stated). Protein samples were prepared as described previously (Section 2.4.1) under anaerobic conditions and transferred into EPR tubes. Samples were then rapidly frozen in liquid nitrogen.

2.4.18 Crystallisation and X-ray crystallography

2.4.18.1 Crystallisation by hanging drop method

The hanging drop variant of the vapour diffusion method of protein crystallisation (McPherson, 1976) was employed. Proteins were purified using nickel affinity chromatography as previously described (Section 2.4.1), and subjected to further purification by FPLC, with the largest homogenous peak (correlating to a heterodimeric species) being selected for crystallisation. Protein was then concentrated to 8-20 mg mL^{-1} using an Amicon Ultra-15 centrifugal filter unit (Merck Millipore) with a molecular weight cut-off of 10,000 Da.

Initial crystallisation screens were performed using pre-formulated Structure Screen 1™, Structure Screen 2™ and MIDAS™ (Molecular Dimensions Ltd.) in an attempt to cover a wide range of precipitants to determine the conditions in which crystals would grow. Further

development of crystallisation conditions was performed using Clear Strategy Screen I™, Clear Strategy Screen II™ (Molecular Dimensions Ltd.) and Additive Screen (Hampton Research) according to the manufacturers descriptions.

For a standard aerobic crystallisation experiment 1 mL reservoir solution was placed in each well of a 24-well XRL plate (Molecular Dimensions Ltd.). Protein was mixed with reservoir solution on a siliconised coverslip (Hampton Research) either at 2 µL : 2 µL or 1 µL : 1 µL, typically with 2 drops per well and a final protein concentration of 4-10 mg mL^{-1}. The coverslip was then inverted over the well and sealed to the plate using a thin layer of high-vacuum grease (Dow Corning). Crystallisation was carried out in a 19 °C constant temperature room. Conditions producing crystalline precipitant were selected and components were varied in concentration and pH leading to improvements in crystal growth. The original screen compositions are listed in the Appendix.

Anaerobic crystallisation experiments were carried out in an anaerobic glove box using identical techniques and reagent volumes. However, experiments were set up in EasyXtal 15-well plates (Qiagen) which use plastic screw capped lids instead of glass and grease, and 3 reagent/protein drops per well. All plates and reagents were taken into the glove box 24 hours prior to use. Reagents were left open in the glove box for 2-3 hours to allow oxygen to diffuse out of the solutions while reducing the amount of evaporation as degassing all solutions was impractical. Co-crystallisation experiments were performed using a 10-fold excess of sirohaem to protein.

2.4.18.2 Crystal seeding experiments

To enhance the growth of crystals seeding experiments were performed. For this a ‘seed’ crystal was taken from a previous drop and transferred into 10 µL of mother liquor (well solution). The crystal was then crushed using a pipette tip before the addition of 90 µL mother liquor. The crystal was further broken down by vigorous vortexing and serially diluted in mother liquor to form the following seed stocks 10^0, 10^1, 10^2, 10^3 and 10^4. Crystallisation experiments were then performed as previously described using an altered drop formulation. For seeding experiments 1 µL protein was mixed with 0.3 µL seed stock and 0.7 µL well solution with 2 drops per well. These experiments were then incubated as previously described.

2.4.18.3 Tetrapyrrole soaking experiments

Native protein crystals were grown in an anaerobic chamber using the anaerobic crystallisation and seeding methods (Section 2.4.17.1 and Section 2.4.17.2). The crystal seed was grown aerobically before being transferred into the anaerobic chamber. Sirohaem was prepared using the method previously described in Section 2.4.8 and anaerobically freeze dried from a sufficient volume such that resuspension in well solution with 10% (v/v) glycerol would yield a 10 mM soaking solution. After crystals had grown to a sufficient size (~3-5 days) 1 µL sirohaem soaking solution was added to the drop. The crystals were incubated for 1 day in the sirohaem solution prior to being harvested and frozen (Section 2.4.17.5). Soaking of sirohaem into the crystals was confirmed by the crystal turning purple (previously colourless).

2.4.18.4 Preparation of heavy metal soaked crystals

In an attempt to solve the AhbA/B structure using anomalous scattering experiments heavy atom derivatives were produced using the quick-soak method as previously described (Sun *et al.*, 2002). Native protein crystals were selected and soaked in a range of heavy metal solutions (see Appendix) for 10 min to 2 hours before being harvested and frozen (Section 2.4.17.5).

2.4.18.5 Crystal harvest and storage methods

Crystals were removed from the drop and immersed in cryo-protectant [typically well solution with 10% (v/v) glycerol, although a range of glycerol was tested] using appropriate sized Litholoops™ (Molecular Dimensions Ltd.) or CyroLoops™ (Hampton Research). After cryo-protection the crystals were flash frozen in liquid nitrogen and subsequently stored under liquid nitrogen until data collection.

2.4.18.6 X-ray diffraction experiments

The native and SeMet datasets were collected on beamlines IO4-1 and IO2, respectively, at the Diamond Light Source (Oxfordshire). The tetrapyrrole soaked dataset was collected on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France). Structure determination and refinement statistics are described in Chapter 4.

2.4.19 Thermal shift assay

Protein stability was determined by thermal shift assay. This technique uses a fluorescent dye (SYPRO Orange®) that is quenched when free in solution but fluorescence increases upon binding to hydrophobic surfaces of proteins. Protein is incubated with a fluorescent dye and the temperature is increased stepwise. As the protein becomes denatured the dye binds and a fluorescent signal is produced. This allows the production of a melting curve and the melting temperature (T_m) can be determined. By using a range of buffers, salt concentrations and additives the ideal buffer conditions can be determined.

Initially, *D .desulfuricans* was incubated at a final concentration of 0.2 mg mL^{-1} in 100 mM NaCl with a range of buffers from pH 4.2 to 9.8 (see Appendix for list of buffers). The protein was heated from 25 to 85 °C in a 96 well PCR plate using an iCycler iQ™ Real-Time PCR Detection System (Bio-Rad) and fluorescence was measured. Data was normalised and melting points were determined with the aid of Dr. Stephen Irving (Argenta). Once an optimum buffer and pH was determined a second screen was performed to analyse the effect of salt concentration, glycerol and reducing agents using identical experimental conditions (see Appendix).

Chapter 3

Biochemical characterisation of AhbA and AhbB

3.1 Introduction

Evidence for the alternative haem biosynthesis pathway was revealed more than 20 years ago with the fortuitous discovery that the methyl groups of haem (at carbons C2 and C7) are derived from L-methionine (rather than ALA) in the sulphate reducing bacterium *D. vulgaris* (Akutsu *et al.*, 1993). Furthermore, labelled tetrapyrroles, including the novel intermediate 12, 18-didecarboxysirohydrochlorin, were isolated from *D. vulgaris* that indicated the alternative pathway proceeded via precorrin-2 (Ishida *et al.*, 1998). Incorporation of methionine into haem was also demonstrated in the archaeon *M .barkeri* (Buchenau *et al.*, 2006), suggesting a variety of organisms may synthesise haem using this pathway via precorrin-2. With the release of many sequenced genomes, candidates for the alternative haem biosynthesis pathway were discovered in *D. vulgaris* and the archaea (Lobo *et al.*, 2009, Storbeck *et al.*, 2010). Due to sequence similarity the candidate genes were annotated as members of the *nir* family (*nirD*, *H* and *J*). NirDLGH are involved in haem d_1 synthesis (Fig. 3.1). However, *D. vulgaris* and many of the archaea do not require haem d_1 as they do not contain the genes encoding the cytochrome cd_1 nitrite reductase (NirS). The *nirD/H/J* homologues were therefore suspected to be involved in haem synthesis. The gene products of *nirD/H/J1/J2* from *D. vulgaris* were shown to convert sirohaem to haem, constituting the alternative haem biosynthesis pathway [Fig. 3.1; (Bali *et al.*, 2011)]. As such the genes were re-annotated *ahb* (**a**lternative **h**aem **b**iosynthesis) *A*, *B*, *C* and *D*.

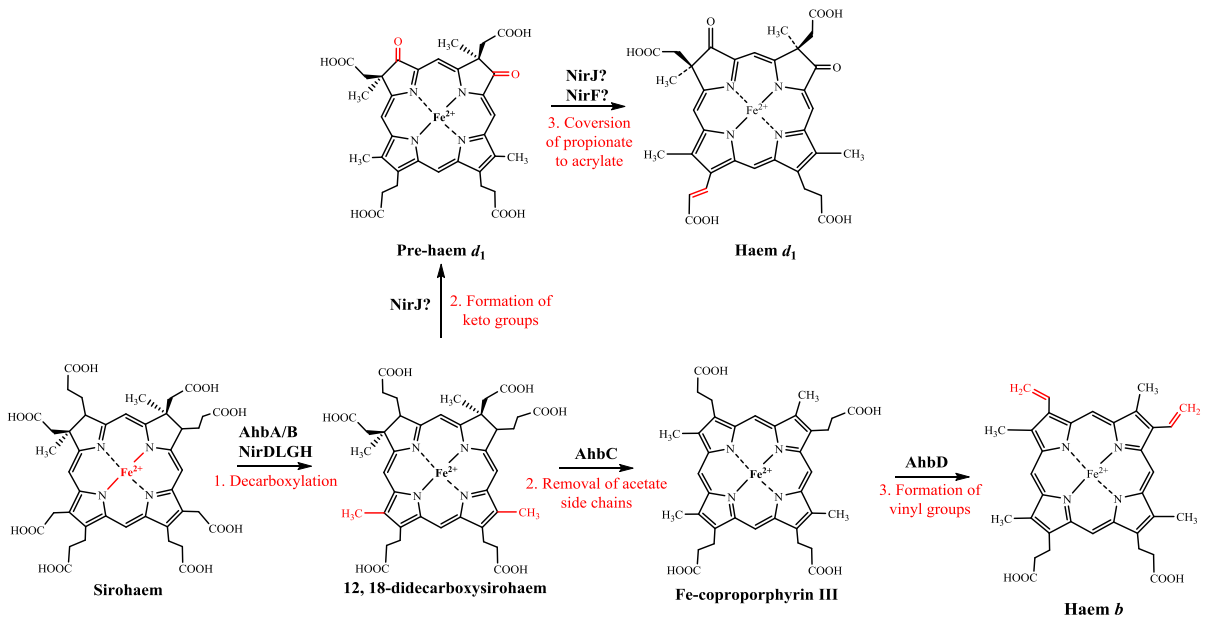


Figure 3.1. The alternative haem and proposed haem d_1 pathways. Chemical modifications are highlighted in red.

The activity of AhbA and AhbB as sirohaem decarboxylases was confirmed when recombinant *E. coli* cell lysates containing either *D. vulgaris* or *D. desulfuricans* AhbA and AhbB were shown to quantitatively convert sirohaem into didecarboxysirohaem (Bali *et al.*, 2011). Subsequently, incubations with purified AhbA and AhbB also resulted in the complete transformation of sirohaem to didecarboxysirohaem (Bali *et al.*, 2011). These results were unexpected for two reasons. Firstly, the substrate contained iron chelated in the macrocycle, which was expected to be inserted in the final stage of the pathway, analogous to the classical pathway. Secondly, the same study also found that this initial reaction is conserved in the haem d_1 pathway catalysed by the enzymes NirDLGH. AhbC from *M. barkeri* was shown to be active using *E. coli* cell lysates containing the enzyme and purified didecarboxysirohaem, leading to the production of Fe-coproporphyrin III. Finally purified AhbD from *D. vulgaris* demonstrated turnover of Fe-coproporphyrin III to haem.

The decarboxylations of the C12 and C18 acetic acid side chains carried out by AhbA and AhbB are analogous to two of the decarboxylations catalysed by HemE (UROD) from the classical haem synthesis pathway (Whitby, 1998). HemE catalyses the removal of the four acetate side chains of uro'gen III at carbons C2, C7, C12 and C18 in a cofactor independent fashion (Dailey, 2002). Analysis of the non-enzymatic decarboxylation of a substrate model (pyrrolyl-3-acetate), suggests that the spontaneous reaction proceeds with a half-life of $2.3x10^9$ years (Lewis & Wolfenden, 2008). A comparison of the non-enzymatic reaction to the enzyme catalysed reaction shows HemE enhances the rate of substrate decarboxylation by a factor of 1.2×10^{17}, one of the largest rate enhancements reported for an enzyme that acts independently of cofactors, leading to HemE being described as a 'benchmark for the catalytic proficiency of enzymes'. In the enzyme catalysed reaction the pyrrole nitrogens are used as electron sinks, with a protonated basic residue (Arg37 in human HemE) aiding carboxylate removal (Silva *et al.*, 2010). Like HemE, AhbA and AhbB decarboxylate a kinetically stable tetrapyrrole (sirohaem) independently of any cofactors or metal ions (Bali *et al.*, 2011). Therefore, the further understanding of these enzymes is of great interest.

The work described in this chapter was undertaken in order to gain further insights into the AhbA and AhbB proteins. A comparative analysis of the proteins from three different organisms: *D. desulfuricans, D. vulgaris* and *M. barkeri* was performed. Experiments were carried out to determine if the individual proteins are capable of both decarboxylation reactions and whether the order of side chain decarboxylation could be ascertained. The formation of a quaternary AhbA/B complex was also investigated, revealing cross-species variation in oligomeric states, haem binding capabilities and redox regulation of the complexes. Methods were developed to allow the determination of kinetic parameters for the

double decarboxylation reaction. Finally, sites of decarboxylation were confirmed by NMR analysis of the AhbA/B reaction product didecarboxysirohaem.

3.2 Results

3.2.1 Overproduction and purification of AhbA and AhbB

The *ahbA* and *ahbB* genes from *D. desulfuricans* and *D. vulgaris* were amplified by PCR from template DNA (pSL201-4) supplied by Dr. Susana Lobo (ITQB, Lisbon), whereas the *M. barkeri ahbA* and *ahbB* genes were amplified from genomic DNA provided by Dr. Evelyne Deery (University of Kent). The genes were cloned individually into the vector pET14b using *NdeI* and *SpeI* restriction sites (Section 2.3.7; Table 2.2). The pET14b vector encodes an N-terminal His_6-tag allowing purification of the fusion protein by nickel affinity chromatography.

In order to overproduce the individual AhbA and AhbB proteins *E. coli* BL21 (DE3) was transformed with a vector containing the gene of interest (*i.e.* pET14b-*ahbA* or pET14b-*ahbB*). The resulting strain was grown in LB media supplemented with ampicillin and constant shaking (~160 rpm) at 37 °C to a cell density of $A_{600} \approx 0.6$. Protein overproduction was induced with the addition of 0.4 mM IPTG and the cells were left to grow overnight at 19 °C before harvesting by centrifugation (Section 2.2.7). After harvesting by centrifugation the cell pellets were resuspended in Tris buffer and lysed by sonication (Section 2.2.9). The soluble proteins were removed from cell debris by centrifugation. Protein purification was achieved using nickel affinity chromatography. Briefly, the resulting supernatant was passed over a nickel charged sepharose column and non-specific proteins were removed by the addition of Tris buffer containing a low concentration of imidazole. Purified protein was eluted in 400 mM imidazole in 2 ml fractions before buffer exchange using a PD10 column.

M. barkeri AhbA was insoluble and remained in the cell pellet after sonication and centrifugation (Fig. 3.2). The AhbA proteins from *D. desulfuricans* and *D. vulgaris* were soluble; however, a high level of precipitation occurred while the proteins were immobilised on the nickel column. Low yields of purified *D. desulfuricans* and *D. vulgaris* AhbA proteins were obtained (~ 4 mg L^{-1} each). After purification both proteins continued to precipitate over time. The presence of the purified proteins was confirmed using SDS-PAGE, although both aggregation and degradation of the proteins can be observed (Fig. 3.2).

As with AhbA, the *M. barkeri* AhbB protein was insoluble and the majority remained in the cell pellet (Fig. 3.2). Attempts to purify the *M. barkeri* AhbB protein from the soluble fraction did not realise any protein. The AhbB proteins from *D. desulfuricans* and *D. vulgaris* were soluble, but precipitated on the nickel column in an identical fashion to AhbA,

and continued to precipitate after purification. Slightly higher yields of the AhbB proteins were obtained in comparison to the AhbAs, with ~ 13 mg L^{-1} of the *D. desulfuricans* AhbB produced and ~ 7 mg L^{-1} of the *D. vulgaris* homologue. Again, confirmation of the purified proteins was obtained using SDS-PAGE, although fewer bands relating to aggregation states were observed for these proteins (Fig. 3.2). Due to the instability of the individual AhbA and AhbB proteins, no gel filtration studies were performed.

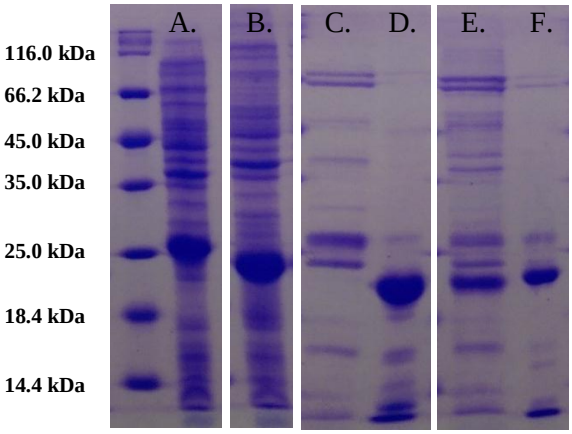


Figure 3.2. 12.5% (v/v) SDS gels of AhbA and AhbB proteins. A. Cell pellet containing *M. barkeri* AhbA (23 kDa), **B.** Cell pellet containing *M. barkeri* AhbB (20 kDa), **C.** Purified *D. desulfuricans* AhbA (21 kDa), **D.** Purified *D. desulfuricans* AhbB (20 kDa), **E.** Purified *D. vulgaris* AhbA (22 kDa), **F.** Purified *D. vulgaris* AhbB (21 kDa). Both aggregation and degradation of the purified protein can be observed.

3.2.2 *In vitro* activity of AhbA and AhbB

3.2.2.1 Production of substrate

Sirohaem was produced and purified as outlined in Section 2.4.8. In brief, the enzymes ALA dehydratase (HemB), porphobilinogen deaminase (HemC), uro'gen III synthase (HemD), SUMT (CobA) and precorrin-2 dehydrogenase (CobA) were purified aerobically and buffer exchanged into anaerobic buffer inside an anaerobic chamber (<2 ppm O_2). These enzymes were then incubated overnight at room temperature with 1 mL cofactor mix which contained SAM, NAD and ALA. Sirohaem was produced after the addition of a 10-fold excess of $FeSO_4$ to the sirohydrochlorin and incubation at room temperature for 4 hours. In the presence of excess iron chelation into the macrocycle occurs without the need for chelatase enzymes. Sirohaem was purified by ion exchange chromatography using a DEAE column. The column was washed with the stepwise addition of NaCl from 100 mM to 800 mM, and finally sirohaem was eluted in 1 M NaCl. All buffers contained Tris-HCl, pH 8, to maintain

the pH of the solutions. DEAE purification of sirohaem resulted in a final volume of ~7 mL and at a concentration of ~60 μM.

3.2.2.2 *In vitro* assays of individual AhbA and AhbB from *D. desulfuricans* and *D. vulgaris*

To test *in vitro* activity of the individual proteins sirohaem (25 μM) was incubated overnight (~16 hours) with an equimolar amount of the protein of interest (*i.e* an AhbA or AhbB protein). After incubation the remaining soluble protein was precipitated by the addition of an equal volume of 100% acetonitrile at 4 °C for 10 min. The precipitate was removed from the sample by centrifugation at 2000 x *g* prior to analysis. Sirohaem and its derivatives were analysed by HPLC-MS using an ACE 5AQ column as described in Section 2.4.14. Using the sirohaem derivative gradient on the HPLC sirohaem is eluted first (~16 min), followed by monodecarboxysirohaem (~19.5 min) and finally didecarboxysirohaem (~25 min).

HPLC chromatograms of the individual *in vitro* assays are shown in Figure 3.3. *D. desulfuricans* AhbA displayed a significant ability to convert sirohaem into didecarboxysirohaem (85%) with a small quantity of monodecarboxysirohaem remaining (10%). AhbA from *D. vulgaris* also demonstrated the ability to carry out both decarboxylation reactions. Although a lower quantity of didecarboxysirohaem was present (44%), a higher quantity of monodecarboxysirohaem remained (19%).

D. desulfuricans AhbB was capable of converting the majority of sirohaem to didecarboxysirohaem (58%) and a high level of monodecarboxysirohaem (27%). However, *D. vulgaris* AhbB displayed very little activity for both decarboxylation reactions with only 16% monodecarboxysirohaem present and a miniscule 1% didecarboxysirohaem.

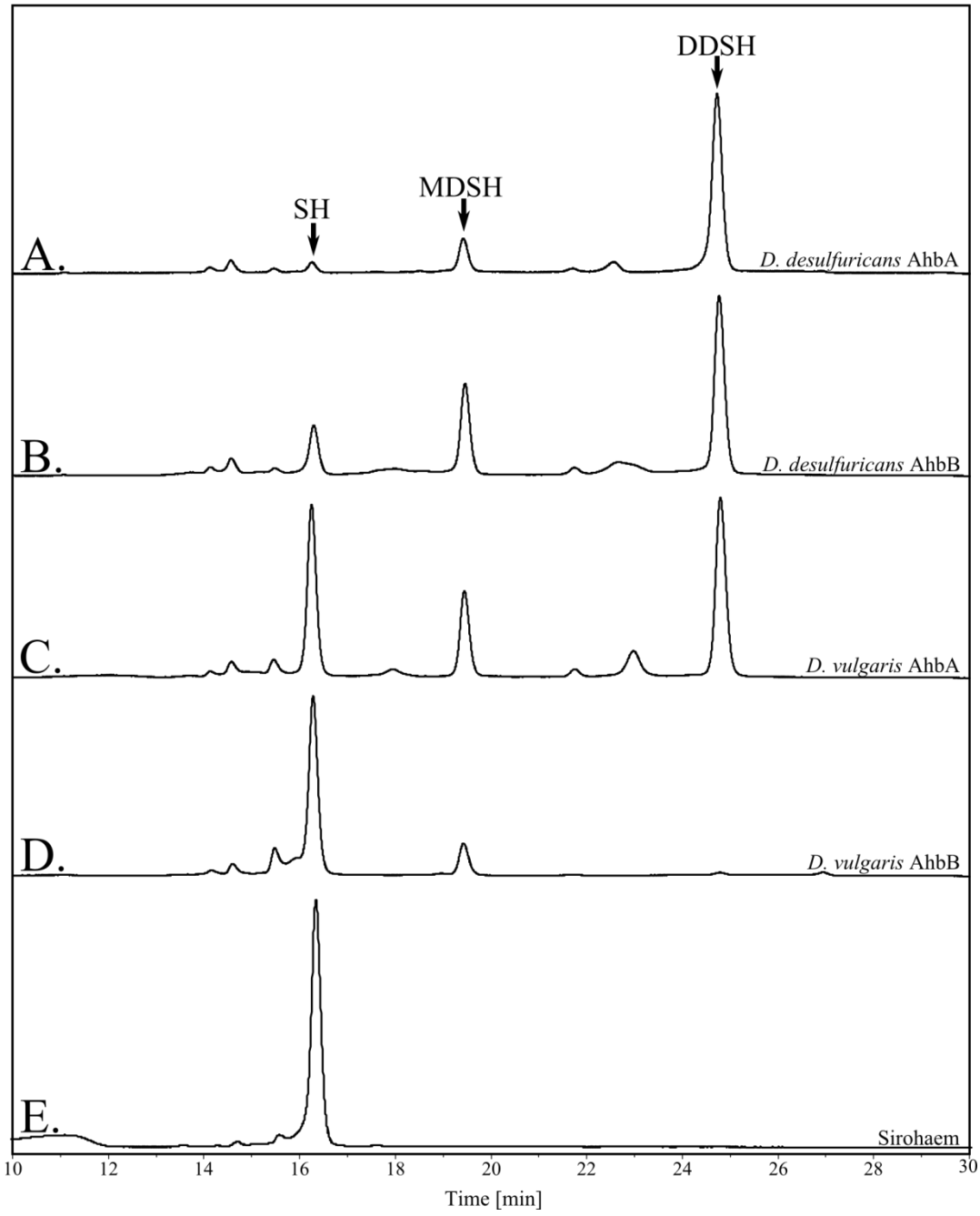


Figure 3.3. Reaction profiles of individual AhbA and AhbB proteins. HPLC traces, recorded at a wavelength of 380 nm, after incubation of sirohaem with purified proteins. **A.** *D. desulfuricans* AhbA, **B.** *D. desulfuricans* AhbB, **C.** *D. vulgaris* AhbA, and **D.** *D. vulgaris* AhbB. **E.** shows a no enzyme control. The arrows indicate the peaks relating to sirohaem (SH, ~16 min), monodecarboxysirohaem (MDSH, ~19.5 min) and didecarboxysirohaem (DDSH, ~25 min).

The problem with this method is that, due to the instability of the proteins, it is unclear whether this is a true representation of activity or an anomaly caused by the differing levels of soluble protein in each reaction, *i.e.* AhbB may have denatured quicker than the other proteins and therefore displayed lower activity. It does however represent the first use of the AhbA and B proteins in an *in vitro* assay and demonstrates that all of the proteins tested are capable of double decarboxylations of sirohaem to varying degrees. Unfortunately, it was not possible to separate monodecarboxysirohaem from the solutions so the order of decarboxylation could not be ascertained.

To test if the combination of AhbA and AhbB produced an enhanced reaction the activity assays were repeated with a 50:50 mix of the two proteins (from a single organism). In both cases (*i.e.* using AhbA and AhbB proteins from *D. desulfuricans* or *D. vulgaris*), incubations of sirohaem with the two proteins resulted in almost complete conversion of sirohaem to didecarboxysirohaem (Fig. 3.4). The increase in activity observed when both AhbA and AhbB are present in the reaction demonstrates that they are capable of acting cooperatively, facilitating the double decarboxylation of sirohaem.

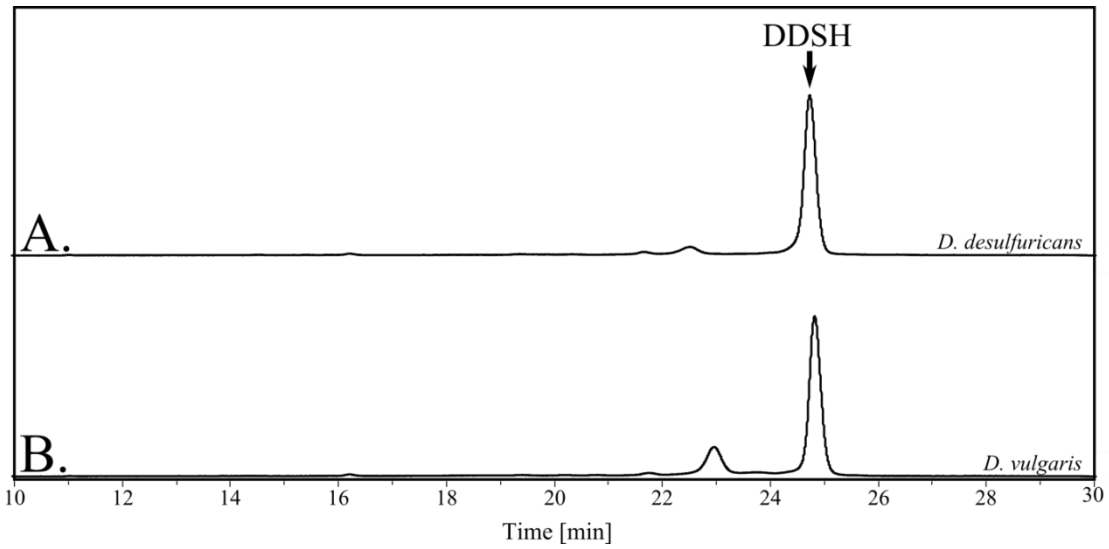


Figure 3.4. Reaction profiles of combined AhbA and AhbB proteins. HPLC traces, recorded at a wavelength of 380 nm, after incubation of sirohaem with purified proteins. **A.** *D. desulfuricans* AhbA and AhbB and **B.** *D. vulgaris* AhbA and AhbB. The arrow indicates the peak relating to didecarboxysirohaem (DDSH, ~25 min).

3.2.3 Overproduction and purification of the AhbA/B complex

The high level of activity observed when the AhbA and AhbB proteins were combined led to the hypothesis that they may interact and that this may also stabilise the proteins. It has been reported previously that in most archaea harbouring these genes, *ahbA* and *ahbB* are found as a fusion (Storbeck *et al.*, 2010), consistent with the idea that the two proteins may form a complex together.

A construct was made containing *D. desulfuricans ahbA* and *ahbB* as described in Section 2.3.7, allowing overexpression of both genes from a single vector. Using the Link and Lock technique a fragment containing a ribosome binding site followed by *ahbB* fused to an N-terminal His_6-tag was inserted into pET3a downstream of *ahbA* to form a single vector encoding the two *ahb* genes with independent ribosome binding sites (Fig. 2.2). Using the *D. desulfuricans ahbA/B* Link and Lock construct, a pull-down assay was performed. As only AhbB is produced with an N-terminal His_6-tag using the *ahbA/B* plasmid, AhbA should not interact with the nickel column and should therefore be removed in the wash steps of purification. However, if AhbA forms a complex with AhbB, both proteins should remain bound to the nickel column and be eluted in the final stage of purification. The AhbA/B proteins were overproduced and purified as described for the individual enzymes (Section 3.2.2.1).

A high yield (~ 53 mg L^{-1}) of colourless protein was purified from *E. coli* transformed with the *D. desulfuricans* pET3a-*ahbA/B* vector. SDS-PAGE analysis showed two individual bands with estimated masses corresponding to AhbA (~19 kDa) and AhbB with a His_6-tag (~20 kDa; Fig. 3.5). This demonstrated that AhbA and AhbB from *D. desulfuricans* co-purify and form a stable heteromeric complex. Gel filtration studies revealed that the AhbA/B complex from *D. desulfuricans* forms a single heterodimeric species with a mass of ~40 kDa which correlates closely with the expected mass of 39.2 kDa (Fig. 3.6). Native polyacrylamide gels were attempted with this complex (data not shown). However, excessive smearing was observed and bands were not clearly resolved, therefore this technique was not repeated.

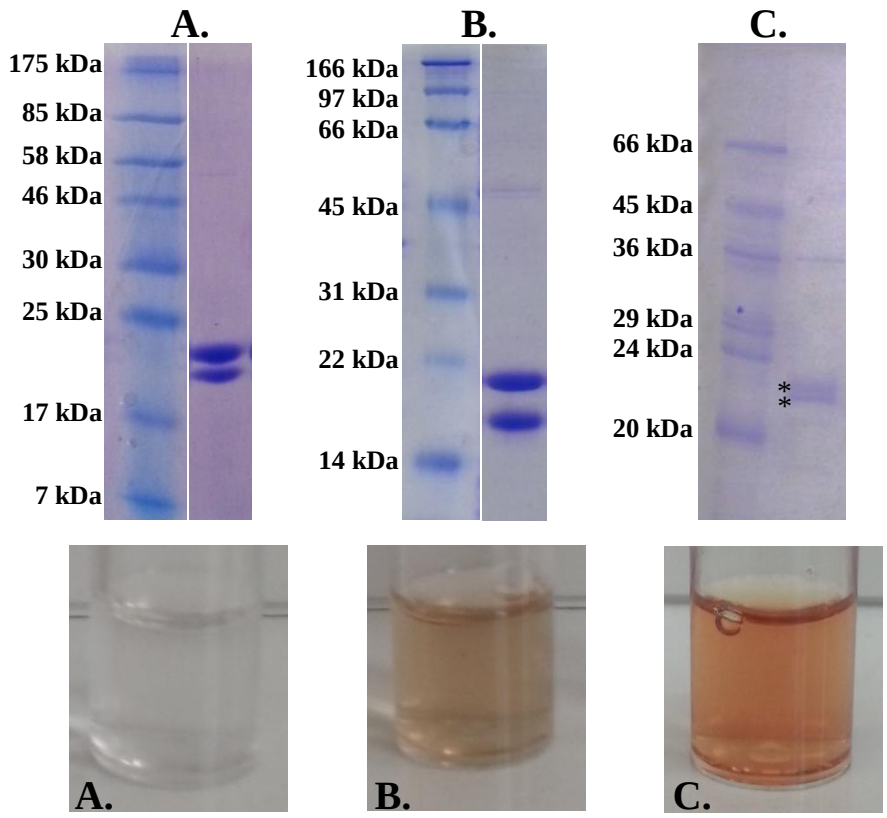


Figure 3.5. 12.5% SDS gels of purified AhbA/B complexes and pictures of the purified protein solutions. The AhbA/B complexes from *D. desulfuricans* (**A.**), *D. vulgaris* (**B.**) and *M. barkeri* (**C.**) were purified using the pull-down method. In each case the higher band in the SDS gel is AhbB and the lower is AhbA (individual *M. barkeri* protein bands are labelled with asterisks).

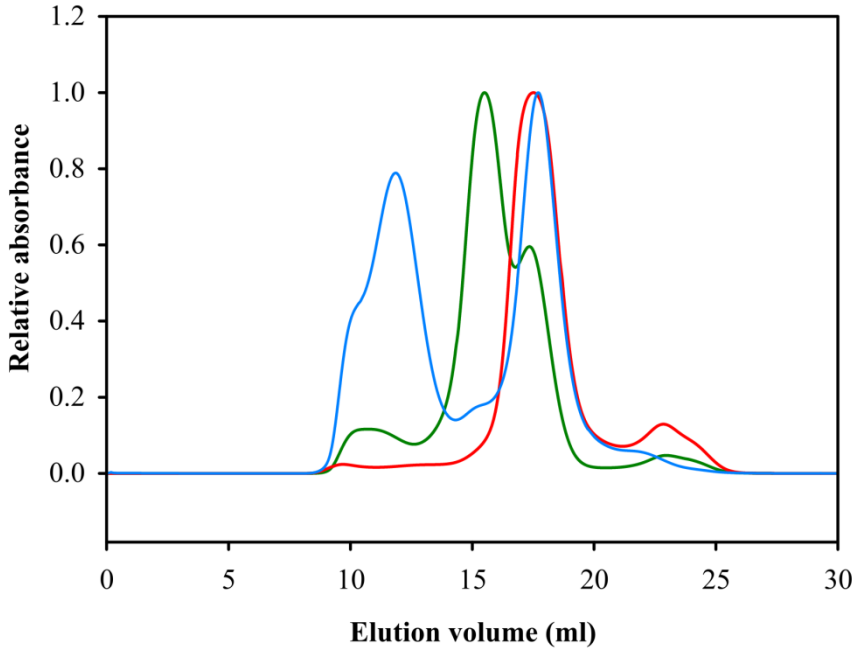


Figure 3.6. Gel filtration UV trace recorded at a wavelength of 280 nm of purified AhbA/B complexes. Major peaks correspond to a single dimeric *D. desulfuricans* AhbA/B complex (red), dimeric and tetrameric *D. vulgaris* AhbA/B complexes (green), and a dimeric *M. barkeri* AhbA/B complex (blue) with high molecular weight aggregates. Peaks after 20 minutes are low molecular weight contaminants.

In light of the discovery of the *D. desulfuricans* AhbA/B complex the *ahbA/B* genes from *D. vulgaris* were cloned into a single vector in an identical manner to those from *D. desulfuricans* using the Link and Lock technique, with only AhbB fused to an N-terminal His_6-tag. Again, co-expression of the two genes yielded much higher quantities of stable protein (~ 79 mg L^{-1}) after purification by nickel affinity chromatography. The co-purification of *D. vulgaris* AhbA/B was confirmed by SDS-PAGE (Fig. 3.5). Surprisingly, gel filtration of the AhbA/B complex from *D. vulgaris* revealed the presence of a dimeric complex as well as a tetrameric complex with masses of ~44 kDa and ~94 kDa, respectively, close to the estimated masses of 40 kDa and 81 kDa (including the His_6-tag; Fig. 3.6). Interestingly, *D. vulgaris* AhbA/B purified with a light brown colour in contrast to the colourless isolated *D. desulfuricans* protein complex (Fig. 3.5).

UV-visible analysis of the *D. vulgaris* AhbA/B complex revealed an absorption peak at 420 nm, suggesting the presence of either an Fe-S complex or a haem group (Fig. 3.7). However, analysis of the protein sequences showed no CX_3CX_2C motif which is conserved in many iron sulphur centre containing proteins (Sofia *et al.*, 2001). Furthermore, EPR analysis of the

protein complex did not show any signal consistent with an iron sulphur centre (data not shown). It was, therefore, postulated that the absorption at 420 nm may be due to a bound haem group in low occupancy.

To investigate this further haemin (6 mM) was added to the resuspended cell pellet of the strain overproducing the *D. vulgaris* AhbA/B prior to sonication in an attempt to load the complex with the prosthetic group. This approach resulted in purified protein with a significantly stronger brown/orange colouration. The UV-visible spectrum of haem-loaded *D. vulgaris* AhbA/B had a Soret band at 423 nm and αβ bands at 527 nm and 555 nm (reduced spectrum; Fig. 3.7). The haem was retained with the protein during gel filtration and did not affect the oligomeric state of the complex (data not shown). Supplementation of cultures with ALA (20 mg L^{-1}) prior to induction of protein production also led to an increased colouration of the purified complex and a subsequent increase in the A_{420} (data not shown). However, the haem occupancy using ALA supplementation was over 50% lower than that produced using haemin addition.

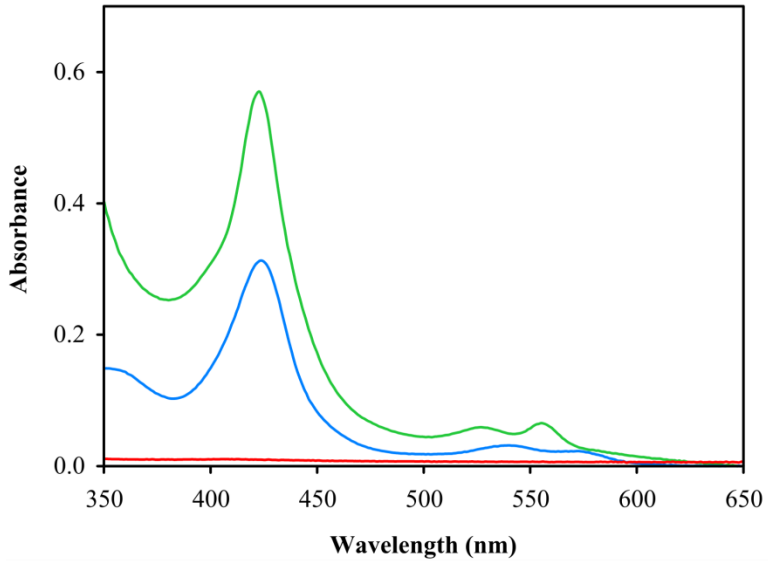


Figure 3.7. UV-visible spectra of purified sirohaem decarboxylases. The spectra of *M. barkeri* AhbA/B (green), *D. vulgaris* AhbA/B (blue), and *D. desulfuricans* AhbA/B (red) are shown.

The observation that co-expression of AhbA and AhbB from *D. desulfuricans* and *D. vulgaris* causes enhanced stability and co-purification of the proteins led to the investigation as to whether this would be the case for the homologues from *M. barkeri* which, as mentioned earlier, had been found to be highly insoluble when produced individually. As with the genes from the sulphate reducers, a single plasmid construct was produced using the

Link and Lock technique, containing *M. barkeri ahbA* and *ahbB* with the latter encoding an N-terminal His_6-tag. The AhbA/B proteins were overproduced in *E. coli* and a pull-down assay was used to test complex formation.

Surprisingly, a distinct red band was observed during the nickel affinity purification. The colour eluted from the column with the protein. The *M. barkeri* AhbA/B complex was purified at a slightly lower yield of ~31 mg L^{-1}. SDS-PAGE analysis showed that both proteins co-purified (Fig. 3.5). The UV-visible spectra of the purified protein solution had a distinctive haem profile with peaks at 426 nm, 530 nm, and 559 nm (Fig. 3.7). The haem-loaded *D. vulgaris* AhbA/B haem UV-visible spectrum is slightly shifted in comparison to the *M. barkeri* AhbA/B. The *M. barkeri* complex displays an intense haem signal without the addition of haemin, suggesting it has a higher affinity for haem. Size exclusion chromatography of *M. barkeri* AhbA/B revealed it forms a single dimeric species with an observed mass of 37 kDa (estimated mass 40 kDa including His_6-tag) as well as high molecular weight aggregates (Fig. 3.6).

In order to confirm the presence of haem in the *D. vulgaris* and *M. barkeri* complexes and to ascertain the type of haem bound (*a/o, b,* or *c*), pyridine haemochrome assays were undertaken [Section 2.4.11; (Berry & Trumpower, 1987)]. These assays were performed with both the *M. barkeri* AhbA/B and the exogenous haem-loaded *D. vulgaris* AhbA/B proteins. First the haemichromes were formed by the addition of 0.5 mL 200 mM NaOH, 40% (v/v) pyridine to 0.5 mL protein (~10 mg mL^{-1}), oxidised with 3 µL 0.1 M potassium ferricyanide, and the spectra were recorded from 520-620 nm. Next the haemochromes were produced by the reduction of the haemichromes with 2-5 mg sodium dithionite and the spectra were recorded at 520-620 nm. The difference spectrum of haemochrome (reduced) minus haemichrome (oxidised) relates to the type of haem present in the sample. The pyridine haemochrome from *M. barkeri* displayed a peak at 556 nm (Fig. 3.8), which corresponds to a *b* type haem (Berry & Trumpower, 1987). The haem-loaded complex from *D. vulgaris* again showed a slightly shifted spectrum with a haemochrome peak at 554 nm, shorter than that reported for haem *b*, but longer than that of a haem *c* (≈ 550 nm; Fig. 3.8).

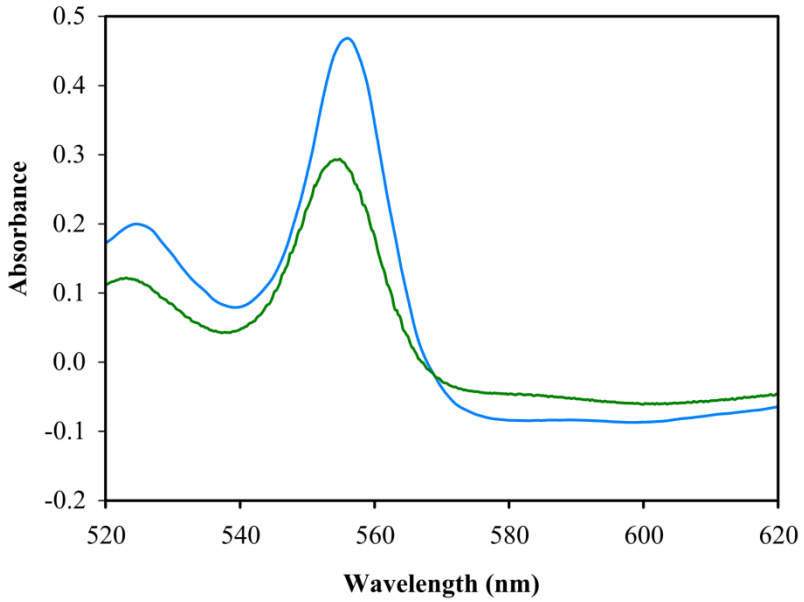


Figure 3.8. Pyridine haemochrome difference spectra (reduced-oxidised) of *M. barkeri* and *D. vulgaris* AhbA/B complexes. Exogenous haem-loaded *D. vulgaris* AhbA/B (green) and *M. barkeri* AhbA/B (blue). Peaks are observed at 554 nm and 556 nm, respectively.

To obtain further confirmation of the presence of haem, extraction methods were employed for analysis of the compound by mass spectrometry. Haem extractions were carried out as previously described [Section 2.4.12; (Lubben & Morand, 1994)]. Using acidified acetone, non-covalently bound haem molecules can be extracted from membrane suspensions and protein solutions. Addition of ice-cold dH_2O and ethyl acetate causes the tetrapyrrole to separate into the organic layer which can then be recovered and the solvent removed using a vacuum centrifuge. The residues can then be analysed by mass spectrometry after resuspension in dH_2O (Section 2.4.14.1).

Low levels of haem could be removed from the *D. vulgaris* AhbA/B complex using haem extraction methods. For this experiment the haem content of *D. vulgaris* AhbA/B was increased by adding ALA (20 mg L^{-1}) to the growth media of the transformed *E. coli* at the point of induction of protein overproduction. The presence of the extracted haem was confirmed by mass spectrometry (Fig. 3.9); however, the majority of the haem remained bound to the denatured protein. Samples were resolved on an ACE 5AQ column with a gradient of 0.1% (v/v) TFA and acetonitrile. The gradient used for analysis of haem is different to that used for the separation of sirohaem and its derivatives due to the increased hydrophobicity of haem. The haem gradient starts with the column equilibrated in 20% acetonitrile allowing earlier elution of haem at ~16.5 min.

Protein bound haem could not be detected on an SDS gel using a haem staining protocol (Section 2.4.6.3) suggesting that it is likely to be a haem *b* bound to *D. vulgaris* AhbA/B. Further attempts to remove the haem using urea and SDS were not successful with the haem remaining bound to the protein. The tight binding of haem to the enzyme complex may explain the shift observed in the UV-visible spectra of the proteins and their pyridine haemochrome derivatives. In contrast, it was much easier to remove the haem from the *M. barkeri* AhbA/B complex using heat or urea (Fig. 3.9). Dimer formation and haem binding properties of the *M. barkeri* AhbA/B was recently published confirming the results reported in this thesis (Kuhner *et al.*, 2014).

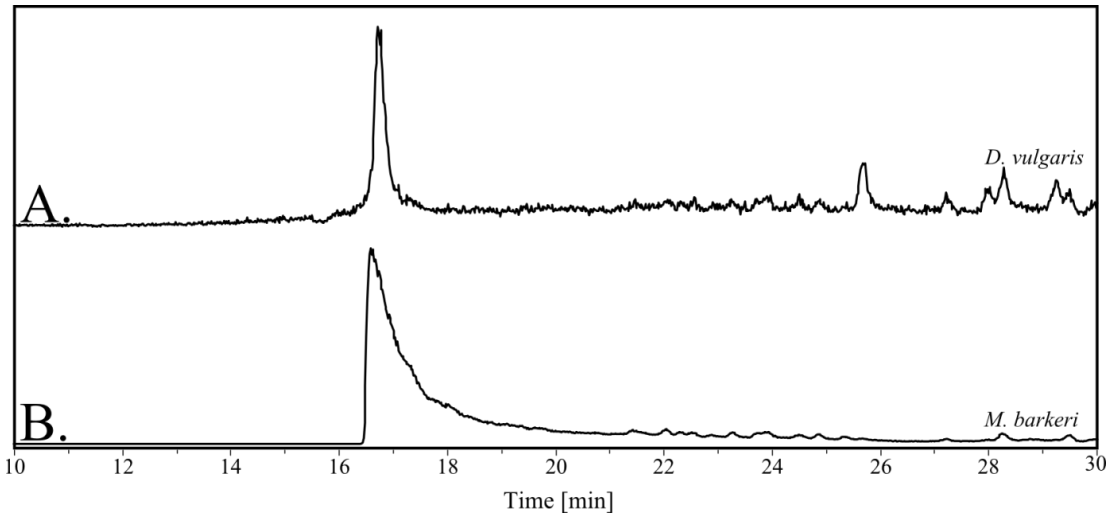


Figure 3.9. Haem extractions from *D. vulgaris* and *M. barkeri* AhbA/B proteins. HPLC traces, recorded at a wavelength of 420 nm, after haem extractions from purified *D. vulgaris* AhbA/B (**A.**) and *M. barkeri* AhbA/B (**B.**). The peaks at ~17 minutes relate to haem with a *m/z* of 616. For this experiment haem content of *D. vulgaris* AhbA/B complex was increased by adding ALA to *E. coli* growth media during protein overproduction.

3.2.4 *In vitro* activity of the AhbA/B complex

To demonstrate that the AhbA/B proteins are active as a complex they were incubated with a ten-fold excess of purified sirohaem (Section 2.4.8) overnight (≈ 16 hours) at 37 °C in an anaerobic chamber (<2 ppm O_2). Protein was precipitated by addition of acetonitrile and centrifugation before samples were analysed using HPLC-MS. Both AhbA/B complexes from sulphate reducing bacteria showed quantitative conversion of sirohaem to didecarboxysirohaem (Fig. 3.10). The haem-loaded *D. vulgaris* AhbA/B complex was

equally as active as the apo-protein suggesting that the presence of the haem does not affect the activity of the enzyme from this organism (Fig. 3.11).

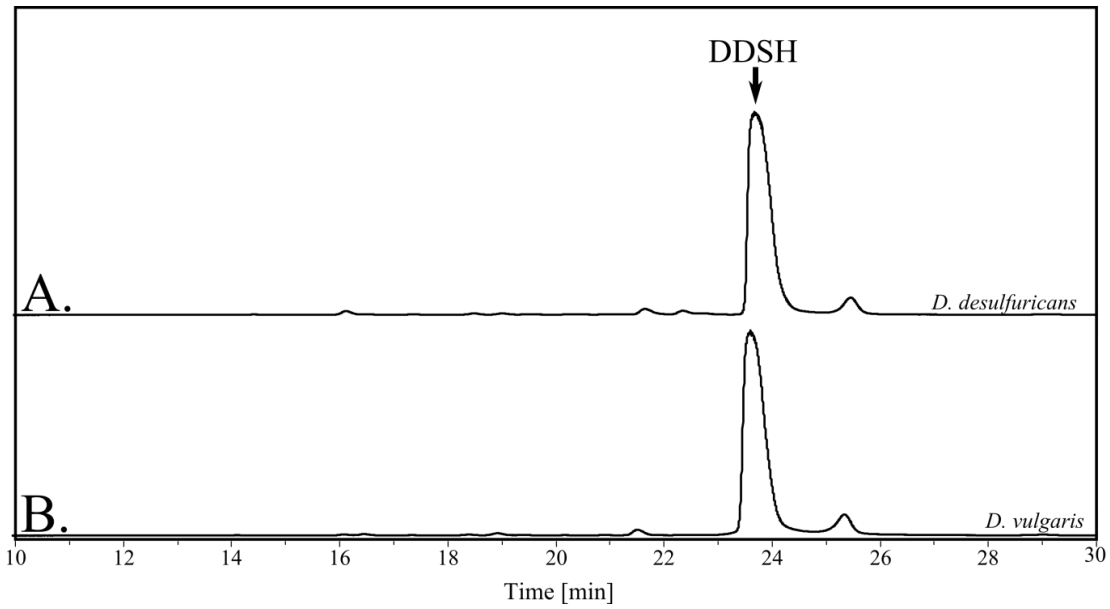


Figure 3.10. Reaction profiles of AhbA/B complexes from sulphate reducing bacteria. HPLC traces, recorded at a wavelength of 380 nm, after incubation of 10-fold excess of purified sirohaem with purified **A.** *D. desulfuricans* AhbA/B, **B.** *D. vulgaris* AhbA/B. The arrow indicates the peak relating to didecarboxysirohaem (DDSH, ~24 min).

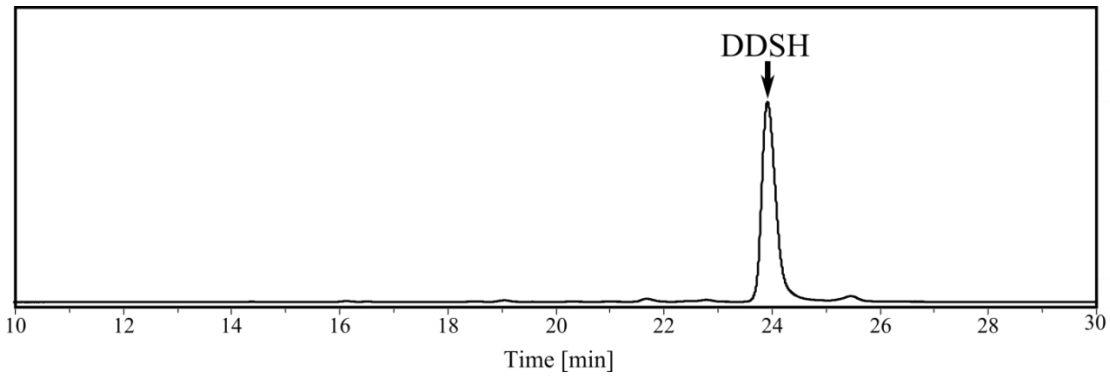


Figure 3.11. Reaction profile of haem-loaded AhbA/B complex from *D. vulgaris*. HPLC trace, recorded at a wavelength of 380 nm, after incubation of 10-fold excess of purified sirohaem with haem-loaded AhbA/B complex from *D. vulgaris*. The arrow indicates the peak relating to didecarboxysirohaem (DDSH, ~24 min).

Surprisingly, similar incubations of *M. barkeri* AhbA/B with sirohaem showed low turnover (Fig. 3.12). Repeating the assay in the presence of a ten-fold excess of sodium dithionite gave a 6.5-fold increase in activity (quantified using HPLC peak integration), indicating that the haem bound to the AhbA/B complex must be reduced in order for the enzyme to be most active. To investigate this further the reaction was repeated with AhbA/B protein that had been oxidised by exposure to potassium ferricyanide prior to buffer exchange. Again low activity was observed with the oxidised protein, which could be recovered by the addition of sodium dithionite (Fig. 3.12). These data demonstrate that the redox state of the haem *b* in the *M. barkeri* AhbA/B modulates the activity of the enzyme. Addition of sodium dithionite to the haem-loaded *D. vulgaris* AhbA/B complex did not affect activity (data not shown). Also, a high degree of activity was displayed using aerobically purified haem-loaded protein (Fig. 3.11), suggesting that the oxidation state of the haem in this complex does not regulate activity.

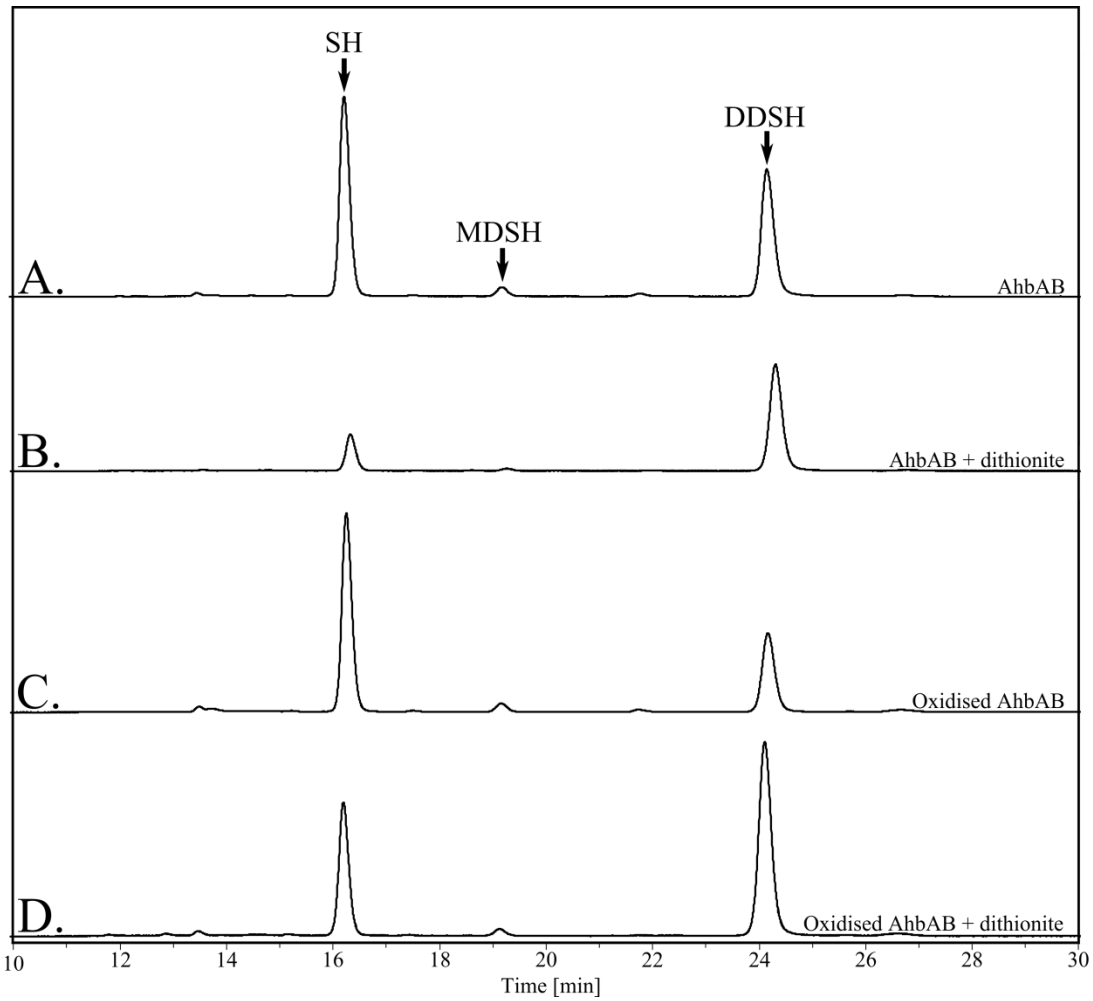


Figure 3.12. Reaction profile of AhbA/B complex from *M. barkeri* in different oxidation states. HPLC traces, recorded at a wavelength of 380 nm, after incubation of 10-fold excess of purified sirohaem with **A.** purified protein, **B.** purified protein with a 10 fold excess of sodium dithionite, **C.** protein oxidised with addition of a 10 fold excess ferricyanide which was removed prior to reaction by buffer exchange, **D.** the same oxidised protein with a 10 fold excess of sodium dithionite. The arrows indicate the peaks relating to sirohaem (SH, ~16 min), monodecarboxysirohaem (MDSH, ~19.5 min) and didecarboxysirohaem (DDSH, ~25 min).

3.2.5 Redox titration of *M. barkeri* AhbA/B

Redox potentiometry of the *M. barkeri* AhbA/B complex was carried out by Dr. Kirsty McLean (University of Manchester). The *M. barkeri* complex was purified using the pull-down method as previously described (Section 3.2.3). The purified protein was buffer exchanged in an anaerobic chamber (<2 ppm oxygen) into degassed phosphate buffer (100 mM), pH 7.0, containing 200mM KCl and 10% (v/v) glycerol. Subsequently, the proteins were reversibly reduced and oxidised using sodium dithionite as a reducing agent and potassium ferricyanide for re-oxidation (Section 2.4.16). Mediators (phenazine methosulphate, 2-hydroxy-1,4-naphthoquinone, methyl viologen and benzyl viologen) were included to mediate the range between +100 to -480 mV. Spectra were recorded at different applied potentials. The maximal overall absorption changes between oxidised and reduced species were plotted versus the applied potential (mV versus the normal hydrogen electrode, NHE). Data were fitted using the Nernst equation for a single electron reduction, yielding the respective midpoint reduction potentials. The proteins remained stable throughout the titrations. A redox potential of -98 ± 3 mV at 426 nm was observed for the haem bound to the *M. barkeri* AhbA/B complex (Fig. 3.13).

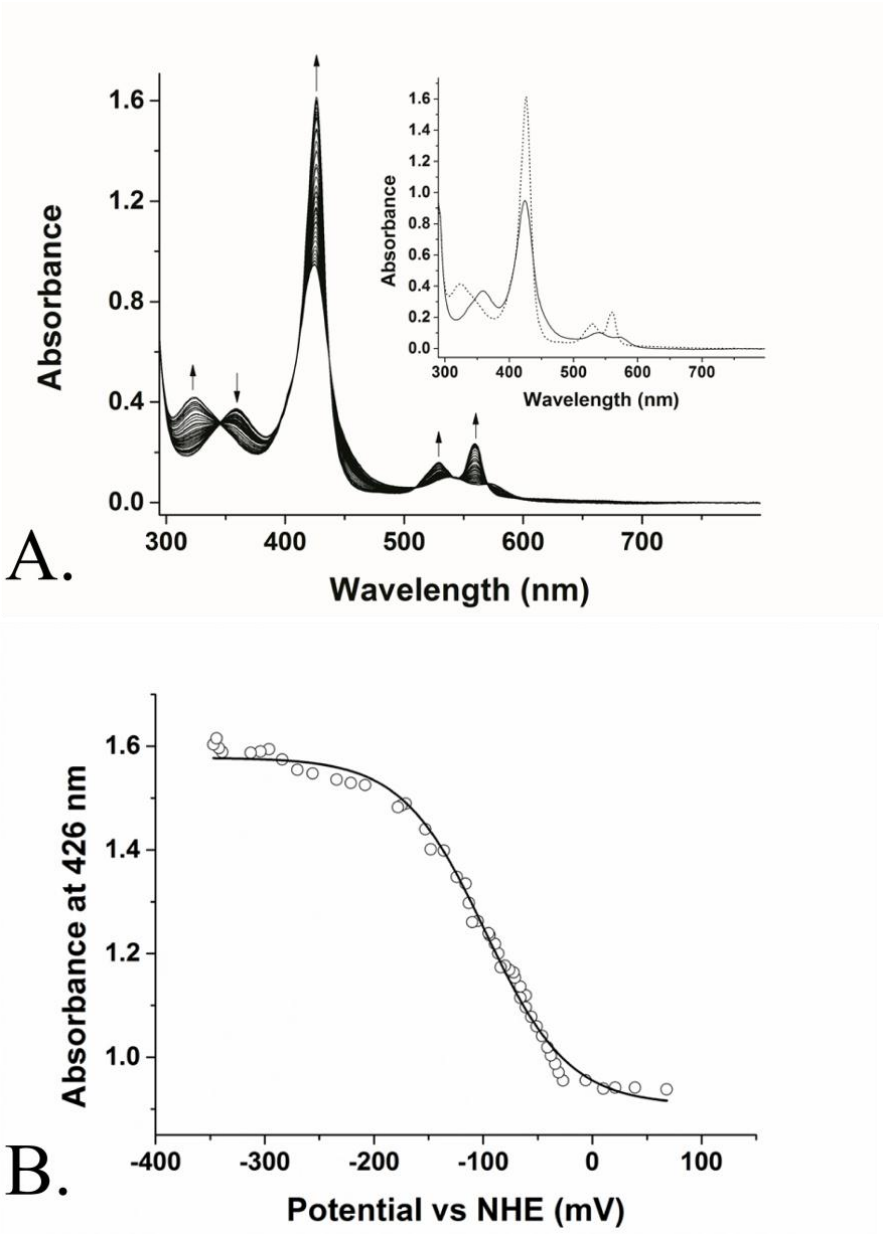


Figure 3.13. Redox titration of *M. barkeri* AhbA/B complex. A. UV-visible spectra recorded during the redox titration. The arrows show direction of maximal changes during the reductive titration. **B.** Plot of absorbance at 426 nm against redox potential versus normal hydrogen electrode (NHE). The redox potential for the Fe^{3+}/Fe^{2+} couple of the bound haem iron is -98 ± 3 mV.

3.2.6 *D. desulfuricans* AhbA/B kinetics

The decarboxylation of sirohaem to 12,18-didecarboxysirohaem has little impact on the conjugation of the macrocycle and therefore causes no noticeable change in the UV-visible absorbance spectrum between sirohaem, monodecarboxysirohaem and didecarboxysirohaem. Sirohaem and its derivatives are also oxygen sensitive and the use of alternative non-spectroscopic techniques to measure reaction rates in a continuous fashion, such as isothermal calorimetry, are impractical. Therefore the only method of quantifying the reaction was to perform stopped assays inside an anaerobic glove box and measure the reaction intermediates by HPLC-MS.

Kinetic parameters for *D. desulfuricans* AhbA/B were determined using the methods outlined in Section 2.4.10. Sirohaem (23 µM) was incubated at 37 °C for 10 minutes prior to the reaction before *D. desulfuricans* AhbA/B was added to a final concentration of 2.3 µM. The mixture was incubated at 37 °C for the duration of the reaction. Samples were taken from a bulk reaction and mixed with an equal volume of acetonitrile to stop the reaction and precipitate the protein. The sample was then left at 4 °C for 10 min before being centrifuged to remove any precipitated protein. The supernatant was analysed using HPLC-MS. To determine the quantity of the tetrapyrroles in the sample the elution peaks were integrated using Compass™ data analysis software (Bruker) and normalised. Normalising the data allowed for variations in peak intensity which would otherwise cause large outliers. Normalisation worked on the principal that the sum of all peaks must represent the total concentration of tetrapyrrole in the system (*i.e.* 23 µM) and therefore individual peaks represented an equivalent fraction of the total concentration.

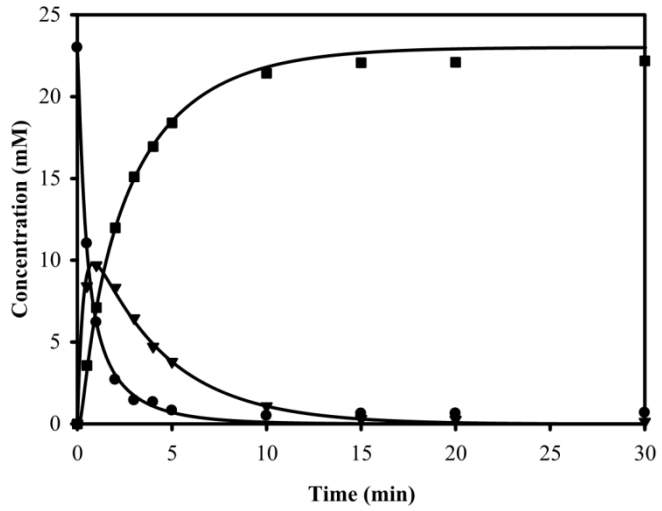
Initial experiments were performed to ascertain the time period over which the reaction is completed (data not shown). From this it was noted that the majority of the reaction occurred during the first 10 min, and the reaction was near completion after 30 min. Therefore time points were taken at 0, 0.5, 1, 2, 3, 4, 5, 10, 15, 20, 30 min to allow an enhanced view of the initial phase of the reaction. The results were fitted to a single active site model with two competitive substrates. A further inhibitory constant was added which gave a better fit for the data and represents the tight binding nature of the product (discussed in Chapter 5). Data were fitted using in house software developed by Dr. Andrew Lawrence (University of Kent) to the following equations:

$$v_1 = \frac{V_{max1}\, S_1}{K_{m1}\,(1 + S_2/K_{m2} + S_3/K_i) + S_1}$$

$$v_2 = \frac{V_{max2}\, S_2}{K_{m2}\,(1 + S_1/K_{m1} + S_3/K_i) + S_2}$$

where v_1 is the rate of sirohaem decarboxylation, S_1 is the concentration of sirohaem, v_2 is the rate of monodecarboxysirohaem decarboxylation, S_2 is the concentration of monodecarboxysirohaem, S_3 is the concentration of didecarboxysirohaem, K_i is the tight-binding constant of didecarboxysirohaem to the enzyme and V_{max1} / V_{max2} and K_{m1} / K_{m2} are the kinetic parameters of each reaction respectively.

From this analysis the following constants were determined:
K_{m1} = 11.74 ± 3.36 μM, V_{max1} = 44.63 ± 4.78 μM min^{-1}, K_{m2} = 9.94 ± 1.19 μM, V_{max2} = 33.22 ± 3.78 μM min^{-1}, K_i = 1.77 ± 0.32 μM. The data were averaged from three replicates (Fig. 3.14). These data show that both decarboxylation reactions have very similar parameters. Although the initial reaction has a higher V_{max}, this is compensated for by the intermediate having a slightly increased affinity for the enzyme, which prevents the build-up of monodecarboxysirohaem. The binding constant for the product is significantly lower than that of the substrates, concurrent with the tight binding of the product to the enzyme (described in Chapter 5).



K_{m1}	11.74 +/ - 3.36 μM
V_{max1}	44.63 +/ - 4.78 μM min^{-1}
K_{m2}	9.94 +/ - 1.19 μM
V_{max2}	33.22 +/ - 3.78 μM min^{-1}
K_i	1.77 +/ - 0.32 μM

Figure 3.14. Kinetic analysis sirohaem decarboxylase. Data from the reaction of sirohaem (23 μM) with *D. desulfuricans* AhbA/B (2.3 μM). Initially, sirohaem (circles) is decarboxylated to monodecarboxysirohaem (triangles), which is in turn decarboxylated to didecarboxysirohaem (squares). These reactions follow a competitive inhibition model for both substrates and the product. Data were fitted using in-house software.

3.2.7 AhbA/B decarboxylates at carbons 12 and 18

It had been postulated that AhbA and AhbB would remove the acetic acid side chains of sirohaem at carbons C12 and C18 (Storbeck *et al.*, 2010), although the decarboxylation of acetate side chains at carbons C2 and C7 would give an equal mass difference. Confirmation of the location of decarboxylations was achieved using NMR.

Sirohaem was incubated with AhbA/B from *D. vulgaris* in excess for ≈ 16 hours at 37 °C in an anaerobic chamber (<2 ppm O_2). The resulting product was then purified from the protein using a DEAE column. The column was equilibrated with Tris HCl, pH 8, containing 100 mM NaCl prior to the loading of the sample. After sample loading, the column was washed with acidified dH_2O with a pH gradient from 8 to 3. Finally, the tetrapyrrole was eluted from the column in dH_2O, pH 2.5. The purified sample was freeze-dried before being resuspended anaerobically in 0.5 ml 100% deuterium oxide and sealed in a septum screw-capped 5 mm NMR tube.

NMR data were collected by Dr. Mark Howard (University of Kent) as described in Section 2.4.15. All NMR data were obtained at 298 K using a 14.1 T (600 MHz 1H) Bruker Avance III NMR spectrometer equipped with a TCI cryoprobe. NMR assignments were obtained using ^{1}H homonuclear DQF-COSY, NOESY (500 ms), rotating-frame Overhauser effect spectroscopy [ROESY (200 ms)], and total correlation spectroscopy [TOCSY (80 ms)] and ^{13}C, ^{1}H heteronuclear sequential quantum correlation (HSQC), computed tomography-HSQC (27 ms), heteronuclear multiple bond correlation, and HSQC-TOCSY (80 ms) experiments. Mixing/constant time periods are shown in parentheses. NMR data processing was achieved using TopSpin 3.0 and NMR data analysis using the didecarboxysirohaem analysis suite, version 2.0.

Data analysis and assignment was performed by Dr. Andrew Lawrence (University of Kent). Figure 3.15 shows the structure of 12,18-didecarboxysirohaem labelled with the atom definitions used for assignment and the NMR data including the NOE contacts observed in the ROSEY spectrum, as well as the two regions of the 14.1 T ^{13}C, ^{1}H HSQC with major resonances labelled. These data confirmed that the AhbA/B complex does indeed decarboxylate sirohaem at carbons C12 and C18.

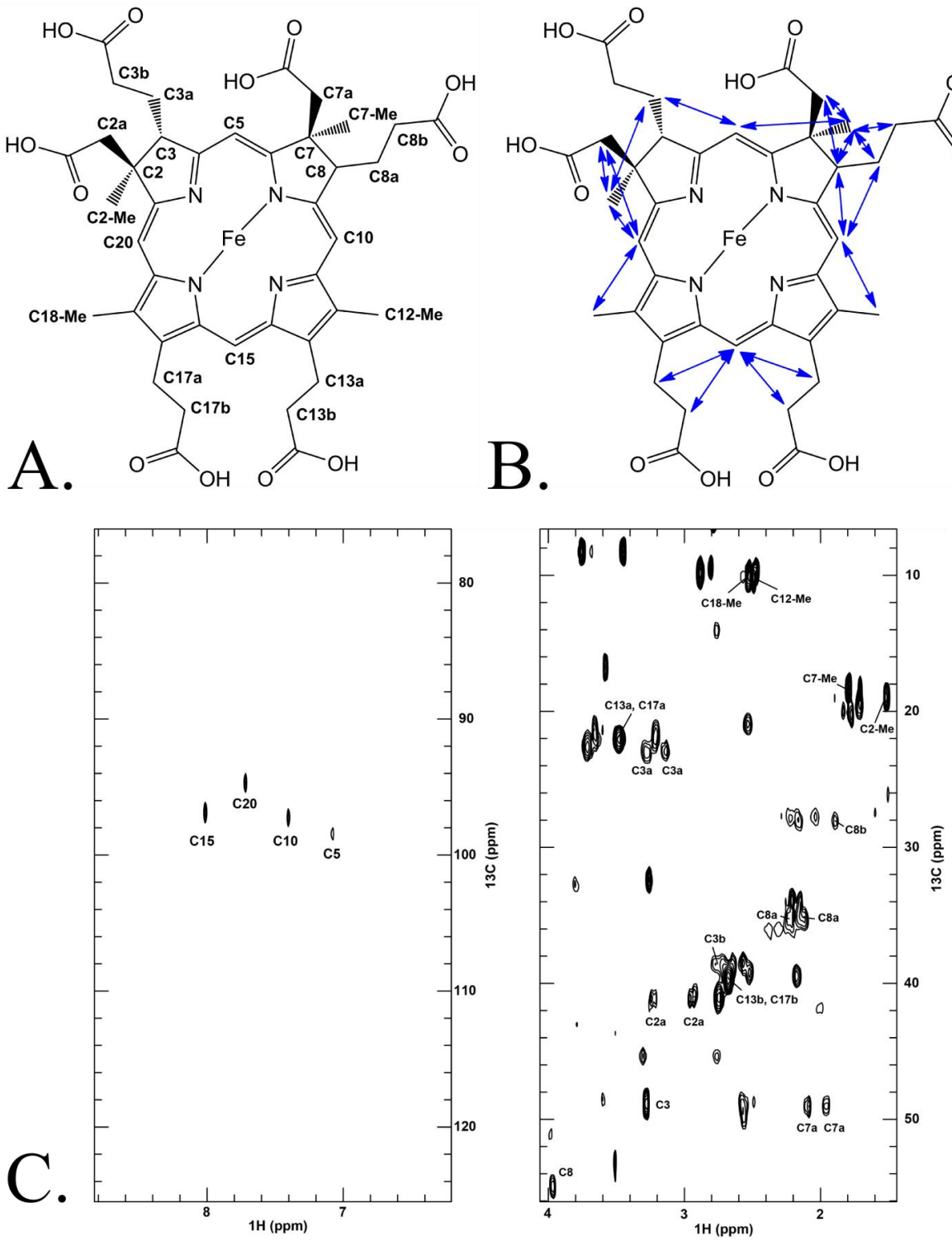


Figure 3.15. NMR analysis of 12,18-didecarboxysirohaem. A. The structure of 12,18-didecarboxysirohaem labelled with the atom definitions used for the assignment of the NMR chemical shifts (Table 3.1). **B.** The structure of 12,18-didecarboxysirohaem with the NOE contacts observed in the ROESY spectrum indicated with blue arrows. **C.** Two regions of the 14.1 T ^{13}C, ^{1}H HSQC of didecarboxysirohaem with major resonances labelled.

Table 3.1. NMR chemical shift assignments for 12,18-didecarboxysirohaem.

Assignment	δ (^{1}H) (ppm)	δ (^{13}C) (ppm)
C2A	2.94	41.1
C2A	3.23	41.1
C2-Me	1.52	18.9
C3	3.28	48.9
C3A	3.14	22.9
C3A	3.28	22.9
C3B	2.72	38.7
C3B	2.78	38.7
C5	7.08	98.5
C7A	1.96	49.0
C7A	2.09	49.0
C7-Me	1.79	18.3
C8	3.97	55.0
C8A	2.16	35.1
C8A	2.23	35.1
C8B	1.89	28.1
C10	7.40	97.3
C12-Me	2.49	10.3
C13A	3.48	22.0
C13B	2.67	39.6
C15	8.01	96.9
C17A	3.48	22.0
C17B	2.67	39.6
C18-Me	2.53	10.1
C20	7.72	94.7

3.3 Discussion

The AhbA and AhbB proteins from the sulphate reducing bacteria *D. vulgaris* and *D. desulfuricans* and the archaeon *M. barkeri* have been overproduced individually with N-terminal His_6-tag fusions in *E. coli*. However, the *M. barkeri* AhbA and AhbB proteins were found to be insoluble. Furthermore, the AhbA and AhbB proteins from both sulphate reducing bacteria were unstable and only low concentrations of purified samples could be obtained after metal affinity purification. Incubation of excess AhbA and AhbB proteins from the sulphate reducers with purified sirohaem led to the production of monodecarboxysirohaem and didecarboxysirohaem in varying quantities depending on the enzyme. The relative activity of the individual enzymes in isolation may be an artefact caused by differences stability rather than a true characteristic of the enzymes, although this would be extremely difficult to prove experimentally. Nonetheless, this has demonstrated that both AhbA and AhbB are able to perform both decarboxylation reactions on their own in an *in vitro* assay. Interestingly, mixing individually purified AhbA and AhbB proteins together in the same incubation with sirohaem resulted in nearly 100% conversion to didecarboxysirohaem. The enhanced activity observed when the proteins are combined demonstrates co-operativity between the two proteins. It may be that the observed activity of the single proteins is due to the action of poorly formed homomeric complexes, as opposed to a single monomeric species. Evidence to support this view would have to come from structural studies, as undertaken in Chapter 4.

Plasmid constructs were produced allowing the simultaneous recombinant overexpression of both *ahbA* and *ahbB* from all three sources in *E. coli*. In these constructs only the *ahbB* genes were fused to a sequence encoding an N-terminal His_6-tag to test if AhbA could be co-purified with AhbB. Purification of protein from *E. coli* overexpressing these constructs led to the co-purification of AhbA and AhbB in all three cases. The production of high yields of stable protein suggested that these proteins naturally form complexes, and explains the poor stability of the individual proteins. Purification of the three homologous complexes led to the discovery of cross-species variation within this protein family. The AhbA/B complexes can be found in different oligomeric states. The complexes from *D. desulfuricans* and *M. barkeri* purify exclusively as dimeric species, whereas the complex from *D. vulgaris* is present in both dimeric and tetrameric species. Similarly, the complexes have differing haem binding capabilities, with the *M. barkeri* complex co-purifying with a high occupancy of haem, but *D. vulgaris* purifying with a much lower level of haem. *D. desulfuricans* AhbA/B on the other hand does not co-purify with haem at all.

The AhbA/B complexes from sulphate reducers are highly active. However, the *M. barkeri* complex demonstrates poor turnover in the absence of a reducing agent. The haem in the *M. barkeri* complex appears to modulate the activity, with the enzyme being more active when the haem is in the reduced state. The midpoint redox potential of the haem group found in the *M. barkeri* AhbA/B complex was found to be -98 ± 3 mV (at 426 nm), suggesting that the haem group would normally be present in the reduced form *in vivo*. These results suggest haem regulation may be present in archael AhbA/B complexes. In the case of *D. vulgaris* haem binding has no effect on activity suggesting there may be a different role for the haem bound to this enzyme complex. A comparison of the redox potential of the haem bound to *D. vulgaris* to that of *M. barkeri* would be interesting and may shed further light on the unusual properties of this enzyme. Binding constants of haem to the *D. vulgaris* and *M. barkeri* AhbA/B complexes would also further the understanding of their association. However, due to the advantageous binding of haem to the complexes during production and purification, the isolation of apo-protein has not been possible.

The kinetic parameters of the *D. desulfuricans* AhbA/B complex were determined via a stopped assay. The results fit a single active site model with both sirohaem and monodecarboxysirohaem being competitive substrates. The initial reaction appears to have a higher V_{max} but this is compensated by the higher affinity for the monodecarboxysirohaem. A greater fit was achieved when a third inhibitory constant was added which, represents the tight binding of the product to the enzyme (discussed in Chapter 5). The fact that monodecarboxysirohaem can be observed at high concentrations, and that it fits a single active site model, suggests that the intermediate may be released and rebinds before the second decarboxylation is performed. Therefore, there may be a specific order of decarboxylation, although this order has not been investigated. By employing NMR it has been shown that the AhbA/B complex selectively decarboxylates sirohaem at carbons C12 and C18 to form 12,18-didecarboxysirohaem. This provides further confirmation that the next enzyme in the pathway, AhbC, removes the acetate side chains found at carbons C2 and C7 to form iron coproporphyrin.

In summary, the results presented in this chapter have contributed to the current understanding of the function, and properties of the AhbA and AhbB enzymes. It has been demonstrated that these proteins form a complex that enhances the stability and activity of the individual subunits. Cross-species variations between complexes have been discovered including differences in oligomerisation state, haem binding and activity. The haem bound to *M. barkeri* AhbA/B appears to be regulatory although this is not the case with *D. vulgaris* AhbA/B. Further work is required to understand the role of haem in the *D. vulgaris* complex.

Analysis of AhbA/B complexes from other organisms may provide a more coherent view of these proteins and their properties.

Chapter 4

Crystallisation and structural analysis of the AhbA/B complex

4.1 Introduction

In 1914 Max von Laue was awarded the Nobel Prize in Physics for the discovery of the diffraction of X-rays by crystals. Now, 100 years on with over 20 Nobel Prizes related to it, X-ray crystallography lies at the forefront of structure determination of proteins and small molecules. To date, over 100,000 protein structures have been deposited in the Protein Data Bank (PDB), which has been facilitated by the development of new and enhanced tools for crystallisation, data collection and structure determination.

All of the haem biosynthesis proteins from the classical pathway with the exception of oxygen-independent protoporphyrinogen IX oxidase (HemG) and the recently discovered protoporphyrinogen IX oxidase HemJ, have had their structures solved by X-ray crystallography (Layer *et al.*, 2010). Solution of the three dimensional structures of these proteins has led to an enhanced view of the function of these enzymes, allowing the determination of many reaction mechanisms, as well as the structural basis of diseases caused by mutations in the human pathway.

The AhbA/B proteins have been annotated as members of the Lrp/AsnC family of transcriptional regulators based on sequence homology for a helix-turn-helix domain (Pfam 13404). Further evidence supporting the idea that the AhbA/B proteins may be regulators comes from the closely related protein NirL from *Heliophilum fasciatum*. NirL was shown to be able to bind the putative promoter of the *nir* operon (Xiong *et al.*, 2007). It has been demonstrated that NirL is essential for the production of didecarboxysirohaem during haem d_1 synthesis by acting as a decarboxylase (Bali *et al.*, 2011) meaning that this protein plays both regulatory and catalytic roles.

Lrp/AsnC proteins are involved in the regulation of a variety of metabolic processes including catabolism and anabolism of amino acids (Brinkman *et al.*, 2003). AsnC is a specific regulator of the *asnA* gene required for the production of asparagine synthetase, which converts aspartate to asparagine in an ATP-dependent manner (Kolling & Lother, 1985). AsnC binds asparagine causing auto-repression of *asnA*. Lrp proteins are global regulators and are able to control expression of a wide range of operons (Ren *et al.*, 2007). Similarly, Lrp proteins are regulated by amino acid binding; however, this can be either in an agonistic or antagonistic manner and a number of amino acids are employed depending on the protein including leucine, asparagine, alanine, valine, lysine and proline (Brinkman *et al.*, 2003). Members of the Lrp/AsnC family display a range of oligomeric species in solution

including dimers, tetramers, octamers and hexadecamers, formed from repeats of the same subunit (Brinkman *et al.*, 2000, Chen *et al.*, 2001, Madhusudhan *et al.*, 1995, Thaw *et al.*, 2006, Willins *et al.*, 1991). Several structures of Lrp/AsnC proteins have been published with *E. coli* AsnC being the first bacterial homologue to be crystallised as a regulator-effector complex (Fig. 4.1).

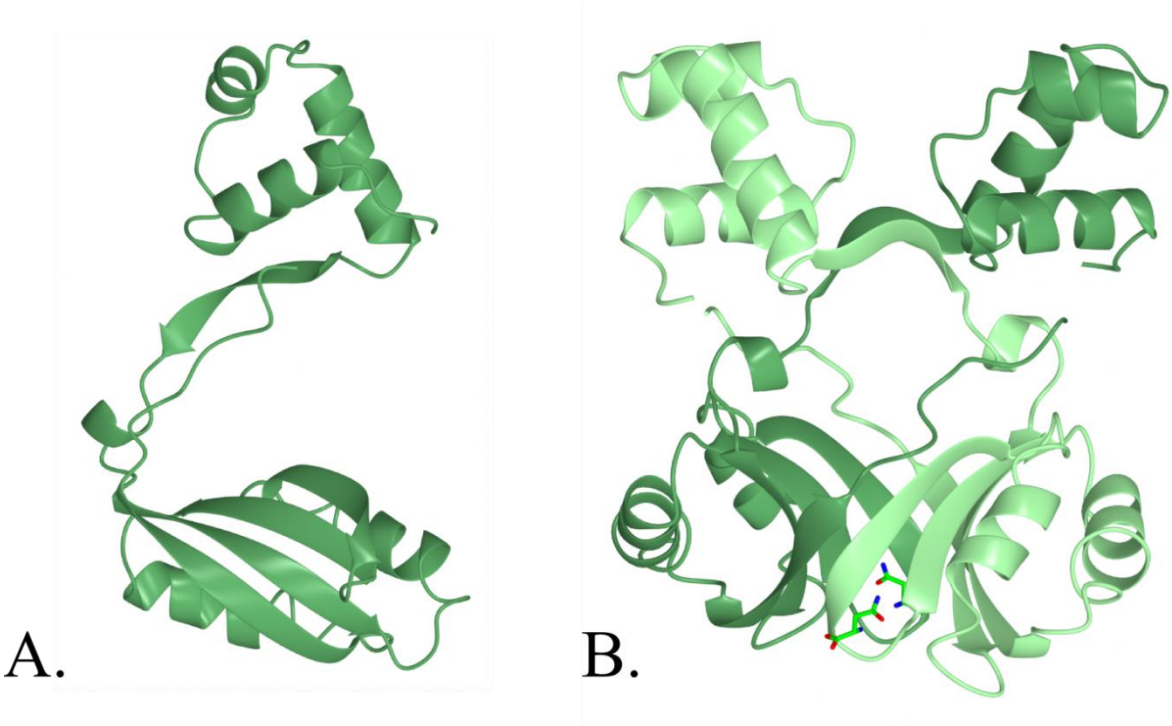


Figure 4.1. *E. coli* AsnC, PDB 2CG4. A. Monomeric *E.coli* AsnC, **B.** Homodimer of *E. coli* AsnC with bound asparagine (Thaw *et al.*, 2006).

No structural information has been published for proteins from either the haem d_1 or alternative haem pathways leaving questions as to whether the AhbA/B proteins contain the helix-turn-helix domain and if they have any structural homology to proteins belonging to the Lrp/AsnC family of transcriptional regulators.

The results outlined in this chapter describe crystallisation experiments carried out on AhbA/B homologues from *D. vulgaris*, *M. barkeri* and *D. desulfuricans*. Optimisation of crystallisation conditions, data collection and phasing experiments led to the solution of the AhbA/B complex from *D. desulfuricans*. Furthermore, substrate soaking was achieved allowing the determination of a product-bound complex. The structural data from this chapter were submitted to the PDB with accession codes 4CZD and 4UN1.

4.2 Results

4.2.1 Crystallisation of AhbA/B complexes

All proteins were overproduced in *E. coli* BL21 (DE3) and purified using nickel affinity chromatography as outlined in Sections 2.2.7 and 2.4.1. Composition of the original crystal screens used can be found in the Appendix.

4.2.1.1 Crystallisation of *D. vulgaris* AhbA/B

Crystallisation was attempted using purified AhbA/B from *D. vulgaris*. The protein was concentrated to 8-20 mg mL^{-1} prior to crystallisation and hanging drop vapour diffusion experiments were set up as described in Section 2.4.18.1. Structure Screen 1™ and Structure Screen 2™ (Molecular Dimensions Ltd.) were used to find conditions that produced crystals or crystalline precipitant. Initially, all screens were performed anaerobically in EasyXtal 15-well plates (Qiagen).

The best initial hit was obtained in 0.2 M lithium sulphate, 0.1 M Tris, pH 8.5, containing 30% (w/v) polyethylene glycol (PEG) 4000. Globular structures were observed with needle-like protrusions and a high degree of precipitant (Fig. 4.2). Optimisation was performed using Clear Strategy Screen I™ and Clear Strategy Screen II™ (Molecular Dimensions Ltd.), resulting in a range of conditions that produced similar needle-like crystals. Interestingly these hits could not be reproduced aerobically using siliconised coverslips (Hampton Research) and 24-well XRL plates (Molecular Dimensions Ltd.). After a wide variety of custom screening, including separation of dimeric and tetrameric fractions by FPLC prior to screening and use of Clear Strategy Screen 1™ and Clear Strategy Screen 2™ (Molecular Dimensions Ltd.), the thickness of the needles and high level of background precipitation could not be altered to produce diffraction quality crystals.

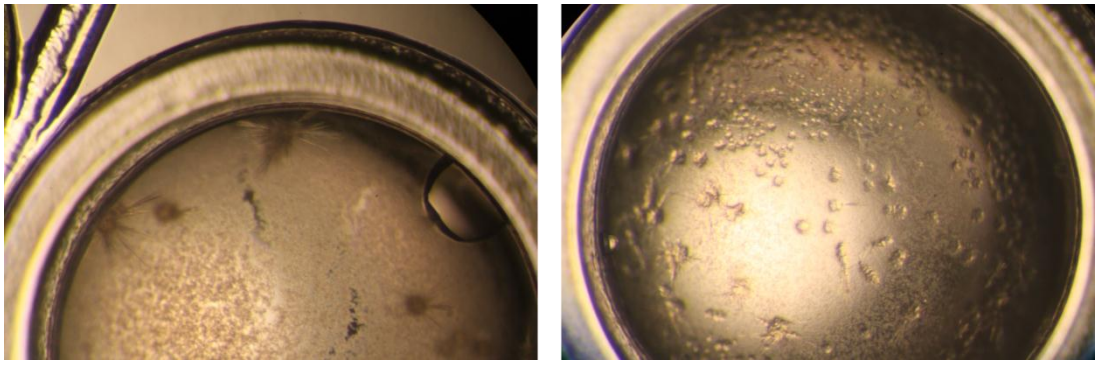


Figure. 4.2. *D. vulgaris* AhbA/B crystals. Needle-like crystals of AhbA/B from *D. vulgaris* produced in 0.2 M sodium acetate trihydrate, 0.1 M Tris, pH 8.5, containing 30% (w/v) PEG 4000. No further improvement of crystal size or thickness was achieved.

Re-screening was performed with the addition of haemin as the protein was previously shown to bind haem (Chapter 3). Addition of ligands often leads to the stabilisation of proteins and improvements in crystallisation. Unfortunately rescreening with haemin did not provide any hits. Further attempts were made using the MIDAS™ screen (Molecular Dimensions Ltd.). Several hits were obtained but further optimisation did not yield any protein crystals. Due to the problems encountered with these proteins other protein homologues were purified and screened in the hope that they may be more amenable to crystallisation.

4.2.1.2 Crystallisation of *M. barkeri* AhbA/B

Crystallisation was attempted using purified AhbA/B from *M. barkeri*. The proteins were further purified by FPLC with the dimeric fraction being purified from the aggregates. The protein was then concentrated to 8-20 mg mL^{-1} prior to crystallisation and hanging drop vapour diffusion experiments were set up aerobically using siliconised coverslips and 24-well XRL plates. Structure Screen 1™ and Structure Screen 2™ were used to find conditions that produced crystals or crystalline precipitant. Unfortunately, no hits were observed during initial screening and no further attempts to crystallise these proteins were made.

4.2.1.3 Crystallisation of *D. desulfuricans* AhbA/B

4.2.1.3.1 Initial crystallisation of *D. desulfuricans* AhbA/B

Prior to screening the AhbA/B complex from *D. desulfuricans* was further purified by FPLC to remove any potential contaminants. The protein was then concentrated to 8-20 mg mL^{-1} prior to crystallisation and hanging drop vapour diffusion experiments were set up aerobically as described for *M. barkeri* AhbA/B (Section 4.2.1.2).

Initial hits were observed with conditions 7 [30% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M tri-sodium citrate, pH 5.6] and 43 [30% (w/v) PEG 4000, 0.2 M ammonium sulphate] from Structure Screen 1™. Custom follow-up screens led to the production of a cluster of thin crystals from a derivative of condition 7 and high quantities of tiny crystals from derivatives of condition 43. After further screening the high levels of nucleation observed in condition 43 derivatives could not be sufficiently reduced and crystal size could not be increased, therefore this condition was not investigated further.

Using condition 7 as a starting point, a custom pH screen was performed in the pH range 5.0-7.6. This resulted in the formation of thin plate like crystals at pH 5.0 and 5.2 (Fig. 4.3). High throughput screening was attempted using a Mosquito® Crystal liquid handling system (TTP Labtech; in collaboration with Prof. David Brown, University of Kent / Argenta) and 100 nL: 100 nL drop ratios. However, the conditions did not respond well to drop size reduction and no crystals were observed. Therefore screening was continued using manual methods.

In an attempt to increase crystal quality further screening was performed in the pH range 5.0-5.4 with variables such as drop size, PEG mass, PEG concentration, dilution seeding (described in Section 2.4.18.2) and acid used for buffer pH adjustment (glacial acetic acid, HCl or citric acid). These screens resulted in individual crystals with sufficient size and thickness for diffraction experiments. The best conditions used were as follows: 10 mg mL^{-1} protein, 26% (w/v) PEG 3350, 0.2 M ammonium acetate, 0.1 M tri-sodium citrate (glacial acetic acid), pH 5.2 using dilution seeding. To prepare the crystals for data collection they were picked and cryoprotected using well solution containing 10% glycerol before being flash frozen in liquid nitrogen. Unfortunately, the diffraction limit of the crystals produced was 3.6 Å and a high degree of anisotropy was observed. Further attempts to produce higher quality crystals using these conditions were unsuccessful.

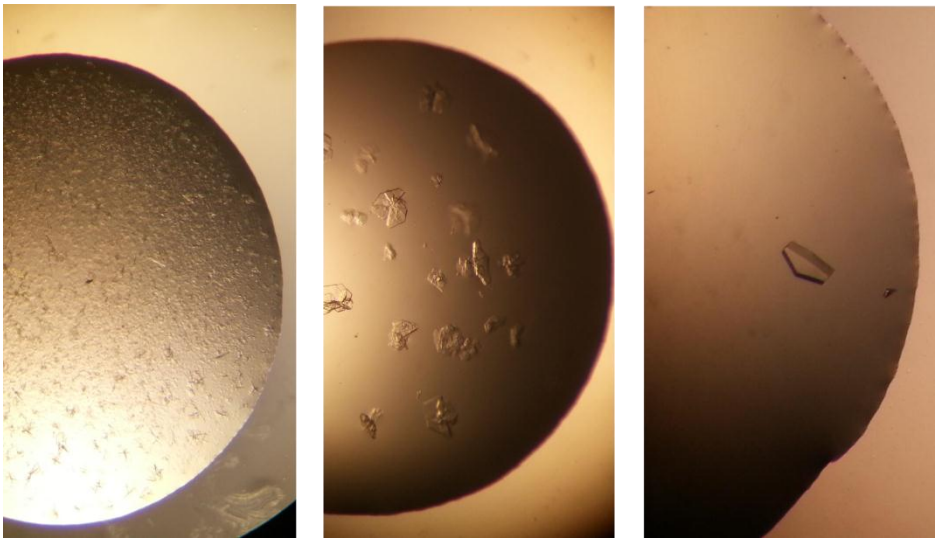


Figure 4.3. Optimisation of *D. desulfuricans* AhbA/B crystallisation. From initial crystalline precipitant (left), to production of stacked plates (middle) and finally single diffraction quality crystals (right). Optimised condition (right): 26% (w/v) PEG 3350, 0.2 M ammonium acetate, 0.1 M tri-sodium citrate (glacial acetic acid), pH 5.2, protein purified in Tris-HCl, pH 8.0.

4.2.1.3.2 Buffer optimisation using thermal shift assay

Access to the equipment required for a thermal shift assay became available through collaboration with Prof. David Brown (University of Kent / Argenta). By assessing the melting temperature of the complex in a range of conditions, the stability of the *D. desulfuricans* AhbA/B complex could be determined, allowing the selection of the optimum purification buffer.

The thermal shift assay was performed as outlined in Section 2.4.19. In brief, a range of buffers with differing pH (4.2-9.8) were used and the protein melting point was determined using the fluorescent dye Sypro® Orange to measure protein unfolding (Fig. 4.4). The optimum pH range was pH 7.2-8.0. Tris pH 7.2 provided the highest melting temperature; however, HEPES provided stability over a wider range of pH with the average melting temperature of 57 degrees as opposed to 55 for Tris buffers. Therefore HEPES pH 8.0 was selected as the optimum buffer.

After finding the optimum buffer and pH a second screen was performed using differing salt conditions and additives including glycerol and reducing agents (data not shown). No further enhancement of stability was achieved using the additives, therefore a standard buffer of 20

mM HEPES, pH 8, 100 mM NaCl was used to purify all *D. desulfuricans* AhbA/B protein used for crystallography after this experiment. Buffers and additives used for the thermal shift assay can be found in the Appendix.

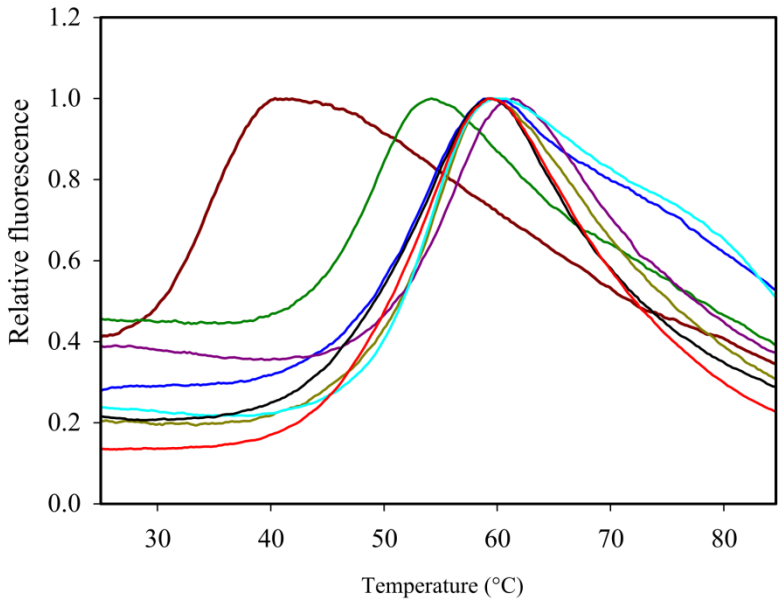


Figure 4.4. Melting curves of *D. desulfuricans* AhbA/B in a range of buffers with differing pH. Sodium citrate pH 4.6 (dark red), pH 5.0 (dark green) and pH 5.4 (gold). Sodium cacodylate pH 6.5 (blue). Tris pH 7.6 (black), pH 8.0 (red). HEPES pH 8.0 (purple). Glycine pH 9.4 (cyan).

4.2.1.3.3 Crystallisation of *D. desulfuricans* AhbA/B in HEPES pH 8

Rescreening of the *D. desulfuricans* AhbA/B complex was performed as described previously (Section 4.2.1.3.1) using protein purified in HEPES, pH 8 and Structure Screen 1™ and Structure Screen 2™ (Molecular Dimensions Ltd.). A number of previously unseen hits were observed with condition 2 [30% (w/v) PEG 4000, 0.2 M ammounium acetate, 0.1 M sodium acetate, pH 4.5], which produced needle like crystals, being the best. Again all conditions were optimised using a custom pH screen from 4.5-8.0, however only condition 2 yielded adequate crystals. Further screening was performed including variation of drop size, PEG mass, PEG concentration, and dilution seeding leading to the optimum condition of 1 µL : 1 µL drop, 26% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5. The Hampton Research Additive Screen was then employed resulting in a large wedge-like crystal with the addition of 0.01 M barium chloride (Fig. 4.5). The crystal was picked and cryoprotected using well solution containing 10% glycerol before being flash frozen in liquid nitrogen. A full data set was collected with a diffraction limit of 2.2 Å (see Section 4.2.2).

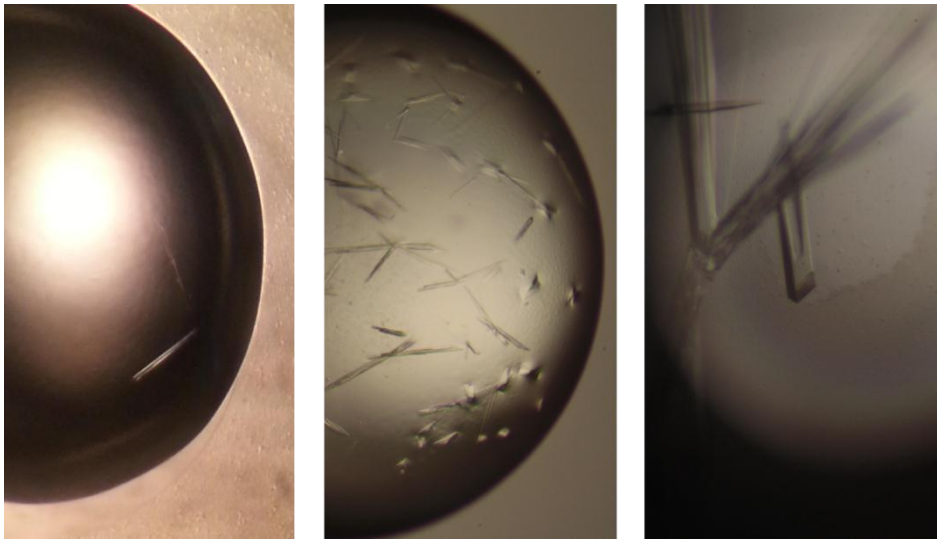


Figure 4.5. Optimisation of *D. desulfuricans* AhbA/B crystallisation using HEPES purified protein. Optimisation from initial needle crystals (left), to increased nucleation of crystals (middle) and finally diffraction quality crystals using additive screen (right). Optimised condition (right): 26% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5, 0.01 M barium chloride, protein purified in HEPES, pH 8.

4.2.1.4 Production of *D. desulfuricans* AhbA/B crystals for experimental phasing

As there are no protein structures published with sufficient sequence homology for molecular replacement, experimental phasing data had to be collected. Two methods were employed in an attempt to obtain data for phasing, isomorphous replacement (using heavy atom protein derivatives) and multiple wavelength anomalous dispersion (MAD; using selenomethionine labelled protein).

For isomorphous replacement native crystals were grown in 26% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5, using 11 mg mL^{-1} protein and 1 µL : 1 µL drop size. Crystals were then soaked in the well solution containing a range of heavy metal solutions (see Appendix) for 10-120 min, before cryoprotecting with 10% (v/v) glycerol and flash freezing in liquid nitrogen.

Selenomethionine labelled protein was produced using the method outlined in Section 2.2.8. After overproduction in *E.coli* KRX the protein was purified in an identical manner as described previously, with the purified protein buffer exchanged into in 20 mM HEPES, pH 8, 100 mM NaCl. Again the protein was purified by FPLC and concentrated to 20 mg mL^{-}

[1].The protein was crystallised in 26% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5, using 11 mg mL^{-1} protein, and the seeding technique described in Section 2.4.18.2. Crystals of selenomethionyl protein grew smaller than those of native protein. Crystals were again cryoprotected using 10% (v/v) glycerol before being flash frozen in liquid nitrogen.

Selenium incorporation was confirmed by HPLC-MS analysis of the protein. An equivalent of 15 µg protein was injected on to a Jupiter® 5 µm C4 column (2 x 50 mm; Phenomenex Inc). The column was developed with a binary gradient at a flow rate of 0.2 mL min^{-1}. Solvent A was 0.05% (v/v) TFA and solvent B was 0.045% (v/v) TFA and 80% (v/v) acetonitrile. The column was equilibrated with 5% solvent B prior to the protein injection of protein on the column. Proteins were eluted as a single peak.

Data was processed using Compass™ Data Analysis (Bruker). As selenomethionine has an increased mass of 46.89 Da in comparison to methionine, incorporation can be determined by comparing the mass of native protein with that of selenomethionyl protein. Selenomethionine labelling efficiency was determined using the following equation:

$$\%\ \text{labelling efficiency} = 100 \times \frac{(MW_{labelled\_experimental} - MW_{unlabelled\_experimental})}{(MW_{labelled\_theoretical} - MW_{unlabelled\_theoretical})}$$

The N-terminal methionines of the proteins were removed during the experiment (a common occurrence during protein MS). Therefore it was not possible to confirm whether the proteins were fully labelled. However, discounting the N-terminal methionines 100% labelling efficiency was achieved for both proteins.

4.2.1.5 Production of tetrapyrrole soaked crystals

All attempts to co-crystallise the AhbA/B complex with sirohaem were unsuccessful, therefore crystal soaking was performed. Crystals were produced as described in Section 2.4.18.1. Identical reagents and volumes were used as above (Section 4.2.1.3.3); however, crystallisation was carried out in EasyXtal 15-well plates. All plates and reagents were taken into the glove box 24 hours prior to use. Reagents were left open in the glove box for 2-3 hours, allowing oxygen to diffuse out of the solutions whilst minimising the amount of evaporation, as degassing all solutions was impractical.

Native crystals were grown in 26% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5, using 11 mg mL^{-1} protein and the seeding technique as described in Section 2.4.18.2. After the crystals had formed they were soaked in well solution with an additional 10% (v/v) glycerol and a 10-fold excess of purified sirohaem. The crystals were left to soak for 24 hours before being directly flash frozen in liquid nitrogen.

4.2.2 Crystallographic data collection and refinement

All native and phasing data were collected at the Diamond Light Source (Oxfordshire, England; on beamlines IO4-1 and IO2, respectively) in collaboration with Prof. David Brown (University of Kent / Argenta). Native crystals diffracted to 2.2 Å and a full data set was collected. Heavy metal soaked crystals did not provide adequate diffraction for phasing data. However, data sets were collected using selenomethionyl crystals at three wavelengths to a resolution of 2.7-2.8 Å. These data allowed phasing using MAD data. Both native and selenomethionine crystals were found to belong to the space group $P2_12_12_1$ and had a Matthews coefficient predicted to have 2 copies of the AhbA/B heterodimer in the asymmetric unit. Selenium substructure and initial chain tracing was determined with Autosolve (Solve/Resolve) software from the PHENIX package (Adams *et al.*, 2010) using the MAD data.

Initial models were rebuilt in Coot (Emsley *et al.*, 2010) using maps and phases calculated by Solve/Resolve to a resolution of 2.8 Å and further validated using Buccaneer (Cowtan, 2006). Initial model building and refinement used loose non-crystallographic symmetry (NCS) restraints, which were relaxed in final refinement. The native AhbA/B model was built using molecular replacement of the selenomethionine model using PhaserMR (McCoy *et al.*, 2007) in the CCP4 suite (Winn *et al.*, 2011), and refined using Refmac5 (Murshudov *et al.*, 1997) and manual model rebuilding in Coot. Final refinement was performed using BUSTER (Global Phasing Ltd). AhbA/B was refined at 2.23 Å to a crystallographic *R*-factor of 19.8% (R_{free} = 23.6%).

Diffraction data from tetrapyrrole soaked crystals was collected in collaboration with Prof. Richard Pickersgill (Queen Mary University of London) at the European Synchrotron Radiation Facility (Grenoble, France; on beamline ID23-1). Substrate soaked crystals were refined using Refmac5 at a resolution of 1.97 Å to an *R*-factor of 20.4% (R_{free} = 26.6%). Soaked crystals had an occupancy of ~60%.

All data collection and refinement statistics can be found in Table 4.1. Data were processed using XDS with a resolution cut-off at CC1/2 0.5. Native and selenomethionine data were processed using the xia2 pipeline. Coordinates and structure factor amplitudes have been deposited in the protein data bank with accession codes 4CZD (AhbA/B) and 4UN1 (AhbA/B bound to didecarboxysirohaem).

Data were validated using the wwPDB X-ray Validation Server as part of the Autodep deposition process. Ramachandran plots of the final structures can be seen in Figures 4.6 and 4.7. For the native structure 96.8% (573/592) of all residues are in favoured regions with 3 outliers, Gln64B, His70B and Ser121B. Residues 64 and 70 are in the disordered linking region of AhbB, while Ser121 is in a loop. For the tetrapyrrole bound structure 95.8% (567/592) of all residues are in favoured regions with 4 outliers, Lys162A, Lys65B, Trp68B and His70B. Again the residues of AhbB are located in the disordered loop region. Lys162A is at the C-terminal end of AhbA in a disordered loop region.

Table 4.1. Summary of crystallographic data and refinement statistics. RMSD - root mean squared deviation.

	Native AhbA/B	**Sirohaem soaked AhbA/B**	**SeMet AhbA/B (peak)**	**SeMet AhbA/B (inflection)**	**SeMet AhbA/B (remote high)**
Data collection					
Resolution (Å)	69.8-2.23 (2.29)	44-1.97 (2.02)	75.3-2.64 (2.71)	78.26-2.65 (2.72)	78.32-2.71 (2.78)
Wavelength (Å)	0.92	0.972997	0.97917	0.97935	0.97854
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell constants (Å)	55.85 x 78.62 x 151.4	55.3 x 78.6 x 150.3	55.63 x 78.22 x 150.6	55.62 x 78.26 x 150.7	55.64 x 78.32 x 150.7
Multiplicity	5.0 (3.8)	5.3 (4.9)	12.7 (9.9)	12.8 (10.1)	13.0 (11.9)
Mean (I)/sd(I)	13.1 (2.0)	13.38 (2.5)	20.4 (3.0)	20.9 (3.0)	21.3 (3.7)
Observations	160189 (8051)	245471 (16175)	254097 (13624)	252537 (13697)	24924 (15998)
Unique observations	32358 (2121)	446648 (3309)	19969 (1383)	19771 (1356)	18595 (1340)
R_{merge}	4.6 (69.3)	8.2 (80.4)	8.9 (70.7)	8.7 (69.6)	8.9 (66)
Completeness (%)	97.6 (89.6)	99.2 (97.0)	99.6 (95.1)	99.6 (95.3)	99.9 (99.3)
Ligand occupancy (%)	-	60	-	-	-
Refinement					
R_{factor} (%)	19.8	20.4	-	-	-
R_{free} (%)	23.6	26.6	-	-	-
Bond length (Å; RMSD)	0.0156	0.017	-	-	-
Bond angles (°; RMSD)	1.6	2.2	-	-	-
No. residues	704	704	-	-	-
No. waters	252	576	-	-	-

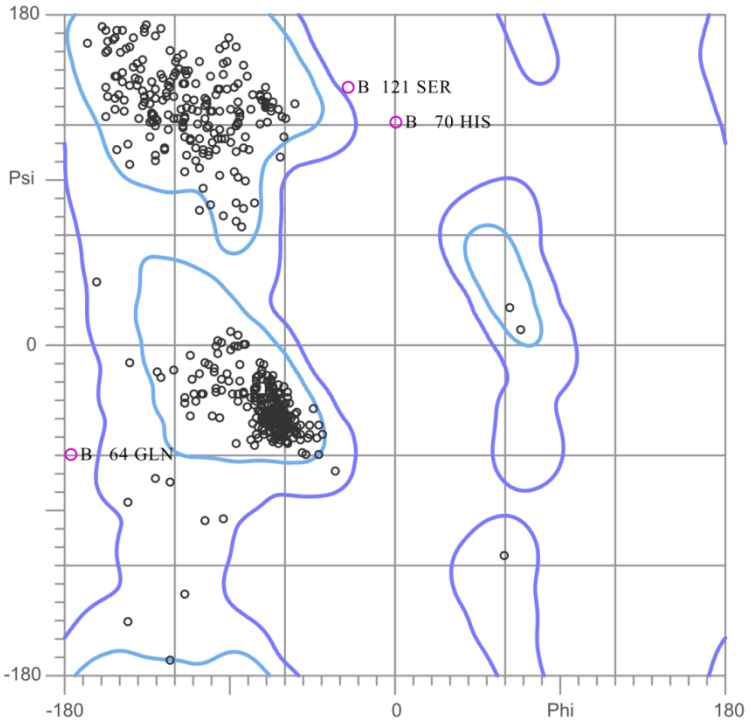


Figure 4.6. Ramachandran plot of native AhbA/B (PDB 4CZD). 96.8% (573/592) of all residues are in favoured regions with outliers labelled on the plot. Plot generated using MolProbity (Lovell *et al.*, 2003).

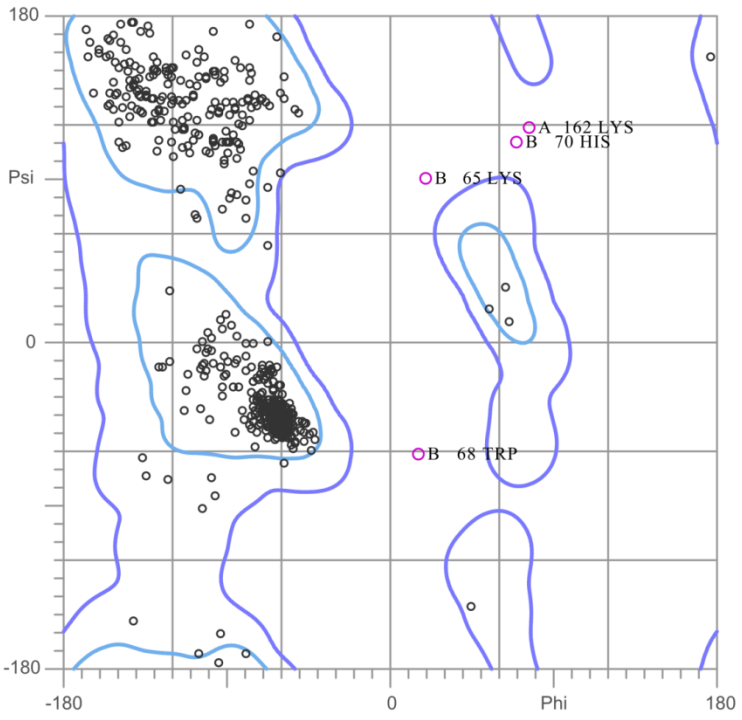


Figure 4.7. Ramachandran plot of product-bound AhbA/B (PDB 4UN1). 95.8% (567/592) of all residues are in favoured regions with outliers labelled on the plot. Plot generated using MolProbity (Lovell *et al.*, 2003).

4.2.3 The structure of the AhbA/B complex

4.2.3.1 Overall structure of the AhbA/B complex

The initial structure confirmed the formation of a heterodimer of AhbA and AhbB (Fig. 4.8) with two copies of the heterodimer in the asymmetric unit. The full sequence of AhbB was correctly traced in the structure except for the histidine tag, which was disordered. The majority of AhbA was also identified with only the first 20 and last 10 amino acids in disordered regions. The AhbA/B heterodimer is formed from two similar protein subunits that share a near identical fold (Fig. 4.9). Both proteins consist of an N-terminal and a C-terminal domain joined by a linking region (Fig 4.8).

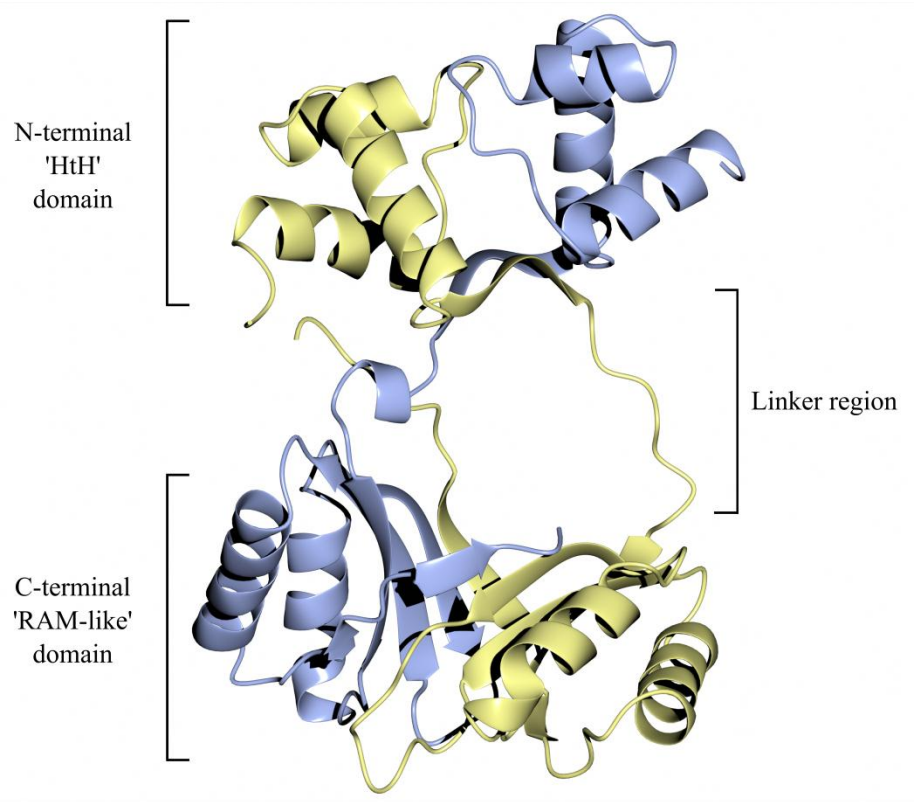


Figure 4.8. The structure of the AhbA/B complex from *D. desulfuricans*. The complex is formed from AhbA (blue) and AhbB (yellow).

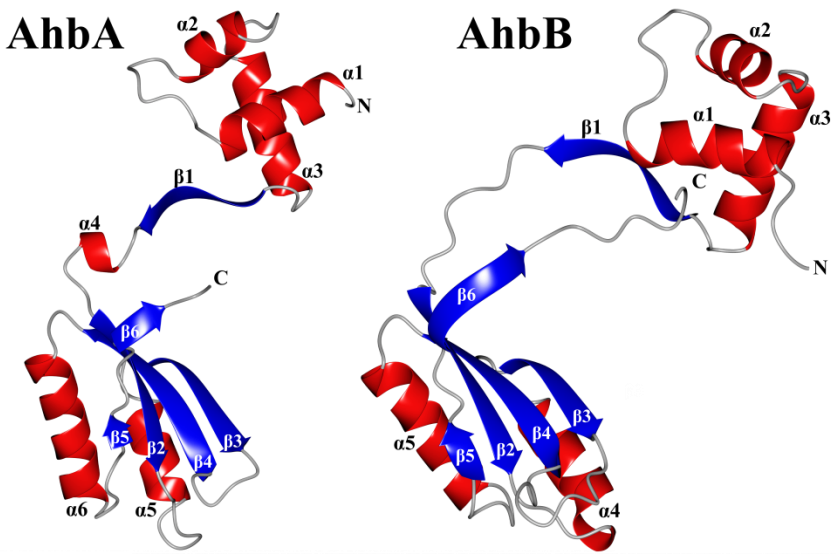


Figure 4.9. Structures of the individual AhbA and AhbB subunits. The monomers of AhbA (left) and AhbB (right) are coloured by secondary structural elements (helices = red, sheets = blue) and labelled by secondary structure. The termini are indicated N and C.

The N-terminal domain of the proteins contain 3 helices (α1–α3), two of which (α2 and α3) form a helix–turn–helix (HtH) motif. Helix α1 interacts with this motif via a hydrophobic core. The N-terminal domain then continues with a β-strand (β1) before the linking region joins the two domains. The AhbA linker contains a single turn helix (α4) that is not observed in AhbB. This AhbA α4-helix is stabilized by the N-terminal methionine of AhbB. The linker region of AhbB consists of a longer loop region the flexibility of which is inferred from the poorer electron density observed.

The C-terminal domains of each protein are reminiscent of the previously described RAM (regulation of amino acid metabolism) domain and the closely related 'ACT domain' with a βαββαβ fold (Chipman & Shaanan, 2001, Ettema *et al.*, 2002). However, as there is no evidence that AhbA/B is related to amino acid metabolism it will be subsequently referred to as a 'RAM-like' domain. A four stranded β sheet is formed from β2-5 with 2 helices packed against one face (AhbA α5 and α6, AhbB α4 and α5) stabilising the β-sheets through hydrophobic interactions. In AhbA α5 interacts with β3 and β4, and α6 with β2. Similarly, in AhbB α4 interacts with β3 and β4, and α5 interacts with β2 and β4. After a short loop, one final strand (β6) is formed at the C-terminal end of each monomer.

The Ahb monomers form a closely associated complex (Fig.4.8). At the base of the N-terminal domain, β1 from each protein forms an anti-parallel β-ribbon by direct hydrogen bonding at the start of the linking region. At the C-terminus the two RAM-like domains

dimerise via hydrophobic interactions at the base of each domain forming a sandwich of β-sheets with a hydrophobic cleft. Finally, the C-terminal β-strand (β6) directly hydrogen bonds to β3 of the opposing monomer to form a tight interaction. The different linkers of each subunit allow one side of the dimer to adopt an open conformation allowing access to the central cavity formed between two RAM-like domains. This cavity forms the substrate-binding domain (see Section 4.2.3 for substrate binding).

4.2.3.2 Comparison of the AhbA/B complex to AsnC/Lrp proteins

As mentioned previously, based on sequence similarity, the AhbA and AhbB proteins were both annotated as members of the Lrp/AsnC transcriptional regulator family, prior to the discovery that they were in fact involved in alternative haem synthesis. It is interesting to note that despite their catalytic role the AhbA and AhbB proteins have an identical 3D structural arrangement as members of the Lrp/AsnC family (Fig. 4.1 and 4.9).

The Lrp/AsnC proteins contain an N-terminal HtH motif and a C-terminal RAM domain linked by a β-sheet. Lrp/AsnC proteins form homodimers which can then be assembled into multimeric complexes via associations at the RAM domain of the subunits. The associations formed between subunits of the dimers are identical to those observed in the AhbA/B complex with a beta sheet interaction at the start of the linker region, hydrophobic interactions at the base of each RAM domain and finally hydrogen bonding by β-strands to the opposing RAM domains.

Structural similarity can be demonstrated by superimposing protein structures and measuring the root mean squared deviation (RMSD) of amino acid Cα atoms. Superimpositions were performed in CCP4 Molecular Graphics (McNicholas *et al.*, 2011) using the secondary structure method (SSM). Using *E. coli* AsnC (PDB entry 2CG4) as a template gives RMSD values of 3.8 Å (for 123 Cα atoms) and 2.6 Å (for 138 Cα atoms) for AhbA and AhbB, respectively. These values indicate that AhbB has a much better secondary structure alignment with AsnC than AhbA, which is clearly visible in Figure 4.10. By splitting AhbA into domains and aligning them with AsnC it is evident that the poor alignment is due to the arrangement of the domains in the protein, not their secondary structure. The HtH domain (amino acids 24-79) has an RMSD of 1.3 Å, whereas the RAM-like domain (amino acids 87-162) has an RMSD of 1.7 Å, showing that both domains share a high degree of structural similarity. The HtH domain of AhbA is angled towards the HtH domain of AhbB. This movement, combined with the flexible linker of AhbB, effectively opens up one side of the

dimer, allowing access to the cleft formed by the C-terminal domains. *E. coli* AsnC and other DNA binding Lrp/AsnC proteins show a much more condensed symmetrical structure.

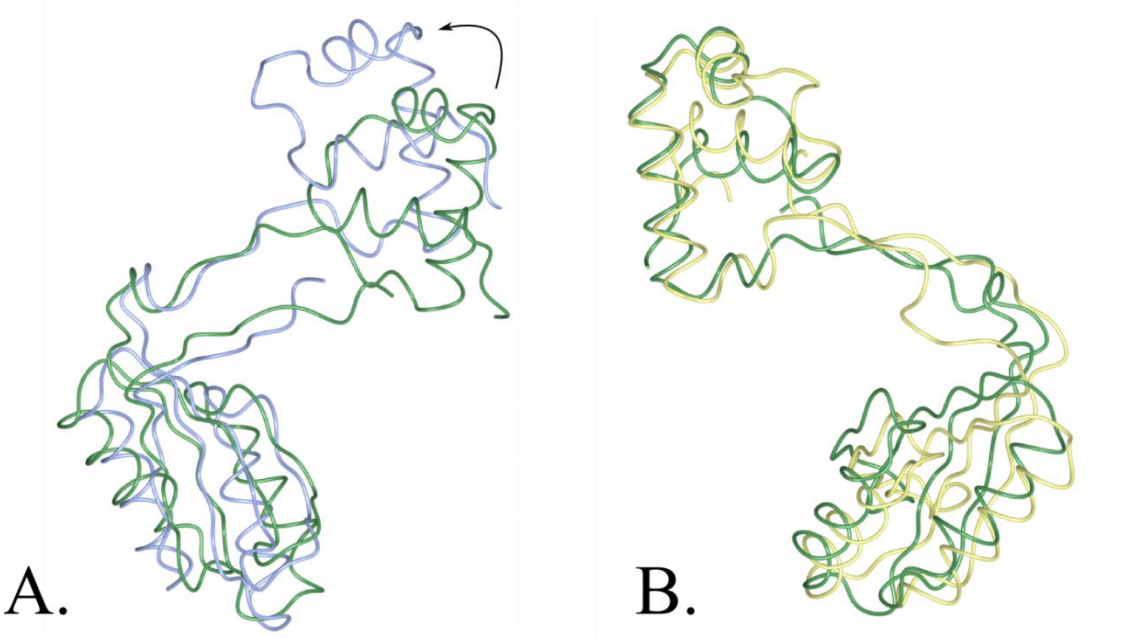


Figure 4.10. Structural alignments of *E. coli* AsnC and *D. desulfuricans* AhbA and B proteins. *E. coli* AsnC (green; PDB 2CG4) aligned with **A.** AhbA (blue; RMSD 3.8 Å) and **B.** AhbB (yellow; RMSD 2.6 Å). Arrow indicates the movement of AhbA HtH domain between the two structures.

The structural similarity of AhbA/B with *E. coli* AsnC suggests that they may still function as transcriptional regulators as well as enzymes. This hypothesis is strengthened by the fact NirL from *H. fasciatum* has been shown to bind DNA; specifically to the putative promoter of *nirJ2*, which encodes a protein homologous to AhbC and AhbD (Xiong *et al.*, 2007). The structure of an Lrp homologue in complex with DNA (the feast/famine transcription factor FL11 from the hyperthermophilic archaeon *Pyrococcus* OT3; PDB 2E1C) was used to model the AhbA/B complex onto DNA (Fig. 4.11). AhbB appears to be in the correct position to bind DNA (Fig. 4.11B), whereas the angle of the HtH domain of AhbA results in a significant clash with the DNA (Fig. 4.11A). Therefore it is unlikely that the AhbA/B complex would bind DNA without major conformational change. Lrp homologues have been noted to bind a variety of amino acids which effect DNA affinity, DNA bending, and the alteration/stabilisation of tertiary and quaternary protein structure (Brinkman *et al.*, 2002, Calvo & Matthews, 1994, Chen *et al.*, 2001, Madhusudhan *et al.*, 1997, Ren *et al.*, 2007, Thaw *et al.*, 2006, Jeong *et al.*, 2013). Therefore, it is possible that effector binding could

induce conformational changes to allow repositioning of the AhbA HtH domain, or induce multimerisation of AhbA/B complexes, which may represent a DNA binding state.

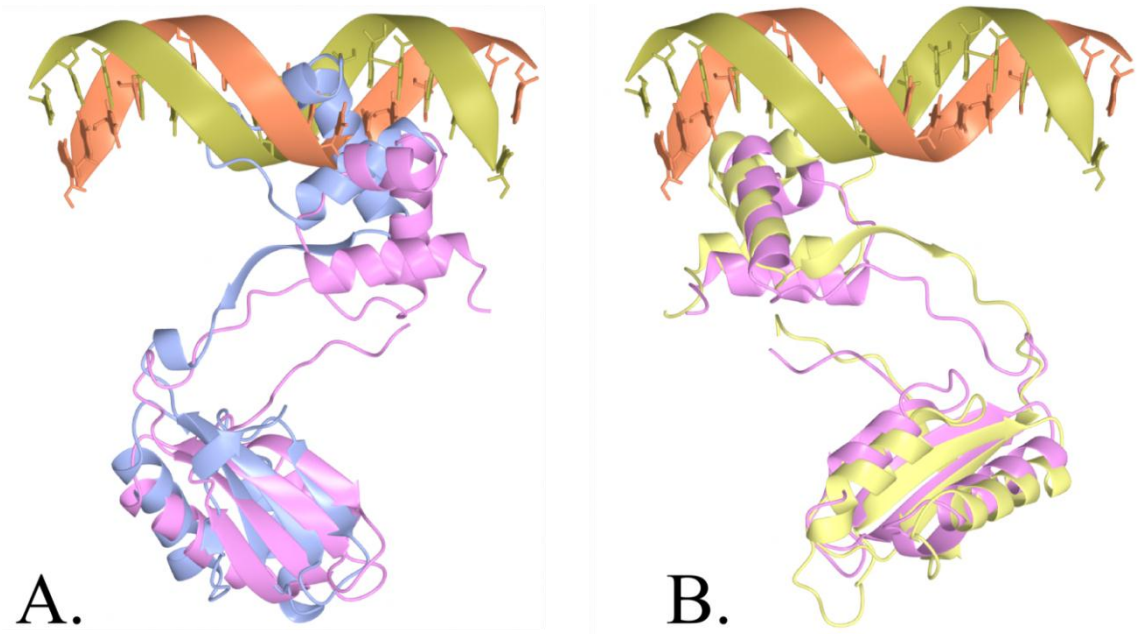


Figure 4.11. DNA binding model of AhbA/B using transcription factor FL11 from *Pyrococcus* OT3 bound to DNA as a template (PDB 2E1C). Superimposed structures of transcription factor FL11 from *Pyrococcus* OT3 (pink) bound to DNA with **A.** AhbA (blue) and **B.** AhbB (yellow). Structures superimposed as a dimer but displayed separately for clarity. While AhbB is in the correct alignment for DNA binding AhbA has large clashes with the DNA molecule.

4.2.4 The structure of the product bound AhbA/B complex

Sequence alignments of AhbA and AhbB and their homologues show a conserved HXYXR motif (Fig. 4.12). These conserved residues face the central cavity supporting their involvement in substrate binding and catalysis. This provides a pseudo-symmetrical binding pocket with positively charged residues suitable for binding the negatively charged tetrapyrrole substrate. To investigate this idea attempts were made at producing crystals with bound substrate. Co-crystallisation attempts were unsuccessful; however native crystals soaked with sirohaem diffracted to 1.97 Å and contained additional density consistent with the presence of a tetrapyrrole. As with the apo-crystals, two AhbA/B dimers were present in the asymmetric unit, both containing the additional electron density.

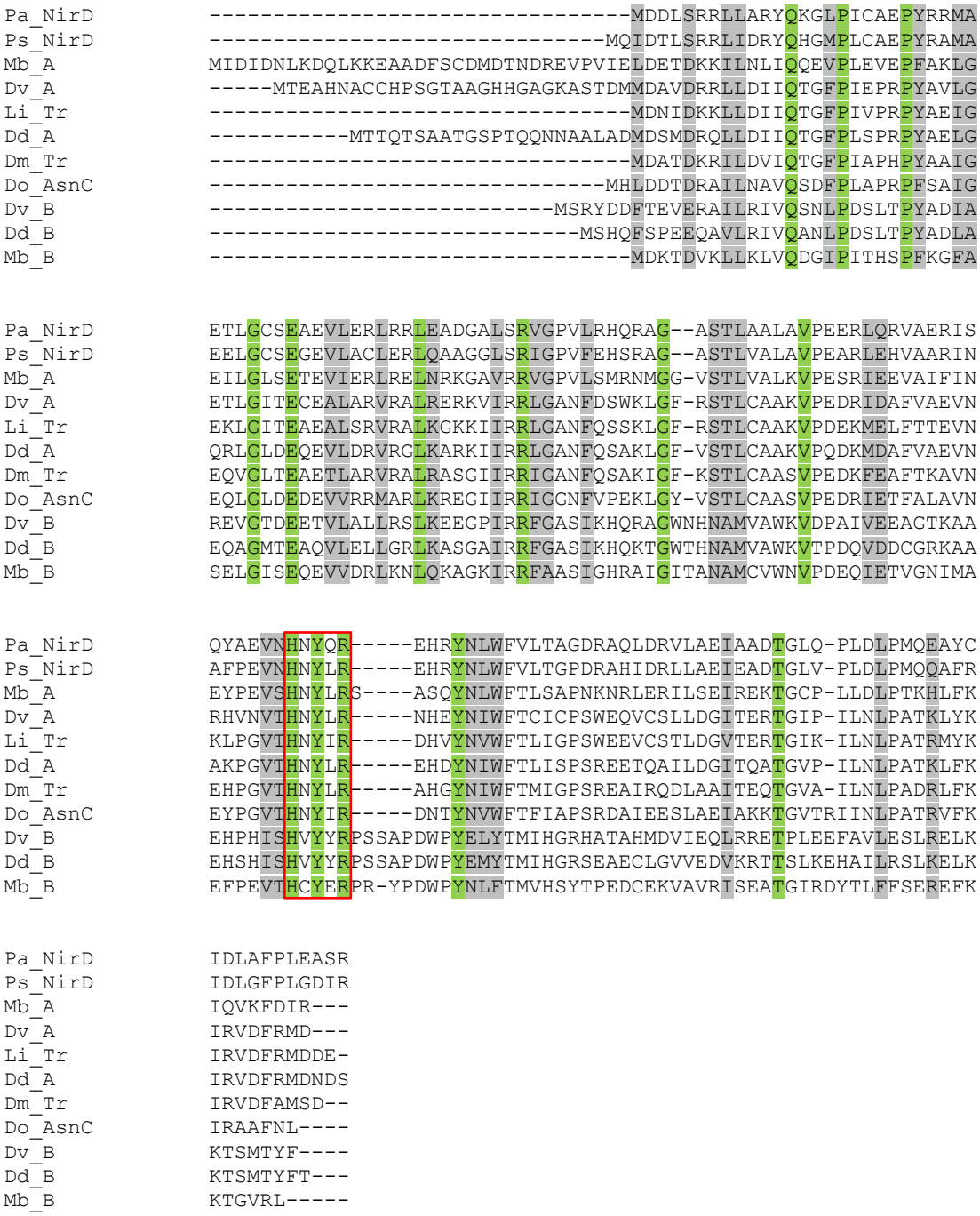


Figure 4.12. Sequence alignment of known and predicted AhbA and AhbB proteins and the haem d_1 biosynthesis protein NirD. Alignments were produced using clustalW. Identical residues are highlighted in green, with residues with similar properties highlighted in grey. The HXYXR motif is highlighted using a red box. Pa = *Pseudomonas aeruginosa*, Ps = *Pseudomonas stutzeri*, Mb = *M. barkeri*, Dv = *D. vulgaris*, Li = *Lawsonia intracellularis*, Dd = *D. desulfuricans*, Dm = *Desulfovibrio magneticus*, Do = *Desulfococcus oleovorans*. Annotations A= AhbA, B = AhbB, Tr = transcriptional regulator.

The electron density of the bound tetrapyrrole suggests the decarboxylations at carbons C12 and C18 have occurred and the bound molecule is actually the product didecarboxysirohaem. The observation also indicates that the crystals are functional as a decarboxylase. Overall, little movement is observed upon substrate/product binding as there is only a RMSD of 0.3 Å between native and product bound complexes. Generally large conformational changes during catalysis would result in cracking of crystals. Therefore, it is likely that the protein undergoes little or no conformational change during catalysis. As predicted the tetrapyrrole is located in the C-terminal beta sheet domain of AhbA/B in the cleft formed between the two protein subunits (Fig. 4.13). This is consistent with the results from the kinetics data (Section 3.2.6), which suggest a single active site mechanism.

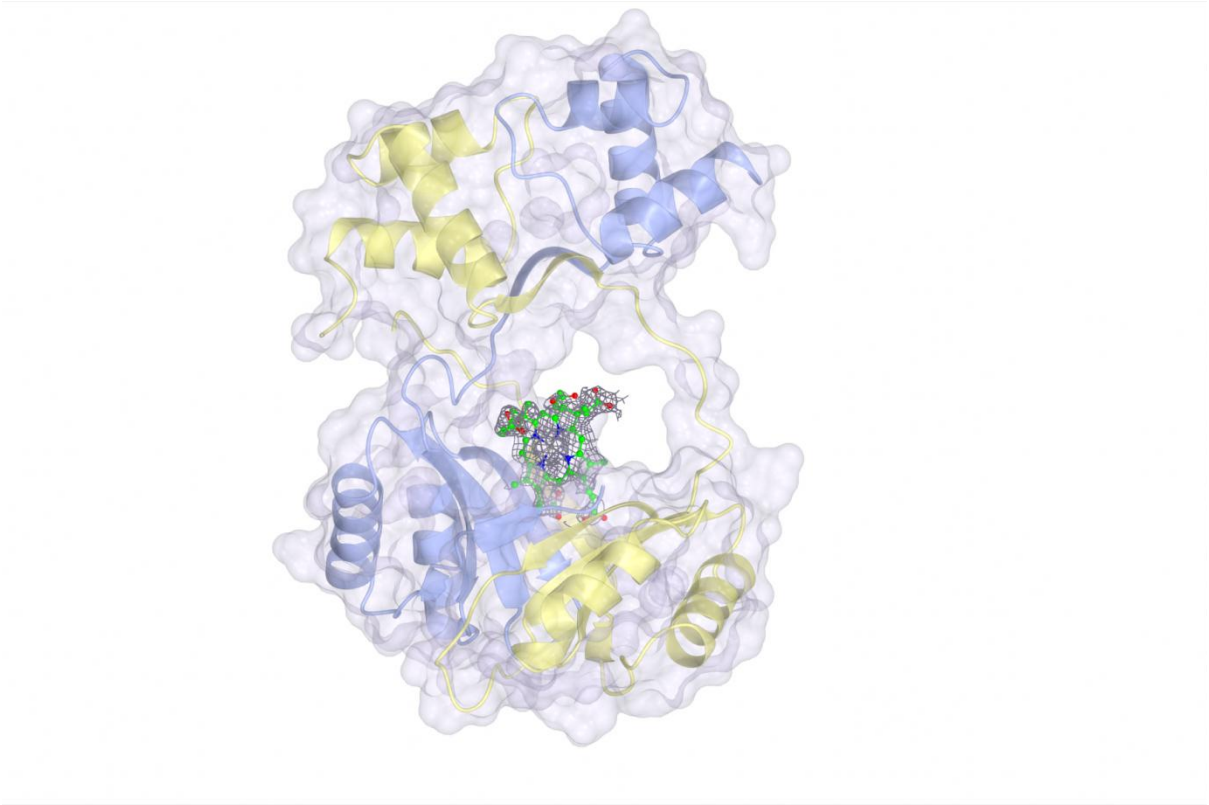


Figure 4.13. Crystal structure of didecarboxysirohaem bound to the AhbA/B complex. Didecarboxysirohaem (atom coloured) binds AhbA (blue) and AhbB (yellow) in the cleft between the two RAM-like domains. Electron density of the tetrapyrrole is displayed as a dark grey mesh, with protein surface rendered in light grey.

The structure shows that His115A coordinates the iron in the macrocycle (2.8 Å) and Arg119A is positioned such that it interacts with the propionate group from ring D of sirohaem (2.8 Å; Fig. 4.14). Lys152B appears to interact with the acetate group from ring A, although the electron density for this side-chain is poor suggesting it is in disordered state in

this product complex. These side-chains help to position didecarboxysirohaem with the methyl group of ring C (at carbon C12) within range of Arg102B.

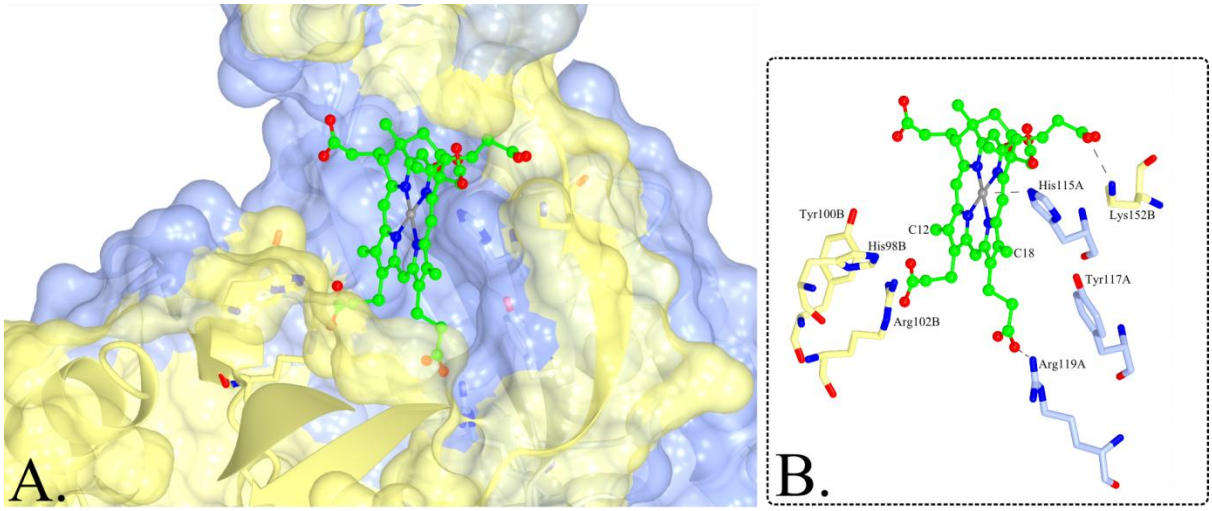


Figure 4.14. Close-up of the active site of AhbA/B and amino acid interactions with didecarboxysirohaem. A. Close up view of the active site formed by AhbA (blue) and AhbB (yellow) with protein rendered in space fill and cartoon with conserved HXYXR residues and Lys152B as sticks. **B.** Conserved HXYXR residues and Lys152 with interactions shown as dashed lines. Residues and carbons 12 and 18 are labelled.

Due to its positioning it is believed that Arg102B is a catalytic residue responsible for initiating the decarboxylation of the acetate side chain at carbon C12. Given that both decarboxylations have occurred in the crystal structure it is likely that this represents the positioning of the tetrapyrrole required for the second decarboxylation event to occur. Therefore the proposed mechanism for decarboxylation of sirohaem is based on acetate side-chain at C18 being decarboxylated first although this is yet to be experimentally proven.

For the initial decarboxylation at C18 the substrate will need to bind in a position such that the tetrapyrrole is flipped over by 180 degrees about the C5-C15 axis to bring C18 into a similar position to that currently occupied by C12. Rearrangement of the double bonds must occur to allow electrons to run to the pyrrole nitrogen which may be facilitated by Tyr100B or His98B. His98B is mobile in the structure, as judged from its poorly defined electron density compared with the adjacent residues, and is in a position where it can act as an acid/base. The role of His98B could be the initial protonation of C19 and later its deprotonation. Arg102B could then stabilise carboxylic acid of the acetic acid on C18, allowing decarboxylation to occur whilst providing a proton for the developing methyl group

(Fig. 4.15). Alternatively, Tyr100B could act as the acid, protonating the side-chain to give the C18 methyl group. After the initial decarboxylation at C18 the substrate ring will need to unbind and rebind flipped over to the position to that is observed in Figure 4.14 before the reaction is repeated for the decarboxylation at C12.

An alternative to tetrapyrrole flipping mechanism is one where sirohaem binds to the second pseudo-symmetrical active site where Arg119A, Tyr117A and His115A may perform equivalent roles to Arg102B, Tyr100B and His98B. After the decarboxylation of C18 the monodecarboxylated intermediate could then migrate ~4 Å to the position observed in Figure 4.14 where the second decarboxylation occurs as described above. Due to the size of the active site cleft and positioning of the HXYXR residues, use of the two active sites would be mutually exclusive. Therefore the migration mechanism would appear to follow a single active site mechanism when studied kinetically.

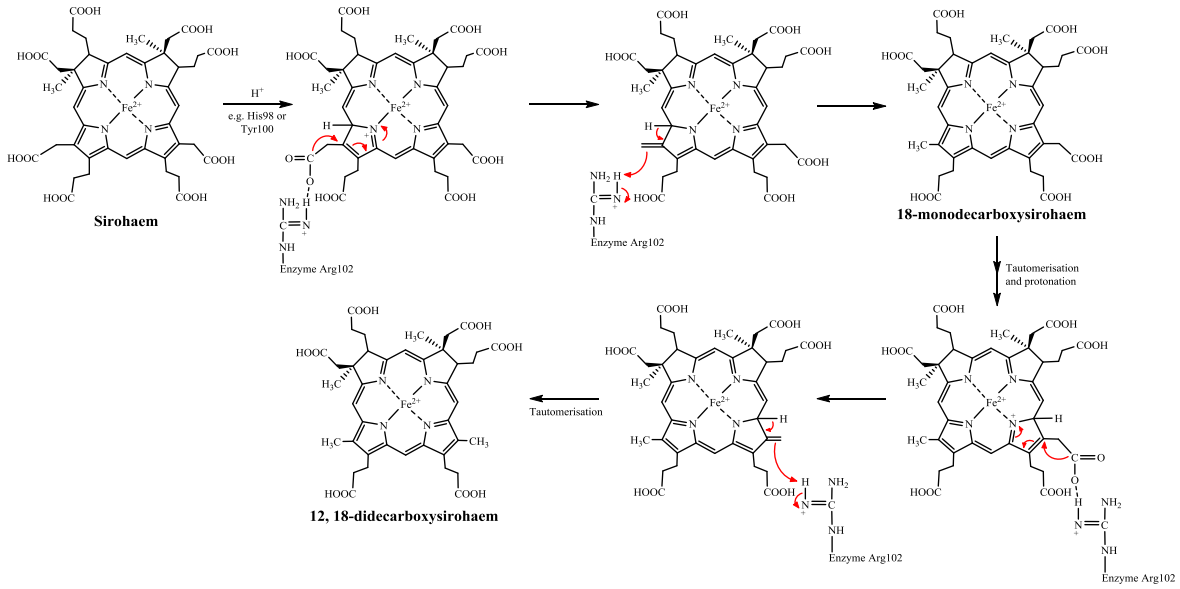


Figure 4.15. A proposed reaction mechanism for sirohaem decarboxylase AhbA/B. This mechanism is based on the flipping model.

From this structure it is unclear where the haem molecule would bind in the complexes from *M. barkeri* and *D. vulgaris*. As mentioned previously the proteins are active with the haem bound and given the size of the binding cleft it cannot be located in the active site with the substrate/product. Therefore further work must be undertaken to locate the haem binding site of these proteins.

4.3 Discussion

Crystallisation of the AhbA/B complexes from *D. vulgaris, M. barkeri* and *D. desulfuricans* was attempted, with only *D. desulfuricans* yielding diffraction quality crystals. Native crystals diffracted to 2.2 Å and the structure was solved using three wavelength selenomethionine MAD data. The structure of the product bound complex was also solved using native crystals soaked with sirohaem to a resolution of 1.97 Å. These represent the first crystal structures of members of the alternative haem biosynthesis pathway, as well as the first structure of a sirohaem decarboxylase.

The crystal structure of the AhbA/B complex confirms the presence of a heterodimeric protein, with the individual subunits having a nearly identical sub-structure. Both monomers contain a HtH domain and a RAM-like domain joined by a linker region. The arrangement of the two monomers forms a cleft in between the two RAM-like domains. The different alignment of AhbA and the flexible linker of AhbB open up the side of the protein complex allowing access to this active site.

The AhbA and AhbB proteins display a striking structural similarity to members of the Lrp/AsnC family of transcriptional regulators which are similarly formed of an HtH and RAM domain joined by a linking region. AhbB has a higher degree of structural similarity to *E. coli* AsnC than AhbA as the HtH domain of AhbA is in a different orientation. It is yet to be demonstrated if the AhbA/B proteins are able to bind DNA; however, it appears likely that a significant conformational change of the HtH domain of AhbA must occur for the correct positioning of the two HtH domains. The movement of the HtH domain of AhbA may restrict access to the active site, thereby reducing activity when in the DNA binding conformation.

The product bound structure revealed that didecarboxysirohaem interacts with several amino acids that are found in a conserved HXYXR motif in AhbA and AhbB proteins. Other members of this conserved motif appear to be within range of one of the sites of decarboxylation, further suggesting their involvement in facilitating the binding and catalysis of the reaction. A reaction mechanism has been proposed with Arg102B having a key role in facilitating the decarboxylation of the acetic acid side chains. The role of this arginine is similar to that of Arg37 in human uroporphyrinogen decarboxylase (HemE), an enzyme that performs analogous decarboxylation reactions on uro'gen III. It is interesting to note that despite the similar reactions these two proteins have drastically different three dimensional structures. Further work must be performed to ascertain the role of the conserved amino

acids in binding and catalysis. The structure also reveals no obvious region that may be involved in haem binding demonstrated by homologues of these proteins.

The solution of the crystal structure of AhbA/B raises many questions about the evolution and functionality of these enzymes, especially in relation to the similarities with DNA binding proteins. For instance, if the AhbA/B complex does not bind DNA, what significant evolutionary pressures could cause the repurposing of a transcriptional regulator to give catalytic function? Alternatively, if the AhbA/B complex does bind DNA, how is this regulated in order to maintain haem synthesis in these organisms and what genes do they regulate? Many of the organisms containing the alternative haem biosynthesis pathway also require sirohydrochlorin for synthesis of cobalamin and coenzyme F_{430} and sirohaem as a cofactor for sulphite and nitrite reductases. Were the AhbA/B complex to regulate the transcription of genes related to the latter half of the pathway (*i.e. ahbC* and *ahbD*) it would represent a very efficient method for controlling the use of sirohaem within the organism.

It should be noted that during the course of this work the structure of another sirohaem decarboxylase, NirDL from *Hydrogenobacter thermophilus*, was published in apo-form and with a substrate analog bound (Haufschildt *et al.*, 2014). As mentioned previously, NirDL is the product of a gene fusion of *nirD* and *nirL*, which, along with *nirG* and *nirH*, are required for the decarboxylation of sirohaem during the synthesis of haem d_1 [Fig. 3.1 (Bali *et al.*, 2014)]. However, *H. thermophilus* does not contain *nirG* and *nirH* and therefore NirDL is responsible for complete decarboxylation of sirohaem to didecarboxysirohaem in this organism. The crystal structure of NirDL displays a high degree of similarity to that of AhbA/B with an identical domain layout, however, the RAM-like domain equivalent to AhbB is contains a shorter β-sheet region and has a slightly different orientation to that of AhbB. The tetrapyrrole bound structure of NirDL shows an alternate binding site to that of AhbA/B. In the NirDL structure the tetrapyrrole is located at the top of the RAM domain just above the cleft, unlike the deeper positioning of didecarboxysirohaem in the AhbA/B complex. The tetrapyrrole is coordinated by His261, which is the equivalent residue to His98 from AhbB. As with AhbA/B, NirDL contains two HXYXR motifs and His261 belongs to the second of the two. The altered binding position may be caused by the difference in the orientation of the RAM-like domain which elevates the position of His261 (NirDL) in comparison to His98 (AhbB). Alternatively, location may be due to the fact the tetrapyrrole is an inactive substrate analog which may have one of two effects: the analogue may bind in a different region to the true substrate, or this structure may represent the initial binding site of the substrate which may be different to that of the product if NirDL uses a migration mechanism between the two HXYXR motifs.

Chapter 5

Further characterisation of sirohaem decarboxylases: enzyme trapping, mutagenesis and novel chimeric proteins

5.1 Introduction

One of the difficulties in analysing tetrapyrrole biosynthesis pathways is that many of the compounds and their intermediates are unstable in laboratory conditions due to temperature, light and oxygen sensitivity. Therefore attempts to overproduce or isolate intermediates using recombinant systems are often unsuccessful. Host organisms have developed various means to overcome these issues, one of which is direct metabolite channelling between enzymes. Direct metabolite channelling refers to the release of an enzyme product when the subsequent enzyme of the pathway is in close proximity (Miles *et al.*, 1999). Not only does this protect the product from oxidation and degradation but it also prevents exposure to other enzymes in the pathway which may cause out of order modifications leading to dead end products. Protection of these compounds is essential as they represent a huge energetic investment by the cell. It has recently been demonstrated that some enzymes from the aerobic cobalamin biosynthesis pathway tightly bind their products, providing a systematic approach for the stabilisation and isolation of pathway intermediates (Deery *et al.*, 2012). Given the oxygen-sensitive nature of sirohaem and didecarboxysirohaem, it was investigated whether the AhbA/B enzymes are also able to bind their products tightly.

Following on from the trapping experiments, site directed mutagenesis was employed to analyse the effect of conserved residues on product binding as well as enzyme activity. The sequence alignments (Fig. 4.12) combined with the crystal structure of *D. desulfuricans* AhbA/B (Chapter 4) allowed the informed selection of potentially important amino acids, especially those found in a conserved HXYXR motif. In the product bound structure His115A clearly coordinates the iron in the macrocycle (Fig. 4.14). Arg102B is positioned such that it would be in range of the acetate leaving group at carbon C12 prior to decarboxylation, and therefore has been proposed to be the catalytic residue for this reaction. Given the pseudo-symmetry of the active site, His98B and Arg119A may perform similar roles as His115A and Arg102B if the migration reaction mechanism is correct. Alternatively Arg119A may be involved in substrate binding as observed in the enzyme/product structure. His98B, Tyr100B and Tyr117A may be involved in the initiation of catalysis by facilitating double bond rearrangement. Site directed mutagenesis was used to examine the roles of these conserved residues and their effect on activity and product binding using AhbA/B from *D. desulfuricans*. Arg73A, Arg74A, Arg55B and Arg56B were also selected as they are highly conserved residues and are in proximity to the top of the active site cleft in the β-strand dimerisation region. Finally, His63B and His118B were selected as they are also conserved in AhbB proteins.

One of the striking things about the sirohaem decarboxylases studied in Chapter 3 was the variation in properties between species. The variety was especially surprising given the high degree of sequence identity for both proteins between the three organisms (summarised in Table 5.1). This led to the question of whether combinations of AhbA and AhbB proteins from different organisms can be used to make novel active chimeric complexes, and subsequently, whether properties such as haem binding or tetramerisation are conferred by a single subunit.

Table 5.1. Sequence identity of AhbA and AhbB proteins from different sources. Sequences were analysed using LALIGN (ExPASy).

Sequence 1	**Sequence 2**	**AhbA Identity (%)**	**AhbB Identity (%)**
D. desulfuricans	*D. vulgaris*	72	67
D. vulgaris	*M. barkeri*	41	37
D. desulfuricans	*M. barkeri*	40	35

The results described in this chapter aim to further elucidate the properties of the AhbA/B proteins. Enzyme trapping experiments were carried out with AhbA/B complexes from all three species to determine if they tightly bind their products. Site directed mutagenesis was used to assess the roles of conserved amino acids in *D. desulfuricans* AhbA and AhbB proteins with regard to product binding and catalysis. Finally, constructs were made allowing for the overproduction of AhbA and AhbB proteins from two different organisms in a single system and resultant chimeric complexes were biochemically characterised.

5.2 Results

5.2.1 An enzyme-trap approach for isolation of didecarboxysirohaem

In order to determine whether the AhbA/B complexes tightly bind their product *cysG*, which encodes the sirohaem synthase from *E. coli* (Warren et al., 1990), was included in all of the constructs containing *ahbA/B* (Table 2.2). Although the expression strain already contains *cysG*, the gene was included in the construct to prevent sirohaem production becoming a limiting factor within the cell. All necessary constructs were made using the Link and Lock technique with only AhbB containing a His_6-tag. *E. coli* BL21 (DE3) were transformed with the appropriate plasmid (pDP019, pDP032 and pDP038; Table 2.2) and cultures were grown in 2YT media at 28 ºC for 24 hours with constant shaking at ~160 rpm. All trapping strain cultures were supplemented with 20 mg L^{-1} ALA 6 hours prior to harvest in an attempt to increase the production of sirohaem within the cells.

To prevent degradation/oxidation of tetrapyrroles all protein purifications were performed anaerobically (<2 ppm O_2). The AhbA/B complexes were purified using nickel affinity chromatography and the pull-down technique as previously described (Section 3.2.3). The purified AhbA/B complexes were analysed by UV-visible spectroscopy. However, in order to confirm which compound was bound to the complex the tetrapyrrole had to be removed from the protein and analysed using HPLC-MS. A number of methods were used to extract the tetrapyrroles from the proteins. In general protein was removed by acidification with equal volumes of 0.1% (v/v) TFA and purification on a TELOS C18 column, eluting in 100% acetonitrile. Extraction with 1% (w/v) TCA allowed removal of the protein for UV-visible analysis; however, this caused degradation of the tetrapyrrole over time and these samples were not suitable for analysis by HPLC-MS.

After production in a trapping strain the *M. barkeri* AhbA/B complex purified with a red colouration as observed previously (Fig. 3.5C). UV-visible and HPLC analysis revealed a signal for haem (data not shown) demonstrating that this complex does not bind any substrate or product.

When overproduced in the trapping strain *D. desulfuricans* AhbA/B complex copurified with a purple colour, consistent with the binding of a substrate or product (Fig. 5.1A). The protein bound UV-visible spectrum showed quenching of the tetrapyrrole signal with broad peaks at 350 nm, 384 nm and 565 nm. After extraction using 1% (w/v) TCA a UV-visible trace characteristic of sirohaem was observed with a large Soret at 384 nm and small α/β bands at

~527 nm and ~583 nm (Figure 5.1). HPLC-MS analysis confirmed the presence of didecarboxysirohaem, as well as a smaller quantity of both sirohaem (~17%) and monodecarboxysirohaem (~3%; Fig. 5.2).

The *D. vulgaris* complex also copurified with a colour after production in the trapping strain. The protein complex was a much darker brown colour (Fig. 5.1B) in comparison to the protein produced previously (Fig. 3.5B). The difference in colour is likely due to the combination of bound haem and product. As with the complex from *D. desulfuricans*, the *D. vulgaris* complex displayed a quenched UV-visible spectrum with peaks at 384 nm and 565 nm. It is unknown why the peak at 350 nm was absent in this complex and what the cause of this peak was in the *D. desulfuricans* complex. After release of the tetrapyrrole from the protein using 1% (w/v) TCA, a UV-visible spectrum was observed identical to that of the product released from the *D. desulfuricans* complex (Fig. 5.1). HPLC-MS of the released product revealed the majority of the bound tetrapyrrole was didecarboxysirohaem, although a small amount of monodecarboxysirohaem (~10%) was also detected (Fig. 5.2).

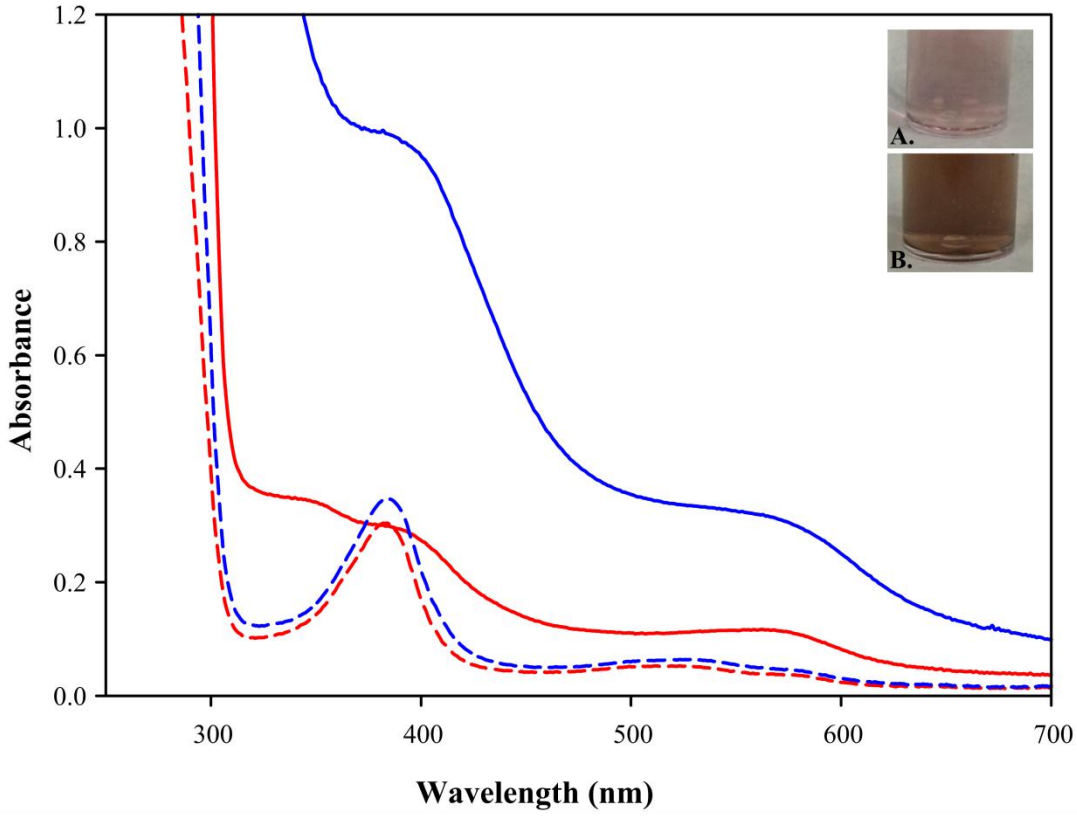


Figure 5.1. UV-visible spectra of *D. desulfuricans and D. vulgaris* AhbA/B complexes from trapping strains. Didecarboxysirohaem absorbance is quenched when bound to the protein (solid) but a characteristic profile is observed upon release using 1% (w/v) TCA (dashed). *D. desulfuricans* AhbA/B = red, *D. vulgaris* AhbA/B = blue. Inset: didecarboxysirohaem bound to *D. desulfuricans* AhbA/B (A.), and *D. vulgaris* AhbA/B.

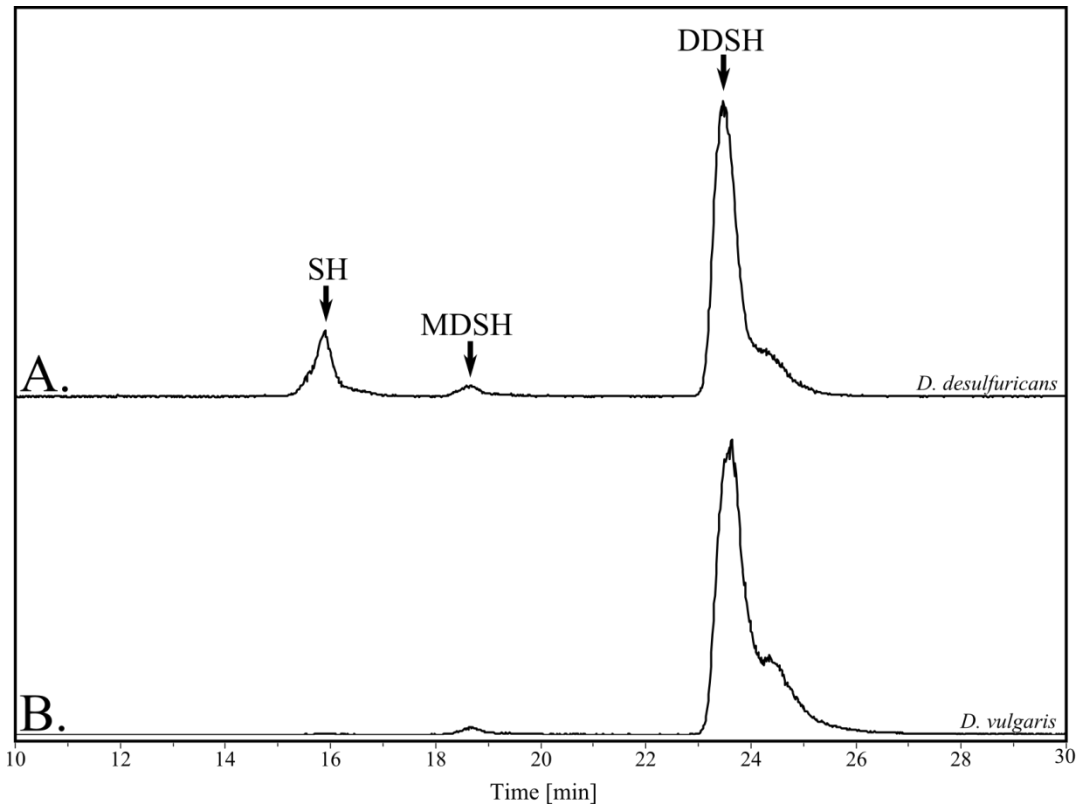


Figure 5.2. HPLC-MS analysis of tetrapyrroles extracted from *D. desulfuricans* and *D. vulgaris* AhbA/B trapping strains. Extracted ion chromatogram traces for masses equivalent to sirohaem (SH), monodecarboxysirohaem (MDSH) and didecarboxysirohaem (DDSH), for samples extracted from *D. desulfuricans* AhbA/B (**A.**) and *D. vulgaris* AhbA/B (**B.**).

5.2.2 Site directed mutagenesis of *D. desulfuricans* AhbA/B

Site directed mutagenesis was performed by Dr. Susanne Schroeder (University of Kent) as described in Section 2.3.7.3. Single amino acid substitutions to alanine were generated individually from pET3a-*ahbA*^{Dd} and pET14b-*ahbB*$^{Dd\text{-}His}$ using the mutagenic primers listed in Table 2.3. Mutations were confirmed by sequencing prior to formation of pET3a-*ahbA/B* and pET3a-*ahbA/B-cysG* constructs using the Link and Lock technique. Constructs were produced with a single mutation in either AhbA or AhbB; however a construct was made with a single mutation in both proteins (Arg119A/Arg102B).

For all activity assays proteins were purified aerobically in Tris-HCl, pH 8 containing 100 mM NaCl using the pull-down method (Section 3.2.3). Elution fractions containing most protein were transferred into the glovebox before buffer exchange into anaerobic Tris buffer. Activity assays were performed with a 10-fold excess of sirohaem that was incubated at 37

ºC for 10 minutes prior to the assay. Reactions were carried out for 10 minutes and overnight at 37 ºC, and stopped with a 1:1 ratio of acetonitrile. Insoluble protein was removed from the sample by centrifugation before analysis by HPLC-MS. Normalisation of data was performed as in Section 3.2.6, with the sum of all peaks representing the total concentration of tetrapyrrole in the system and conversion of each peak into a percentage. A summary of activity data can be found in Table 5.2.

Table 5.2. Activity and binding capabilities of *D. desulfuricans* AhbA and AhbB mutants. Activities given are overall percentage of monodecarboxysirohaem (MDSH) and didecarboxysirohaem (DDSH) after overnight incubations with 10-fold sirohaem. The trapping ratios were excluded for Arg119A and Arg56B due to low concentration. Arg102B and Arg119A/Arg102B were excluded due to haem binding. Wild type *D. desulfuricans* AhbA/B has been included for comparison.

Protein	Site of mutation	Activity (%)		Trapping strain A_{280}:A_{395} x100
		MDSH	DDSH	
AhbA	Arg73	15	23	5
AhbA	Arg74	15	31	4
AhbA	His115	1	0	3
AhbA	Tyr117	22	1	5
AhbA	Arg119	2	0	-
AhbB	Arg55	12	43	3
AhbB	Arg56	10	15	-
AhbB	His63	17	83	6
AhbB	His98	3	0	5
AhbB	Tyr100	21	42	12
AhbB	Arg102	11	0	-
AhbB	His118	39	39	5
AhbA/B	Arg119/Arg102	2	0	-
Wild type AB	-	0	96	20

After overnight incubation mutations Arg73A and Arg74A displayed 23% and 31% conversion to didecarboxysirohaem, respectively, showing that they formed an active complex. Mutations to the conserved HXYXR motif in the active site of AhbA (His115A, Tyr117A and Arg119A) displayed virtually no turnover (Tyr117A ~1%) to didecarboxysirohaem. The Tyr117A mutant still retained the ability to perform a single decarboxylation with 22% monodecarboxysirohaem being present after overnight incubation. His115A and Arg119A showed minimal conversion of sirohaem to monodecarboxysirohaem (1% and 2%, respectively).

Homologous mutations in the β1 region of AhbB (Arg55B and Arg56B), retained activity, although with reduced turnover (43% and 15% conversion to didecarboxysirohaem, respectively). As with AhbA, mutations of His98B and Arg102B from the HXYXR resulted in the inhibition of didecarboxysirohaem formation. His98B displayed a small degree of monodecarboxysirohaem production, while Arg102B produced a higher quantity (11%). Unlike Tyr117A, mutation of Tyr100B did not prevent the production of didecarboxysirohaem with 42% conversion after overnight incubation. His63B and His118B also resulted in an active decarboxylase capable of both decarboxylation reacions. Surprisingly, the His63B mutant complex exhibited a unique property. After 10 minutes the His63B mutant had converted 90% of the sirohaem to monodecarboxysirohaem. Overnight incubation resulted in majority conversion to didecarboxysirohaem. The double arginine mutant, Arg119A/Arg102B, was also inactive with only 2% monodecarboxysirohaem being produced overnight.

All mutant complexes for enzyme trapping experiments were grown as described previously using culture volumes of 500 mL or 1 litre (Section 5.2.1). The majority of the mutant proteins purified with coloured compounds, with only Arg119A appearing colourless. The lack of colour was likely due to extremely low protein concentration. The low protein concentration is believed to be caused by an experimental error during purification as this protein complex was overproduced previously for activity assays without issue. Due to time constraints repetition of the growth and purification could not be performed.

A range of colours was observed for all of the mutant complexes other than Arg119A (Fig. 5.3), with the majority being purple or blue in colour. Arg102B purified with a light brown colour and the double mutant Arg119A/Arg102B purified with a deep red, reminiscent of *M. barkeri* AhbA/B.

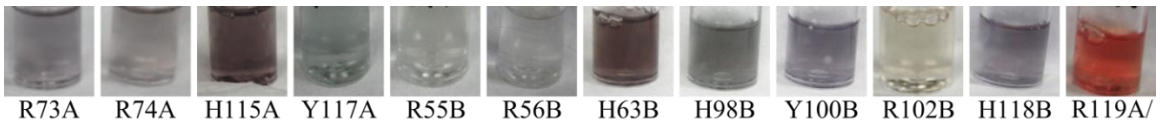


Figure 5.3. Purified mutant *D. desulfuricans* AhbA/B complexes from trapping strains. Mutations labelled with mutated amino acid, residue number and protein subunit (*i.e.* AhbA or AhbB). Arg119A was excluded due to low protein concentration.

UV-visible analysis revealed the majority of proteins had a Soret peak at 385-395 nm. Many of the mutants displayed more prominent Soret peaks in comparison to those observed for wild type *D. desulfuricans* AhbA/B, with some of these peaks slightly shifted in comparison (Fig. 5.4). The ratio of A_{280} to A_{395} provided an estimation of the trapping ability of the protein complex. The redox state of the metal in tetrapyrroles frequently affects their absorbance characteristics. This may explain the slight variations in colours and spectra observed for the mutant proteins. Similarly, the A_{280} to A_{395} can only be used as a guideline as redox state may affect these values. Unfortunately, the concentrations of the Arg119A and Arg56B protein complexes were too low to accurately determine an A_{395}. Tyr100B appears to retain the greatest quantity of tetrapyrrole with over half that observed for wild type AhbA/B. The majority of the mutants appear to have reduced binding capabilities with 15-25% binding in comparison to wild type. Interestingly the UV-visible spectra of Arg102B and the double mutant Arg119A/Arg102B revealed the presence of a haem-like molecule, rather than sirohaem (Fig. 5.5). Arg102B has a Soret band at 412 nm and a broad region between 525 nm and 620 nm where α/β peaks are not clearly defined. The Arg119A/Arg102B double mutant has a much more distinct Soret peak at 424 nm and clearly defined α/β peaks at 526 nm and 558 nm. The haem spectrum for the double mutant is closer to that of the haem-loaded *D. vulgaris* AhbA/B complex (peaks: 423nm, 527 nm and 555 nm) than the single mutant which is possibly due to the oxidation state of the haem. Unfortunately, due to technical problems and time constraints HPLC-MS analysis of the bound tetrapyrroles could not be performed.

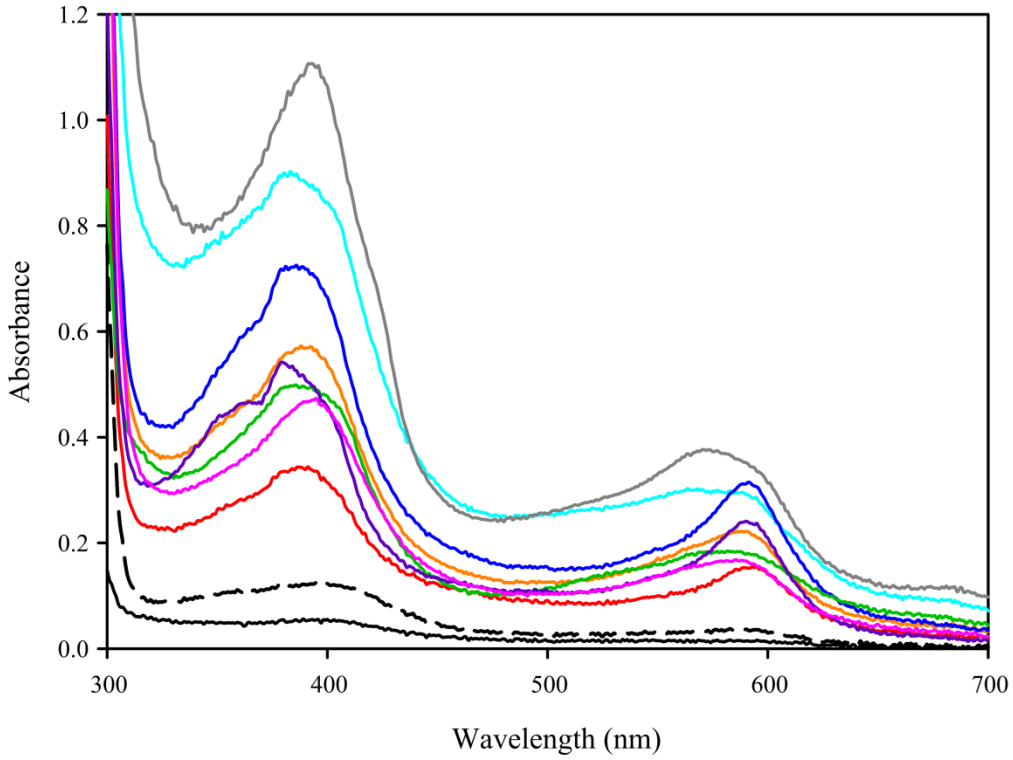


Figure 5.4. UV-visible spectra of mutant *D. desulfuricans* AhbA/B complexes trapping sirohaem derivatives. Arg73A (orange), Arg74A (pink), His115A (grey), Tyr117A (red), Arg55B (black, solid), Arg56B (black, dashed), His63B (yellow), His98B (blue), Tyr100B (green) and His118B (purple). Arg119A was excluded due to low protein concentration.

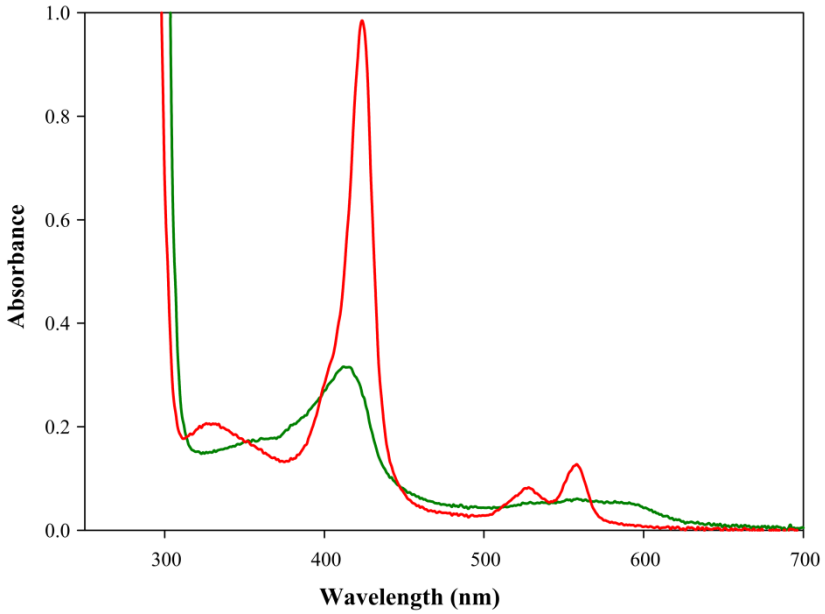


Figure 5.5. UV-visible spectra of mutant AhbA/B *D. desulfuricans* complexes trapping haem. Mutant AhbA/B complexes that trap haem-like compounds: Arg102B (green) and Arg119A/Arg102B (red).

5.2.3 Formation of novel chimeric AhbA/B complexes

There is a high degree of identity between AhbAs and AhbBs from different species (Table 5.1). Therefore, the possibility of chimeric complex formation was investigated. Looking at the protein sequences there were no obvious regions that related to properties such as tetramerisation of the *D. vulgaris* proteins or haem binding of the *D. vulgaris* and *M. barkeri* complexes. The use of chimeric complexes aimed see if these properties could be attributed to a single subunit and if activity could be maintained using an AhbA and an AhbB from different organisms. Constructs were made with pairs of *ahbA* and *ahbB* genes from different organisms using the Link and Lock technique (Table 2.2). In every case the *ahbB* gene was fused to a N-terminal His_6-tag. The proteins were overproduced in BL21 (DE3) as described previously (Section 2.2.7) and purified using the pull-down technique via nickel affinity chromatography using Tris-HCl, pH 8 containing 100 mM NaCl.

D. vulgaris AhbA and *D. desulfuricans* AhbB could be copurified as a complex and exhibited a similar light brown colouration to the native *D. vulgaris* AhbA/B complex (Figs. 5.6 and 3.5). The UV-visible spectrum revealed a haem-like signal with a Soret peak at 423 nm and small α/β bands at 521 nm and 552 nm, respectively (Fig. 5.6). The UV-spectrum is again reminiscent of the native *D. vulgaris* AhbA/B complex with a slight shift in the α/β bands (*D. vulgaris* peaks: 423nm, 527 nm and 555 nm; Fig. 3.7). Gel filtration analysis of the *D. vulgaris*/*D. desulfuricans* AhbA/B complex revealed the presence of both dimeric and tetrameric complexes with peaks at 41 kDa and 97 kDa (Fig. 5.7). Full activity was observed during *in vitro* assays using the *D. vulgaris*/*D. desulfuricans* AhbA/B complex incubated with sirohaem overnight at 37 ºC (Fig. 5.8).

D. desulfuricans AhbA was also found to copurify with *D. vulgaris* AhbB. Intriguingly, the *D. desulfuricans*/*D. vulgaris* AhbA/B complex complex was colourless (Fig. 5.6). Furthermore, gel filtration analysis demonstrated that the *D. desulfuricans*/*D. vulgaris* AhbA/B complex only exists as a dimer with an estimated mass of 41 kDa (Fig. 5.7). As with the other *Desulfovibrio* chimera, the *D. desulfuricans*/*D. vulgaris* AhbA/B complex was active *in vitro* when incubated with sirohaem overnight at 37 ºC (Fig. 5.8). Combinations of *Desulfovibrio* and *M. barkeri* proteins did not produce soluble complexes despite 35-40% identity in the sequences.

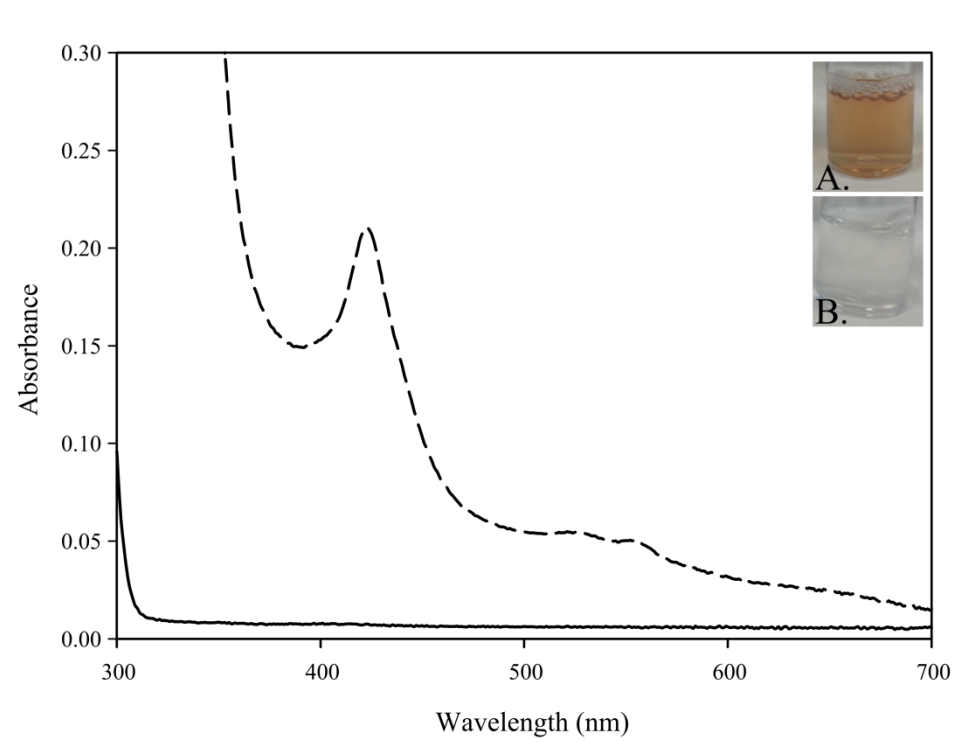


Figure 5.6. UV-visible spectra of the purified chimeric AhbA/B complexes. *D. vulgaris/D .desulfuricans* (dashed) and *D. desulfuricans/D. vulgaris* AhbA/B chimeras reduced using excess sodium dithionite. Inset: pictures of purified protein solutions of *D. vulgaris/D .desulfuricans* (**A.**) and *D. desulfuricans/D. vulgaris* (**B.**) AhbA/B chimeras.

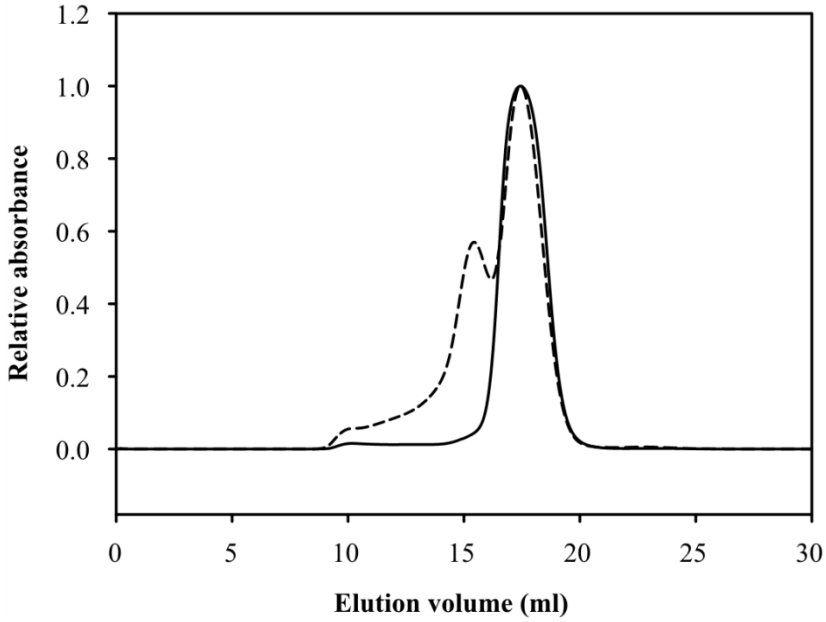


Figure 5.7. Gel filtration UV trace, recorded at a wavelength of 280 nm, of purified chimeric AhbA/B complexes. *D. desulfuricans* AhbA with *D. vulgaris* AhbB (solid) appears as a single heterodimeric species whereas *D. vulgaris* AhbA with *D. desulfuricans* AhbB (dashed) is present as both a dimer and tetramer.

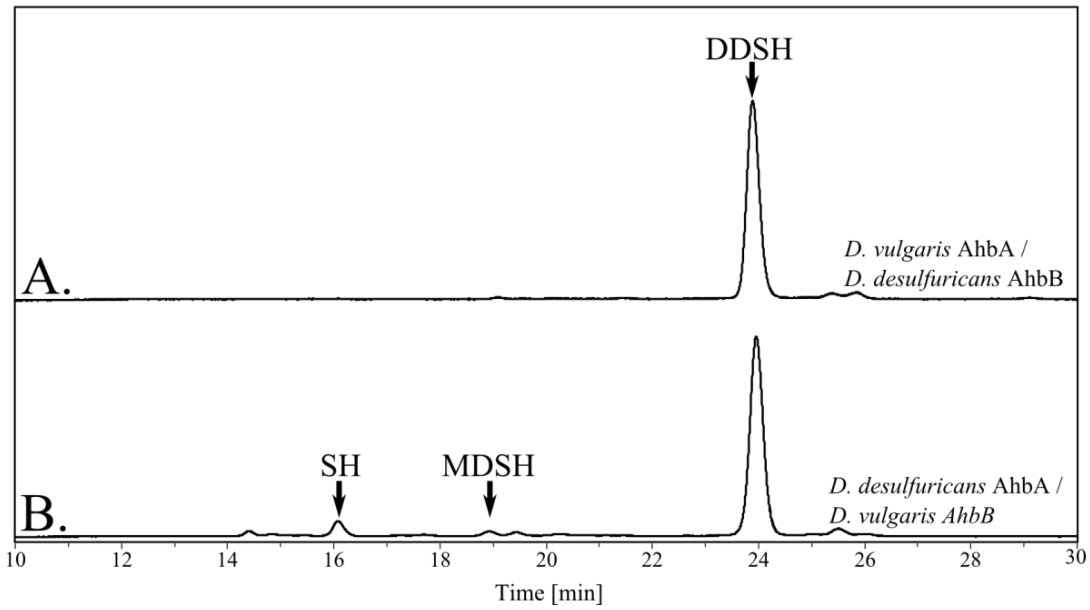


Figure 5.8. HPLC traces, recorded at a wavelength of 380 nm, of sirohaem incubated with purified AhbA/B chimeric complexes. A. *D. vulgaris* AhbA with *D. desulfuricans* AhbB, **B.** *D.desulfuricans* AhbA and *D. vulgaris* AhbB. Arrows indicate peaks for sirohaem (SH; ~16 min), monodecarboxysirohaem (MDSH; ~19 min) and didecarboxysirohaem (DDSH; ~24 min).

5.3 Discussion

The work presented in this chapter has furthered the understanding of the AhbA/B complexes from *D. desulfuricans*, *D. vulgaris* and *M. barkeri*, through enzyme trapping and site-directed mutagenesis experiments as well as production of novel chimeric AhbA/B complexes.

The AhbA/B complexes from *D. desulfuricans* and *D. vulgaris* copurify with didecarboxysirohaem when overproduced with excess CysG. Product binding by the AhbA/B complexes from *Desulfovibrio* species may be related to the need to protect the unstable product from the intracellular environment, especially as sulphate reducers are known to grow in oxygen exposed environments (Risatti *et al.*, 1994). Didecarboxysirohaem may be handed on to the next enzyme in the pathway, AhbC, by substrate channelling to prevent substrate oxidation/degradation. It was shown that, unlike the sulphate reducers, the AhbA/B complex from *M. barkeri* does not bind its product. The lack of product binding in *M. bakeri* could be because this organisms thrives in more anaerobic conditions (Thauer, 1998), and the cellular environment is suitable for didecarboxysirohaem to be in solution without risk of oxidation or degradation.

Site directed mutagenesis has been performed on the AhbA/B enzymes from *D. desulfuricans*. Mutations of the arginines in the dimerisation region (73A, 74A, 55B and 56B) caused reduced activity. However, they do not appear essential and this reduction is likely due to a general perturbation of the system. Similarly, His63B and His118B displayed reduction in activity, most likely for the same reasons. Interestingly, His63B caused a build-up of monodecarboxysirohaem after 10 minutes, indicating a reduction in the rate of the second decarboxylation reaction. His63B is located in the disordered loop region at the top of the active site cleft and is too far away from the sites of decarboxylation (carbons C12 and C18) to be catalytic. Therefore His63B may be involved in facilitating the migration, flipping or positioning of the tetrapyrrole between decarboxylation reactions. The removal of this residue may therefore reduce the efficiency of this step causing a significantly reduced rate of reaction for the second decarboxylation.

Mutations of residues in the HXYXR motif in AhbA (His115A, Tyr117A and Arg119A) prevented the production of didecarboxysirohaem. His115A and Arg119A produced minimal monodecarboxysirohaem (1-2%) after an overnight incubation suggesting they were incapable of performing both reactions. His115A and Arg119A may be involved in both decarboxylations; alternatively they may catalyse the initial decarboxylation which may then prevent the second decarboxylation from occurring. It is currently unknown if there is a specific order of decarboxylation. The initial decarboxylation may be required to allow the tetrapyrrole to flip/migrate into the position for the second decarboxylation. If the first decarboxylation is inhibited by mutation the flipping/migration process may be prevented. Therefore, what appears to be involvement in both reactions may be a secondary effect from inhibition of the first reaction. To confirm this hypothesis, monodecarboxysirohaem could be incubated with the mutants. However, both 12-monodecarboxysirohaem and 18-monodecarboxysirohaem would have to be tested, and production of these intermediates is not currently possible by enzymatic means. Tyr117A appears to accrue monodecarboxysirohaem (22% after overnight incubation) suggesting that mutation of this residue may prevent the second decarboxylation from occurring, although its exact role is yet to be determined.

Some of the mutations in the HXYXR motif of AhbB have a different effect to the homologous mutations in AhbA. For example, Tyr100B maintained 42% activity and therefore does not appear to be essential. Arg102B had greater yield of monodecarboxysirohaem (11%), but did not produce any didecarboxysirohaem. This would suggest the Arg102B protein capable of the first decarboxylation, although at a reduced

level, but it is not able to carry out a second decarboxylation. His98B produced a small quantity of monodecarboxysirohaem (3%), which close to the values observed for His115A and Arg119A (1% and 2%, respectively), suggesting this residue is either essential for both reactions or the initial decarboxylation for reasons described above. The double mutation Arg119A/Arg102B only produced 2% monodecarboxysirohaem after overnight incubation, further supporting the role of Arg119A in the initial decarboxylation reaction.

In Chapter 4 two reaction mechanisms were proposed based on the structure of AhbA/B with didecarboxysirohaem bound. The flipping mechanism (Fig. 4.15) requires Arg102B to be the catalytic residue responsible for initiation of both decarboxylation reactions, with the tetrapyrrole leaving the active site and flipping 180º about the C5/C15 axis between decarboxylations. However, mutation of Arg102B produced monodecarboxysirohaem, suggesting involvement in only one of the two decarboxylation reactions. The migration mechanism uses the catalytic centres provided by both proteins with Arg119A and Arg102B catalysing a single decarboxylation each. The mutant activity data therefore support the active site migration mechanism.

For the first decarboxylation reaction His98B would coordinate with the iron. His115A could facilitate double bond rearrangement and initial protanation/deprotonation of C19. Arg119A could then stabilise carboxylic acid of the acetic acid on C18, allowing decarboxylation to occur whilst providing a proton for the developing methyl group. The monodecarboxylated intermediate could then unbind/migrate to the second active site using His115A to coordinate the iron in the macrocycle, His98B to facilitate protonation/deprotanation of C11 and Arg102B as the catalytic residue. Tyr117A may play a role in migration of the intermediate, hence why the second decarboxlation reaction is affected.

As mentioned in Chapter 4, the structure of the AhbA/B homologue NirD/L from *H. thermophilus* was published during the course of this work (Haufschildt *et al.*, 2014). Site directed mutagenesis was performed on a selection of amino acids within this protein, including equivalent residues to His115A, Tyr117A, His98B and Tyr100B. His93 (His115A equivalent) was shown to be essential for activity, whereas mutations of His226 (His98B equivalent) and the tyrosine residues affected substrate binding but not activity. Surprisingly, the arginine residues from the HXYXR motif in NirDL were not mutated in the study and therefore it is unknown whether these residues are required for the decarboxylase activity of NirDL.

Tetrapyrrole binding was examined using the mutant AhbA/B proteins and the enzyme-trap approach for tetrapyrrole isolation. A variety of colours were observed (Fig. 5.3), most likely due to the varying concentrations of protein and oxidation states of the bound tetrapyrroles. Of the complexes studied the majority displayed occupancies 15-30% to that of the wild type (based on A_{280}:A_{395} ratio). The Arg102B and Arg119A/Arg102B mutants displayed haem-like spectra. The presence of haem suggests advantageous haem uptake during overproduction. Addition of ALA to the cultures naturally causes an increase in haem within the cells, as well as the desired increase in sirohaem, increasing the availability of free haem for these complexes to bind. The alteration of positive arginines to small hydrophobic alanines would affect the overall charge of the active site cleft. The positive charge of the wild type active site would normally draw in the negative charges of the acetate and propionate side chains of sirohaem. Producing a more hydrophobic cleft appears to reduce the specificity of the active site. Haem lacks the acetate side chains and is therefore more hydrophobic, allowing it to bind more easily in the modified active site.

Due to the technical problems experienced and time constraints, analysis of the bound tetrapyrroles could not be performed. As the UV-visible spectra of sirohaem and its decarboxylated derivatives are near identical the only way to determine what is bound is by extraction and HPLC-MS analysis. Analysis of these bound molecules would provide further insight into the reaction mechanisms proposed, especially if some of the mutants trap sirohaem or monodecarboxysirohaem. EPR analysis of histidine mutants would also provide information about the coordination of the iron in the tetrapyrrole/enzyme complexes.

Novel chimeric complexes were produced using the AhbA and AhbB proteins from *D. desulfuricans* and *D. vulgaris*. In both cases the pull-down method of protein purification (using His_6-tagged AhbB) allowed copurification of the chimeric AhbA/B complex. These novel chimeric complexes were equally active in comparison with the native complexes, demonstrating the proteins are capable of acting cooperatively to decarboxylate sirohaem. Whilst this may be expected given the high sequence identity (72% and 67% for AhbA and AhbB, respectively), these experiments allowed the tetramerisation and haem binding properties of the *D. vulgaris* complex (described in Chapter 3) to be attributed to a single subunit. The chimeric complex containing AhbA from *D. vulgaris* forms both tetrameric and dimeric species whereas the complex with *D. desulfuricans* AhbA is only present as a dimer. Similarly, the complex containing AhbA from *D. vulgaris* was able to bind haem during overproduction in *E. coli*. These data suggest that both the haem binding and tetramerisation domains reside in the AhbA subunit of the *D. vulgaris* AhbA/B complex. No chimeric complexes involving Ahb proteins from *M. barkeri* could be purified. The reason for the lack

of compatibility between the *M. barkeri* AhbA/B proteins and those from the *Desulfovibrio* species is unclear. Structural predictions using the Phyre2 server (Kelley & Sternberg, 2009) show a similar fold with no obvious extensions or areas or alternative secondary structure other than the first and final β-sheets of each protein which are predicted to be loops (Fig. 5.9). The loops regions may be due to the protein folds being predicted independently before being superimposed on the known AhbA/B structure. It is likely that the poor interactions are a result of sequence differences in the dimerisation regions. However, due to these regions being poorly predicted in the model it is hard to ascertain if this is the cause.

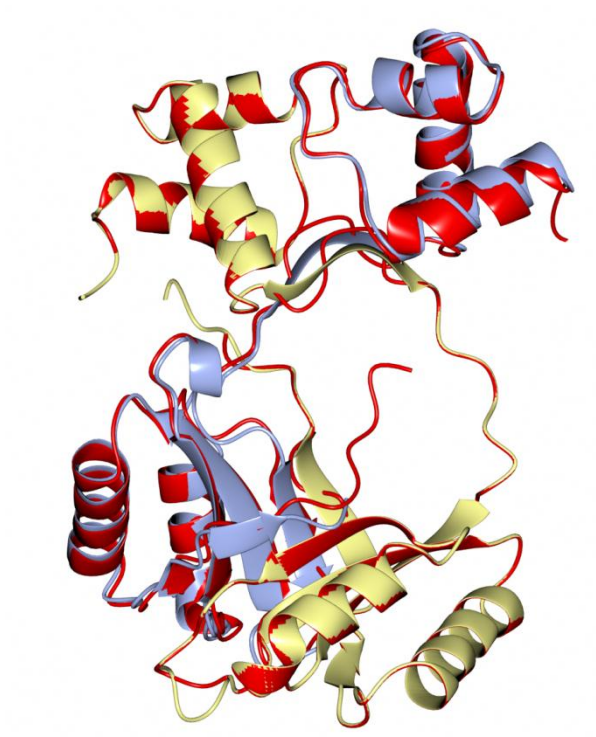


Figure 5.9. Structural prediction of the *M. barkeri* AhbA/B complex (red) produced using the Phyre2 server superimposed on *D. desulfuricans* AhbA/B. Structural predictions were performed using individual AhbA and AhbB protein sequences from *M. barkeri* and therefore are unable to predict β1 and β6 regions that likely form interactions between the two subunits.

In summary, these data have demonstrated that the AhbA/B complexes from the *Desulfovibrio* species are able to tightly bind their product. The mutation of residues in the active site has supported the dual action of pseudo-symmetrical active sites resulting in the double decarboxylation of sirohaem with the migration of a monodecarboxylated intermediate between reactions. It has also been suggested that Arg119A and Arg102B not only play a catalytic role but also allow specificity for the negatively charged groups of the substrate and monodecarboxylated intermediate. Finally, production of chimeric complexes

revealed AhbA from *D. vulgaris* is responsible for formation of tetrameric complexes as well as the ability to advantageously bind haem.

Chapter 6

Discussion

6.1 Which tetrapyrrole is the most ancient cofactor?

The results presented in this work also contribute to the discussion concerning the evolution of the modified tetrapyrroles. There are many arguments as to which tetrapyrrole is the most ancient cofactor, a question which is yet to be given a definitive answer. Cobalamin was once envisaged to be the oldest modified tetrapyrrole; based on the nucleotide loop and involvement in ribonucleotide reduction, attributes that were suggested to be relics of the RNA world (Benner *et al.*, 1989, Dickman, 1977). The main roles associated with cobalamin are methylation reactions and radical generation for catalysis of rearrangement reactions (Roth *et al.*, 1996). Both of these reactions can be carried out by radical SAM enzymes (Atta *et al.*, 2010, Frey *et al.*, 2008) and given the simplicity of SAM and the early evolution of iron sulphur clusters (Mendel *et al.*, 2007), radical SAM enzymes likely predate cobalamin-dependent processes. The earlier evolution of radical SAM enzymes suggests there was no evolutionary driving force for cobalamin in early life. The nickel containing coenzyme F_{430} is found exclusively in methanogens and contains a highly reduced ring system (Thauer, 1998). Whilst it has been considered to have a role in early biotic life (Benner *et al.*, 1989) the enzymes involved in the pathway remain elusive restricting further understanding of the evolution of this pathway.

Though haem and chlorophylls were once considered related to aerobic metabolism, there is evidence of their early (pre-oxygenic) origin. There is a strong indication that photosynthesis evolved three billion years ago, within half a billion years of the origin of life, suggesting very early evolution of chlorophylls (Tice & Lowe, 2006). Haem containing terminal oxidases are likely to predate oxygenic photosynthesis according to sequence and phylogenetic analysis (Schafer *et al.*, 1996). The discovery of the alternative haem biosynthesis pathway, which operates under anaerobic conditions, further supports the existence of haem in ancient life (Akutsu *et al.*, 1993, Bali *et al.*, 2011).

Sirohaem, however, is a prime candidate as the most ancient modified tetrapyrrole. With the shortest biosynthesis pathway of all the modified tetrapyrroles, sirohaem is the easiest to make biosynthetically (Warren & Scott, 1990). Some obligate anaerobes lack cytochromes but contain sirohaem, suggesting the use of sirohaem as a primitive alternative to haem (Crane & Getzoff, 1996). In this way, dissimilatory sulphite reduction coupled to electron transport facilitated energy generation before oxidative respiration was possible. The requirement of sirohaem as an intermediate of alternative haem and haem d_1 synthesis (Bali *et al.*, 2014) further promotes the antiquity of this cofactor. Whilst current research has still

not come to a conclusion on which modified tetrapyrrole represents the most ancient specimen, the evidence for sirohaem being one of the forerunners is undeniable.

6.2 Development of the haem biosynthesis pathways

The discovery of the alternative haem biosynthesis pathway raises questions as to how and why Nature derived two very different methods for the synthesis of the same molecule. Dissimilatory sulphate reduction represents one of the oldest energy conserving respiratory systems on Earth and proceeds via the highly toxic intermediate sulphite (Susanti & Mukhopadhyay, 2012). Organisms using this process employ a dissimilatory sulphite reductase in order to combat this toxicity. Sulphite reductases belong to two groups, the simplest of which utilises sirohaem as a cofactor. Therefore, as mentioned previously, early prokaryotic life probably contained sirohaem. Given the anaerobic nature of alternative haem biosynthesis and the group of organisms that utilise it, the alternative pathway likely represents a primitive pathway for haem biosynthesis. The alternative pathway was probably formed via the patchwork model of pathway evolution. The patchwork model describes the fortuitous development of new intermediates by the action of broad specificity enzymes, followed by duplication and specialisation of advantageous enzymes resulting in formation of novel pathways (Ycas, 1974, Jensen, 1976). Through the action and evolution of promiscuous enzymes, modifications to sirohaem could have occurred eventually leading to the new, more stable tetrapyrrole, haem. With its broader functionality haem has been adopted as the prosthetic group for a plethora of proteins across all domains of life. Prior to classical haem and chlorophyll biosyntheses all modified tetrapyrroles would have stemmed from the intermediates precorrin-2 and sirohydrochlorin. An overview of the branched tetrapyrrole biosynthesis pathways is shown in Figure 6.1. Most of the intermediates of the alternative haem pathway are oxygen sensitive. Therefore, as atmospheric oxygen increased there was a driving force for Nature to evolve a new pathway with aerobically stable intermediates.

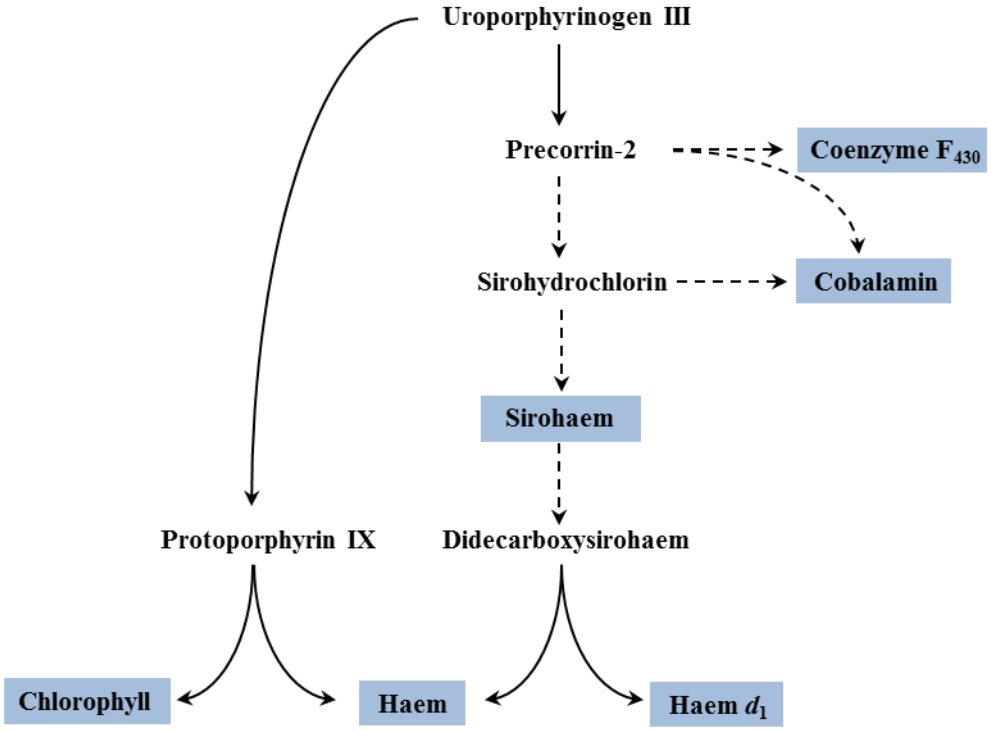


Figure 6.1. The known branch points of the tetrapyrrole biosynthesis pathways. Potential points of regulation between coenzyme F_{430}, cobalamin, sirohaem, haem, and haem d_1 are displayed as dashed lines.

The cobalamin biosynthesis pathway similarly evolved both aerobic and anaerobic routes (Holliday *et al.*, 2007). Cobalamin biosynthesis branches from a common precursor (precorrin-2) and converges again at adenosylcobyric acid before final modifications are made (Raux *et al.*, 1999b). Many of the enzymes in the cobalamin pathways also share a high degree of sequence similarity (Roth *et al.*, 1996, Blanche *et al.*, 1993). In contrast, the haem biosynthesis pathways do not reconvene and none of the enzymes involved share sequence homology. However, the modifications to the ring system are carried out in a near identical fashion during the classical and alternative haem biosynthesis pathways. The similarities in side chain modifications demonstrate a chemical precedence for these reactions, which allowed the evolution of very different enzymes that perform similar reactions.

The evolution of the classical pathway also resulted in a more energetically favourable method of haem synthesis. The alternative pathway first requires production of sirohaem from uro'gen III, consuming 2 SAM molecules and NAD^+. Although the initial action of

AhbA/B does not require any cofactor (at least in sulphate reducing bacteria), completion of haem synthesis via AhbC and AhbD requires further use of SAM (Lobo *et al.*, 2014, Bali *et al.*, 2011). Total SAM consumption from uro'gen III to haem via the alternative pathway lies in the region of 4-6 molecules (quantitative studies with AhbC and AhbD are yet to be performed) and 1 NAD^+. In contrast, the synthesis of haem along the classical pathway requires fewer cofactors. Here, HemE catalyses four decarboxylations of uro'gen III in a cofactor-independent manner, essentially carrying out the roles of AhbA/B and AhbC. Conversion of the northern propionate groups to vinyl groups is either catalysed by the SAM-dependent enzyme HemN or the cofactor-independent membrane associated enzyme HemF. Ring oxidation by HemY/G requires FAD or FMN as a cofactor (it is unknown what cofactor is used by HemJ). Finally, chelation of iron into the macrocycle is another cofactor-free reaction. In total the classical pathway only requires FAD/FMN, with the addition of 2 molecules of SAM for organisms using HemN to convert copro'gen to proto'gen. Therefore there is a significant energetic advantage for organisms using the classical haem synthesis pathway.

6.3 The similarities and differences of HemE and AhbA/B

HemE catalyses four successive decarboxylations of uro'gen III (at carbons C3, C8, C12 and C18) to form copro'gen III (Fig. 1.6). The decarboxylations at carbons C12 and C18 are analogous to those carried out on sirohaem by the AhbA/B complex. HemE functions without use of a cofactor as do the AhbA/B complexes from *D. desulfuricans* and *D. vulgaris*; whilst it is unknown if the haem associates with the *M. barkeri* AhbA/B complex functions as a cofactor or as regulatory mechanism in an indirect fashion. Despite these similarities, the HemE and AhbA/B proteins do not share sequence or structural similarity (Fig. 6.2A/C). HemE forms a homodimer; however each monomer is a catalytic unit capable of binding one uro'gen III molecule (Phillips *et al.*, 2003). Whilst similar in size to the AhbA/B complex, the 40.8 kDa monomeric protein is comprised of a single domain with $(\beta/\alpha)_8$-barrel fold (Whitby *et al.*, 1998). The active site cleft is formed by loops at the C-terminal ends of the barrel strands. The product bound complex (PDB 1R3Y) shows that copro'gen III adopts a domed conformation with the active site with Asp86 (human numbering) coordinating a pyrrole NH group conferring stability. The substrate binding pocket is overall positively charged with several conserved arginine residues, some of which coordinate with the substrate propionate groups (Fig. 6.2B). One of these conserved arginines (Arg37 in humans) has been attributed to catalysis, playing a direct role in the decarboxylation process (Lewis & Wolfenden, 2008).

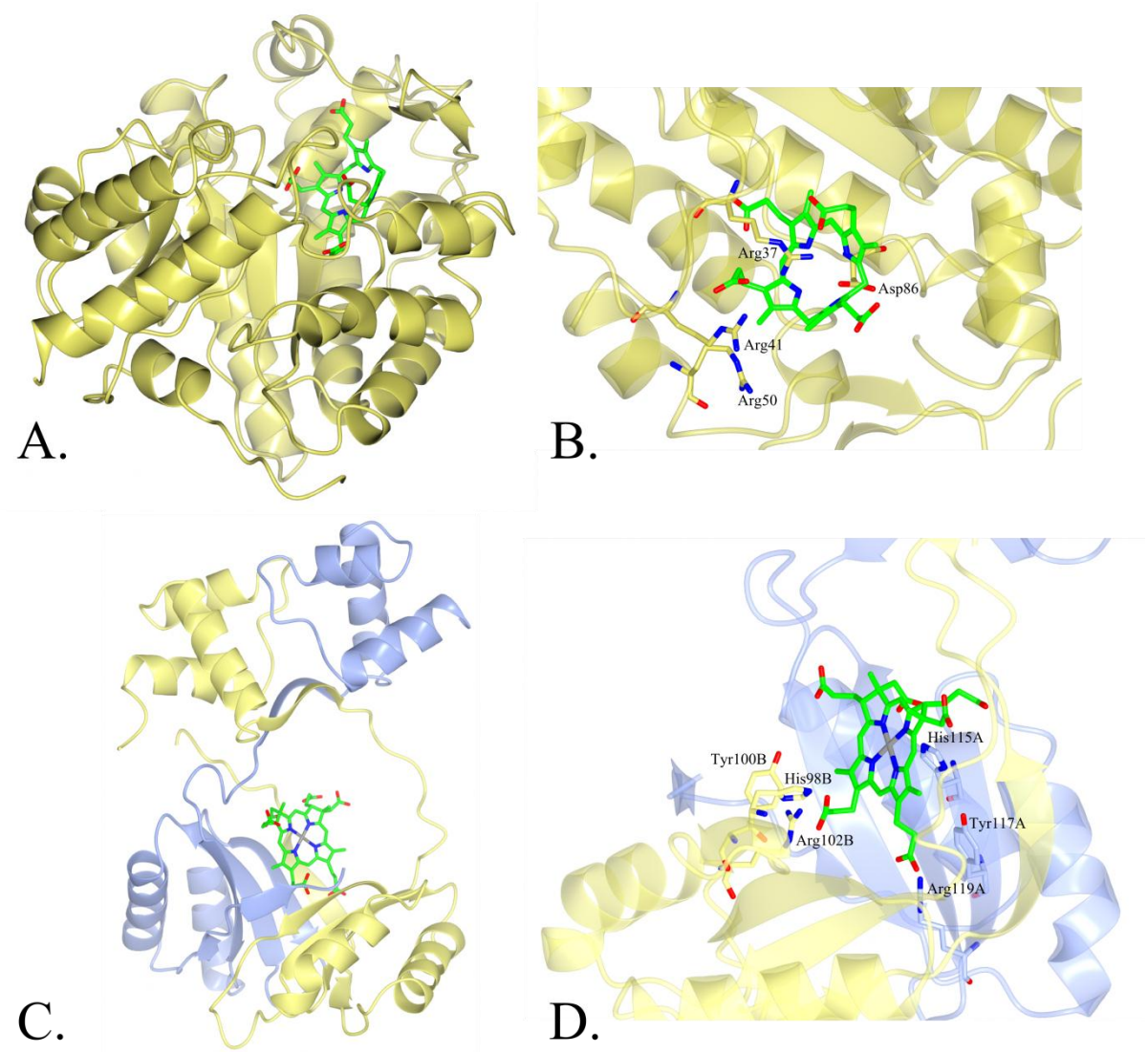


Figure 6.2. Product bound structures of human HemE (PDB 1R3Y) and *D. desulfuricans* AhbA/B (PDB 4UN1). A. Overall view of the human HemE protein with bound copro'gen III (atom coloured), **B.** Close-up of active site of human HemE with conserved residues (Arg37, Arg41, Arg50 and Asp86) and copro'gen III displayed as sticks, **C.** Overall view of *D. desulfuricans* AhbA/B with bound didecarboxysirohaem (atom coloured), **D.** Close-up of active site of *D. desulfuricans* AhbA/B with conserved residues (His115A, Tyr117A, Arg119A, His98B, Try100B and Arg102B) and didecarboxysirohaem displayed as sticks.

The crystal structure of *D. desulfuricans* AhbA/B revealed that is has an entirely different fold to that of HemE. The heterodimer is formed of two domains, an N-terminal head domain consisting of two HtH motifs and a C-terminal RAM-like domain where the active site cleft is located (Fig. 6.2C/D). The two proteins interact at the start of the linker regions,

the base of the RAM-like domains and finally at the side of the RAM-like domains, where the terminal β-strands form a β-sheet with the opposing proteins. Both AhbA and AhbB share a near identical fold with the only exception being a short helical region in AhbA, located in the linker region. The different linkers appear to effect the orientation of the HtH domains, with AhbA being angled towards that of AhbB.

Previous research surveying archaeal genomes for the *ahb* genes revealed that all of the archaea that contain the alternative haem biosynthesis pathway either have both *ahbA* and *ahbB* or and *ahbA/B* fusion (Storbeck *et al.*, 2010), demonstrating the essential requirement for both genes. The presence of *ahbA* and *ahbB* is likely due to a gene duplication event and subsequent sequence variation, producing an asymmetric dimer which provides a match of rigidity on one side of the dimer and an open conformation on the other, allowing access to the active site. The difference in the linker regions appears to be essential in forming this asymmetry. A homodimer of AhbA would cause clashes with the HtH domains as they adopt a more upright conformation with a more ordered linker region. The linker regions would also block access to the active site cleft on both sides of the complex. Conversely the flexible linkers of AhbB would likely produce more unstable homodimer as well as generating a much more solvent accessible active site cleft. The close association of the two Ahb proteins explains why the individual proteins were so unstable during purification (Section 3.2.1). Without their counterpart the individual proteins would have exposed hydrophobic areas causing aggregation of the proteins. Whilst the *in vitro* activity assays of individual AhbA and AhbB proteins demonstrated their ability to perform the decarboxylation reactions to varying degrees of efficiency, combinations of AhbA and AhbB displayed full activity and supports the essential nature of both proteins for a fully functional sirohaem decarboxylase (Section 3.2.2). From the structure of the active site it is likely that the activity observed with individual proteins was catalysed by poorly formed homodimers.

As with HemE, the base of the active site cleft of AhbA/B contains a positive charge provided by Arg119A and Arg102B. Arg119A interacts with the propionate group of ring D in the product bound structure and Arg102B is within range of C12, where the propionate leaving group would be located prior to decarboxylation. Unlike copro'gen III, didecarboxysirohaem adopts a much flatter conformation in the active site with the sites of decarboxylation located within the cleft (Fig. 6.2D). His115A plays a similar role to Asp86 in HemE, however instead of coordinating a pyrrole nitrogen it coordinates the iron in the macrocycle that is not present in copro'gen III. Sequence similarity suggested the role of a conserved HXYXR motif in both AhbA and AhbB. Site directed mutagenesis of the HXYXR motif in both proteins provided evidence of the involvement of all of these

conserved residues except Tyr100B. From these data Arg102B does not appear to be the sole catalytic residue as the mutated protein was capable of producing monodecarboxysirohaem, supporting the migration mechanism of decarboxylation (described in Chapter 4). What is remarkable about the use of Arg102B (and probably Arg119) as catalytic residue(s) is the fact that, despite the differences in sequence and structure between HemE and AhbA/B, Nature has evolved two enzymes that enhance the decarboxylations of kinetically stable substrates by very similar mechanisms.

The work presented in this thesis has provided an insight into the active site of the AhbA/B complex and the residues that are involved in product binding and catalysis. However, some questions remain unanswered, including whether there is a specific order of decarboxylation and the exact reaction mechanism utilised by AhbA/B. The active site mutants generated provide an excellent tool to address these questions. It is unknown whether His115A, Tyr117A and Arg119A are essential for both decarboxylation reactions or if loss of function is due to the inhibition of the initial decarboxylation. To test this hypothesis monodecarboxysirohaem must be isolated and then assayed with these mutant AhbA/B complexes. If turnover is achieved then this would support the migration mechanism of decarboxylation and it may be possible to ascertain an order of decarboxylation. Mutagenesis also revealed the unexpected involvement of His63B located in the linker region of AhbB. Analysis of this mutation demonstrated an overproduction of monodecarboxysiorhaem after 10 minutes, which was then converted to didecarboxysirohaem after overnight incubation. His63B is therefore expected to play a role in substrate migration or flipping. Use of the His63B mutant complex can be employed for the production of monodecarboxysirohaem which can be utilised in the aforementioned experiments to determine the order of decarboxylation. Whilst all mutations tested appear to trap sirohaem or haem derivatives, mass spectrometry data would be essential for the characterisation of the bound compounds. Again, this may represent a convenient method for the determination of the order of decarboxylation.

6.4 Regulation of overlapping biosynthesis pathways

Many organisms have the ability to produce multiple tetrapyrroles which all stem from the common precursor uro'gen III and further branch at intermediates such as precorrin-2 and sirohydrochlorin (Fig. 6.1). Therefore, tight regulation of each tetrapyrrole biosynthesis pathway is required to allow flux along the correct pathway according to the organisms needs. Pathway regulation is particularly relevant for the sulphate reducing bacteria and

methanogenic archaea which require cobalamin, sirohaem, haem and coenzyme F_{430} (methanogens only). These organisms require regulation at each branch point including precorrin-2, sirohydrochlorin and sirohaem. Furthermore, it has been shown that *D. desulfuricans* uses iron coproporphyrin III as a unique cofactor in bacterioferritin (Romao *et al.*, 2000), again requiring a further point of regulation in the alternative haem pathway in this organism. Whilst regulation of anaerobic cobalamin and sirohaem syntheses likely occurs via modulation of chelatase activity (Stroupe *et al.*, 2003), further regulation is required for the conservation of sirohaem within the cell, modulating its use as both an enzyme cofactor and a pathway intermediate.

The redox controlled activity observed with the haem bound *M. barkeri* AhbA/B complex might represent a regulatory mechanism for sirohaem conservation utilised by the Archaea. Purification of other archaeal AhbA/B proteins will help determine if this is a conserved mechanism within the Archaea. *D. vulgaris* AhbA/B was found to bind haem at a much lower occupancy, and aerobically purified protein demonstrated full activity *in vitro* suggesting the haem does not affect activity of this complex. Similar haem-redox regulation would also appear to be absent from the *D. desulfuricans* complex as it did not copurify with haem. Therefore it is suspected that the sulphate reducers have an alternative method of regulation in comparison to the Archaea.

From the product bound AhbA/B structure it is clear that only a single tetrapyrrole could fit in the active site cleft at one time, therefore the haem bound to the *D. vulgaris* and *M. barkeri* complexes must be located at an alternative site. Given that *D. vulgaris* and *M. barkeri* do not contain any conserved histidines that are not also present in *D. desulfuricans* and both haem binding proteins display slightly different characteristics (*e.g.* haemochrome peak differences and tight binding by *D. vulgaris*) it is possible they have different haem binding sites.

The chimeric complexes produced from the proteins of the sulphate reducing bacteria reveal that both the haem binding and tetramerisation capacity of the *D. vulgaris* complex can be attributed to the AhbA subunit, although it is unknown if these are in any way linked. The N-terminal region of *D. vulgaris* AhbA contains 4 histidine residues that are not conserved in the other proteins, which may be a candidate for haem binding. *M. barkeri* AhbA contains a histidine (His171) that is not present in the other AhbA proteins tested. The equivalent residue is from *D. desulfuricans* is Lys159A which is located on sheet β6, the final strand that interacts with the RAM-like domain of the other subunit. Lys159A is exposed to the solvent within range of α4 which may provide a surface suitable for further interactions with

the haem molecule. However, His171 is not conserved among the archaeal proteins. Therefore, if haem binding is an archaeal trait His171 is unlikely to be the coordinating residue.

Site directed mutagenesis or crystallisation of a haem bound AhbA/B complex could be employed to further study these potential binding sites. Similarly, a comparative analysis of the binding/dissociation constants of the two haem binding proteins would be useful to further understand the association of these complexes and the prosthetic group. Whilst *D. vulgaris* has a lower occupancy it appears to have a great binding affinity as the haem is extremely hard to remove. However, these experiments would be difficult to perform as the proteins advantageously bind haem within the cell and therefore purify in the haem bound state. This would affect any binding measurements as there would already be a level of haem occupancy. Unfolding the protein and then removal of the haem by HPLC separation or dialysis followed by reconstitution of the complex may be a means to obtain apo-protein. Haem-free *M. barkeri* AhbA/B would also be extremely interesting for analysis of *in vitro* activity of the apo-protein complex.

Another appealing theory for the regulation of the alternative haem biosynthesis pathway stems from the crystal structure of *D. desulfuricans* AhbA/B. Solution of the structure revealed a striking resemblance to the Lrp/AsnC family of transcriptional regulators. While sequence similarity predicted the HtH motif to be present, the conservation of the lower RAM-like domain was unexpected. The similarity of these proteins furthers the question as to whether AhbA/B carries out both a catalytic and a regulatory role at the level of DNA. AhbB is more structurally similar to *E. coli* AsnC (PDB 2CG4; overall RMSD of 2.6 Å for 138 Cα atoms) than AhbA (overall RMSD of 3.8 Å for 123 Cα atoms). However, the lower structural similarity is clearly caused by the angling of the HtH domain of AhbA towards that of AhbB (Fig. 4.10), causing the opening of the side of the dimer and allowing access to the cleft formed between the RAM-like domains (Fig. 6.2C). Movement of the AhbA HtH domain into a position similar to AhbB would allow both domains to be in the correct orientation for DNA binding, but would likely block or reduce access to the active site cleft. If the AhbA/B complex is also a DNA binding protein, this domain movement would represent an extremely elegant method of both enzyme and DNA regulation. One can envisage that when haem synthesis is required AhbA/B is in the catalytic form, unable to bind DNA but allowing access to the active site cleft. In the event of haem synthesis no longer being required, an HtH domain movement would allow AhbA/B to switch to the catalytically inactive DNA binding form. The domain movement may occur after the binding of an effector molecule. Binding of amino acids (including leucine, asparagine, alanine,

valine, lysine and proline) by members of the Lrp/AsnC family regulates multimerisation and DNA binding properties of these proteins (Brinkman *et al.*, 2003). Therefore, these amino acids represent prime candidates as effector molecules for the AhbA/B complex. An obvious target for repression would be the *ahbC* and/or *ahbD* genes. This hypothesis is strengthened by the fact that the AhbA/B homologue NirL from *H. fasciatum* has been shown to bind DNA, specifically the putative promoter of *nirJ2*, a gene homologous to *ahbC* and *ahbD* (Xiong *et al.*, 2007). Experimental evidence is required to confirm the DNA binding properties of the AhbA/B complex.

The structural similarities of the AhbA/B complex and the Lrp/AsnC family raises many questions about the evolution of this decarboxylase. If the AhbA/B complex does bind DNA, was AhbA/B a transcriptional regulator which gained the catalytic function, or a fusion of two proteins to form a bifunctional enzyme? Alternatively, if the complex does not bind DNA, does this represent a loss of function and a repurposing of the protein? Genome-wide searches may aid the understanding of the origin of these proteins. However, misannotations of many of the AhbA/B proteins as Lrp/AsnC proteins hinders this process. Confirmation of the function of the HXYXR motif in AhbA/B proteins may provide a means for discovering true AhbA/B proteins and allow annotation of many of these family members.

Unlike the complexes from *D. desulfuricans* and *M. barkeri* the *D. vulgaris* complex forms both dimeric and tetrameric structures. The tetrameric complex may have a functional application with regards to regulation of activity or haem/DNA binding. Dimeric and tetrameric fractions were not separated prior to *in vitro* assays, therefore it is unknown if one form is active and the other is not. It is also unknown if these complexes are dynamic, reaching an equilibrium between the two forms in solution, or if an effector molecule is required for their formation. Members of the Lrp/AsnC family have been shown to form a variety of oligomers, some of which are regulated by amino acid binding (Brinkman *et al.*, 2003). Lrp/AsnC proteins form homodimeric subunits which are then arranged into higher order structures by dimer-dimer associations between their C-terminal RAM domains (Ren *et al.*, 2007). Therefore it is likely that the tetrameric form of *D. vulgaris* AhbA/B is produced by similar dimer-dimer interactions. Site directed mutagenesis of potential tetramerisation regions or cross-linking experiments may be useful methods for analysis of the dimeric and tetrameric species.

6.5 Conclusions

The alternative haem biosynthesis pathway likely represents a primitive route for haem synthesis. Although the metabolic pathway has been elucidated much remains unknown about the enzymes involved. The work presented within this thesis has provided major insights into the enzymes that catalyse the decarboxylation of sirohaem, the initial step of the alternative haem biosynthesis pathway. Sirohaem decarboxylase was revealed to be a heteromeric complex formed by AhbA and AhbB. Although only three AhbA/B protein complexes were examined, a surprisingly high degree of cross-species variation was observed raising many questions regarding enzyme regulation, haem/product binding and oligomerisation state of AhbA/B proteins. The structure of the AhbA/B complex from *D. desulfuricans* presented in this work represents the first enzyme from the alternative haem biosynthesis pathway to have its structure solved, and the first structure of a sirohaem decarboxylase. The structure of the AhbA/B complex revealed that whilst AhbA/B and its counterpart HemE have distinctly different sequences and structures, they have both evolved a similar mechanism for catalysis of decarboxylation reactions. Furthermore, a striking similarity with the Lrp/AsnC family of transcriptional regulators was revealed, leading to the question as to whether these proteins have both a catalytic and regulatory function. Whilst many properties of sirohaem decarboxylases have been discovered during this work, the cross-species variations have given rise to many unanswered questions, demonstrating there is much more work to be done to fully understand this family of enzymes.

References

Abicht, H.K., J. Martinez, G. Layer, D. Jahn & M. Solioz, (2012) *Lactococcus lactis* HemW (HemN) is a haem-binding protein with a putative role in haem trafficking. *Biochem J* **442**: 335-343.

Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger & P.H. Zwart, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213-221.

Akutsu, H., J.S. Park & S. Sano, (1993) L-Methionine Methyl Is Specifically Incorporated into the C-2 and C-7 Positions of the Porphyrin of Cytochrome-*c*3 in a Strictly Anaerobic Bacterium, *Desulfovibrio vulgaris*. *J Am Chem Soc* **115**: 12185-12186.

Al-Karadaghi, S., M. Hansson, S. Nikonov, B. Jonsson & L. Hederstedt, (1997) Crystal structure of ferrochelatase: the terminal enzyme in heme biosynthesis. *Structure* **5**: 1501-1510.

Alexander, F.W., E. Sandmeier, P.K. Mehta & P. Christen, (1994) Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families. *Eur J Biochem* **219**: 953-960.

Astner, I., J.O. Schulze, J. van den Heuvel, D. Jahn, W.D. Schubert & D.W. Heinz, (2005) Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *Embo J* **24**: 3166-3177.

Atta, M., E. Mulliez, S. Arragain, F. Forouhar, J.F. Hunt & M. Fontecave, (2010) S-Adenosylmethionine-dependent radical-based modification of biological macromolecules. *Curr Opin Struc Biol* **20**: 684-692.

Bali, S., A.D. Lawrence, S.A. Lobo, L.M. Saraiva, B.T. Golding, D.J. Palmer, M.J. Howard, S.J. Ferguson & M.J. Warren, (2011) Molecular hijacking of siroheme for the synthesis of heme and d_1 heme. *Proc Natl Acad Sci U S A* **108**: 18260-18265.

Bali, S., D.J. Palmer, S. Schroeder, S.J. Ferguson & M.J. Warren, (2014) Recent advances in the biosynthesis of modified tetrapyrroles: the discovery of an alternative pathway for the formation of heme and heme d_1. *Cell Mol Life Sci* **71**: 2837-2863.

Banerjee, R. & S.W. Ragsdale, (2003) The many faces of vitamin B_{12}: Catalysis by cobalamin-dependent enzymes. *Annu Rev Biochem* **72**: 209-247.

Barker, P.D. & S.J. Ferguson, (1999) Still a puzzle: why is haem covalently attached in c-type cytochromes? *Structure* **7**: R281-290.

Battersby, A.R., C.J.R. Fookes, G.W.J. Matcham & E. McDonald, (1979) Order of assembly of the four pyrrole rings during biosynthesis of the natural porphyrins. *J Chem Soc, Chem Commun* **1**: 539-541.

Battersby, A.R., (2000) Tetrapyrroles: the pigments of life. *Nat Prod Rep* **17**: 507-526.

Beale, S.I. & P.A. Castelfranco, (1973) ^{14}C incorporation from exogenous compounds into -aminolevulinic acid by greening cucumber cotyledons. *Biochem Biophys Res Commun* **52**: 143-149.

Beale, S.I., S.P. Gough & S. Granick, (1975) Biosynthesis of delta-aminolevulinic acid from the intact carbon skeleton of glutamic acid in greening barley. *Proc Natl Acad Sci U S A* **72**: 2719-2723.

Benner, S.A., A.D. Ellington & A. Tauer, (1989) Modern metabolism as a palimpsest of the RNA world *Proc Natl Acad Sci U S A* **86**: 7054-7058.

Berry, E.A. & B.L. Trumpower, (1987) Simultaneous determination of hemes *a*, *b*, and *c* from pyridine hemochrome spectra. *Anal Biochem* **161**: 1-15.

Bishop, D.F., A.S. Henderson & K.H. Astrin, (1990) Human delta-aminolevulinate synthase: assignment of the housekeeping gene to 3p21 and the erythroid-specific gene to the X chromosome. *Genomics* **7**: 207-214.

Blanche, F., D. Thibaut, L. Debussche, R. Hertle, F. Zipfel & G. Müller, (1993) Parallels and Decisive Differences in Vitamin B12 Biosyntheses. *Angewandte Chemie International Edition in English* **32**: 1651-1653.

Blankenship, R.E. & H. Hartman, (1998) The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci* **23**: 94-97.

Bollivar, D.W., C. Clauson, R. Lighthall, S. Forbes, B. Kokona, R. Fairman, L. Kundrat & E.K. Jaffe, (2004) *Rhodobacter capsulatus* porphobilinogen synthase, a high activity metal ion independent hexamer. *BMC Biochem* **5**: 17.

Bollivar, D.W., (2006) Recent advances in chlorophyll biosynthesis. *Photosynth Res* **90**: 173-194.

Booker, S.J., (2009) Anaerobic functionalization of unactivated C-H bonds. *Curr Opin Chem Biol* **13**: 58-73.

Boynton, T.O., L.E. Daugherty, T.A. Dailey & H.A. Dailey, (2009) Identification of *Escherichia coli* HemG as a Novel, Menadione-Dependent Flavodoxin with Protoporphyrinogen Oxidase Activity. *Biochemistry* **48**: 6705-6711.

Boynton, T.O., S. Gerdes, S.H. Craven, E.L. Neidle, J.D. Phillips & H.A. Dailey, (2011) Discovery of a gene involved in a third bacterial protoporphyrinogen oxidase activity through comparative genomic analysis and functional complementation. *Appl Environ Microbiol* **77**: 4795-4801.

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

Breckau, D., E. Mahlitz, A. Sauerwald, G. Layer & D. Jahn, (2003) Oxygen-dependent coproporphyrinogen III oxidase (HemF) from *Escherichia coli* is stimulated by manganese. *J Biol Chem* **278**: 46625-46631.

Brindley, A.A., E. Raux, H.K. Leech, H.L. Schubert & M.J. Warren, (2003) A story of chelatase evolution: identification and characterization of a small 13-15-kDa "ancestral" cobaltochelatase ($CbiX^S$) in the archaea. *J Biol Chem* **278**: 22388-22395.

Brinkman, A.B., I. Dahlke, J.E. Tuininga, T. Lammers, V. Dumay, E. de Heus, J.H. Lebbink, M. Thomm, W.M. de Vos & J. van Der Oost, (2000) An Lrp-like transcriptional

regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J Biol Chem* **275**: 38160-38169.

Brinkman, A.B., S.D. Bell, R.J. Lebbink, W.M. de Vos & J. van der Oost, (2002) The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277**: 29537-29549.

Brinkman, A.B., T.J. Ettema, W.M. de Vos & J. van der Oost, (2003) The Lrp family of transcriptional regulators. *Mol Microbiol* **48**: 287-294.

Buchenau, B., J. Kahnt, I.U. Heinemann, D. Jahn & R.K. Thauer, (2006) Heme biosynthesis in *Methanosarcina barkeri* via a pathway involving two methylation reactions. *J Biol Chem* **188**: 8666-8668.

Cabrera, G., R. Perez, J.M. Gomez, A. Abalos & D. Cantero, (2006) Toxic effects of dissolved heavy metals on *Desulfovibrio vulgaris* and *Desulfovibrio* sp. strains. *J Hazard Mater* **135**: 40-46.

Calvo, J.M. & R.G. Matthews, (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol Rev* **58**: 466-490.

Cavallaro, G., L. Decaria & A. Rosato, (2008) Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems. *J Proteome Res* **7**: 4946-4954.

Chen, M.W., D. Jahn, G.P. O'Neill & D. Soll, (1990) Purification of the glutamyl-tRNA reductase from *Chlamydomonas reinhardtii* involved in delta-aminolevulinic acid formation during chlorophyll biosynthesis. *J Biol Chem* **265**: 4058-4063.

Chen, S., Z. Hao, E. Bieniek & J.M. Calvo, (2001) Modulation of Lrp action in *Escherichia coli* by leucine: effects on non-specific binding of Lrp to DNA. *J Mol Biol* **314**: 1067-1075.

Chipman, D.M. & B. Shaanan, (2001) The ACT domain family. *Curr Opin Struct Biol* **11**: 694-700.

Corradi, H.R., A.V. Corrigall, E. Boix, C.G. Mohan, E.D. Sturrock, P.N. Meissner & K.R. Acharya, (2006) Crystal structure of protoporphyrinogen oxidase from *Myxococcus xanthus* and its complex with the inhibitor acifluorfen. *J Biol Chem* **281**: 38625-38633.

Corrigall, A.V., K.B. Siziba, M.H. Maneli, E.G. Shephard, M. Ziman, T.A. Dailey, H.A. Dailey, R.E. Kirsch & P.N. Meissner, (1998) Purification of and kinetic studies on a cloned protoporphyrinogen oxidase from the aerobic bacterium *Bacillus subtilis*. *Arch Biochem Biophys* **358**: 251-256.

Cowtan, K., (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**: 1002-1011.

Crane, B.R. & E.D. Getzoff, (1996) The relationship between structure and function for the sulfite reductases. *Curr Opin Struc Biol* **6**: 744-756.

Dailey, H.A., C.S. Jones & S.W. Karr, (1989) Interaction of free porphyrins and metalloporphyrins with mouse ferrochelatase. A model for the active site of ferrochelatase. *Biochim Biophys Acta* **999**: 7-11.

Dailey, H.A. & T.A. Dailey, (1996) Protoporphyrinogen oxidase of *Myxococcus xanthus*. Expression, purification, and characterization of the cloned enzyme. *J Biol Chem* **271**: 8714-8718.

Dailey, H.A., T.A. Dailey, C.K. Wu, A.E. Medlock, K.F. Wang, J.P. Rose & B.C. Wang, (2000) Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters. *Cell Mol Life Sci* **57**: 1909-1926.

Dailey, H.A., (2002) Terminal steps of haem biosynthesis. In: Colloquium on Tetrapyrroles - Their Life, Birth and Death. Heriot-Watt, Edinburgh, Scotland: Biochem Soc Trans, pp. 590-595.

Dailey, T.A. & H.A. Dailey, (1997) Expression, purification, and characteristics of mammalian protoporphyrinogen oxidase. In: Methods Enzymol. D. McCormick, J. Suttie & C. Wagner (eds). pp. 340-349.

Dailey, T.A., T.O. Boynton, A.N. Albetel, S. Gerdes, M.K. Johnson & H.A. Dailey, (2010) Discovery and Characterization of HemQ: an essential heme biosynthetic pathway component. *J Biol Chem* **285**: 25978-25986.

De Vrieze, J., T. Hennebel, N. Boon & W. Verstraete, (2012) *Methanosarcina*: the rediscovered methanogen for heavy duty biomethanation. *Bioresour Technol* **112**: 1-9.

Deery, E., S. Schroeder, A.D. Lawrence, S.L. Taylor, A. Seyedarabi, J. Waterman, K.S. Wilson, D. Brown, M.A. Geeves, M.J. Howard, R.W. Pickersgill & M.J. Warren, (2012) An enzyme-trap approach allows isolation of intermediates in cobalamin biosynthesis. *Nat Chem Biol* **8**: 933-940.

Dickman, S.R., (1977) Ribonucleotide reduction and possible role of cobalamin in evolution. *J Mol Evol* **10**: 251-260.

Dutton, P.L., (1978) Redox potentiometry: determination of midpoint potentials of oxidation-reduction components of biological electron-transfer systems. *Methods Enzymol* **54**: 411-435.

Elder, G.H., J.O. Evans, J.R. Jackson & A.H. Jackson, (1978) Factors Determining Sequence of Oxidative Decarboxylation of 2-Propionate and 4-Propionate Substituents of Coproporphyrinogen-III by Coproporphyrinogen Oxidase in Rat-Liver. *Biochem J* **169**: 215-223.

Emsley, P., B. Lohkamp, W.G. Scott & K. Cowtan, (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.

Ettema, T.J., A.B. Brinkman, T.H. Tani, J.B. Rafferty & J. Van Der Oost, (2002) A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *J Biol Chem* **277**: 37464-37468.

Ferreira, G.C. & J. Gong, (1995) 5-Aminolevulinate Synthase and the First Step of Heme-Biosynthesis. *J Bioenerg Biomembr* **27**: 151-159.

Frankenberg, N., J. Moser & D. Jahn, (2003) Bacterial heme biosynthesis and its biotechnological application. *Appl Microbiol Biotechnol* **63**: 115-127.

Frere, F., W.D. Schubert, F. Stauffer, N. Frankenberg, R. Neier, D. Jahn & D.W. Heinz, (2002) Structure of porphobilinogen synthase from *Pseudomonas aeruginosa* in complex with 5-fluorolevulinic acid suggests a double Schiff base mechanism. *J Mol Biol* **320**: 237-247.

Frey, P.A., A.D. Hegeman & F.J. Ruzicka, (2008) The radical SAM superfamily. *Crit Rev Biochem Mol* **43**: 63-88.

Friedmann, H.C. & R.K. Thauer, (1992) Macrocyclic tetrapyrrole biosynthesis in bacteria. In: Encyclopedia of Microbiology, Vols. 1-4. Academic Press, pp. 4) 1-19.

Fritz, G., O. Einsle, M. Rudolf, A. Schiffer & P.M. Kroneck, (2005) Key bacterial multi-centered metal enzymes involved in nitrate and sulfate respiration. *J Mol Microbiol Biotechnol* **10**: 223-233.

Gasteiger, E., C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel & A. Bairoch, (2005) Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Springer, pp. 571-607.

Gibson, K.D., A. Neuberger & J.J. Scott, (1954) The enzymic conversion of delta-aminolaevulic acid to porphobilinogen. *Biochem J* **58**: xli-xlii.

Hart, G.J., A.D. Miller, F.J. Leeper & A.R. Battersby, (1987) Biosynthesis of the Natural Porphyrins - Proof That Hydroxymethylbilane Synthase (Porphobilinogen Deaminase) Uses a Novel Binding Group in Its Catalytic Action. *J Chem Soc Chem Comm*: 1762-1764.

Haufschildt, K., S. Schmelz, T.M. Kriegler, A. Neumann, J. Streif, H. Arai, D.W. Heinz & G. Layer, (2014) The crystal structure of siroheme decarboxylase in complex with iron-uroporphyrin III reveals two essential histidine residues. *J Mol Biol* **426**: 3272-3286.

Hederstedt, L., (2012) Heme *a* biosynthesis. *Biochim Biophys Acta* **1817**: 920-927.

Heinemann, I.U., N. Diekmann, A. Masoumi, M. Koch, A. Messerschmidt, M. Jahn & D. Jahn, (2007) Functional definition of the tobacco protoporphyrinogen IX oxidase substrate-binding site. *Biochem J* **402**: 575-580.

Hennig, M., B. Grimm, R. Contestabile, R.A. John & J.N. Jansonius, (1997) Crystal structure of glutamate-1-semialdehyde aminomutase: an alpha2-dimeric vitamin B6-dependent enzyme with asymmetry in structure and active site reactivity. *Proc Natl Acad Sci U S A* **94**: 4866-4871.

Hockin, S.L. & G.M. Gadd, (2007) Bioremediation of metals and metalloids by precipitation and cellular binding. In: Sulphate-reducing Bacteria: Environmental and Engineered Systems. L. Barton, Hamilton, WA (ed). New York: Cambridge University Press, pp. 405-434.

Holliday, G.L., J.M. Thornton, A. Marquet, A.G. Smith, F. Rebeille, R. Mendel, H.L. Schubert, A.D. Lawrence & M.J. Warren, (2007) Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep* **24**: 972-987.

Hoober, J.K., A. Kahn, D.E. Ash, S. Gough & C.G. Kannangara, (1988) Biosynthesis of delta-aminolevulinate in greening barley leaves. IX. Structure of the substrate, mode

of gabaculine inhibition, and the catalytic mechanism of glutamate 1-semialdehyde aminotransferase. *Carlsberg Res Commun* **53**: 11-25.

Iida, K., I. Mimura & M. Kajiwara, (2002) Evaluation of two biosynthetic pathways to delta-aminolevulinic acid in *Euglena gracilis*. *Eur J Biochem* **269**: 291-297.

Ilag, L.L. & D. Jahn, (1992) Activity and spectroscopic properties of the *Escherichia coli* glutamate 1-semialdehyde aminotransferase and the putative active site mutant K265R. *Biochemistry* **31**: 7143-7151.

Ishida, T., L. Yu, H. Akutsu, K. Ozawa, S. Kawanishi, A. Seto, T. Inubushi & S. Sano, (1998) A primitive pathway of porphyrin biosynthesis and enzymology in *Desulfovibrio vulgaris*. *Proc Natl Acad Sci U S A* **95**: 4853-4858.

Jacobs, N.J. & J.M. Jacobs, (1976) Nitrate, fumarate, and oxygen as electron acceptors for a late step in microbial heme synthesis. *Biochim Biophys Acta* **449**: 1-9.

Jaffe, E.K., (2000) The porphobilinogen synthase family of metalloenzymes. *Acta Crystallogr D Biol Crystallogr* **56 (Pt 2)**: 115-128.

Jaffe, E.K., J. Kervinen, J. Martins, F. Stauffer, R. Neier, A. Wlodawer & A. Zdanov, (2002) Species-specific inhibition of porphobilinogen synthase by 4-oxosebacic acid. *J Biol Chem* **277**: 19792-19799.

Jahn, D., E. Verkamp & D. Soll, (1992) Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends Biochem Sci* **17**: 215-218.

Jahn, D., and Heinz, D. W., (2009) Biosynthesis of 5- aminolevulinic acid. In: Tetrapyrroles: birth, life, and death. M.J. Warren, and Smith, A. G. (ed). Austin, Texas: Landes Bioscience, pp.

Jensen, R.A., (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**: 409-425.

Jeong, J.A., E.Y. Baek, S.W. Kim, J.S. Choi & J.I. Oh, (2013) Regulation of the ald gene encoding alanine dehydrogenase by AldR in *Mycobacterium smegmatis*. *J Bacteriol* **195**: 3610-3620.

Jordan, P.M. & J.S. Seehra, (1979) The biosynthesis of uroporphyrinogen III: order of assembly of the four porphobilinogen molecules in the formation of the tetrapyrrole ring. *FEBS Lett* **104**: 364-366.

Jordan, P.M. & P.N. Gibbs, (1985) Mechanism of action of 5-aminolaevulinate dehydratase from human erythrocytes. *Biochem J* **227**: 1015-1020.

Jordan, P.M. & M.J. Warren, (1987) Evidence for a dipyrromethane cofactor at the catalytic site of *E. coli* porphobilinogen deaminase. *FEBS Lett* **225**: 87-92.

Karlberg, T., D. Lecerof, M. Gora, G. Silvegren, R. Labbe-Bois, M. Hansson & S. Al-Karadaghi, (2002) Metal binding to *Saccharomyces cerevisiae* ferrochelatase. *Biochemistry* **41**: 13499-13506.

Kato, K., R. Tanaka, S. Sano, A. Tanaka & H. Hosaka, (2010) Identification of a gene essential for protoporphyrinogen IX oxidase activity in the cyanobacterium *Synechocystis sp*. PCC6803. *Proc Natl Acad Sci U S A* **107**: 16649-16654.

Kawasaki, S., H. Arai, T. Kodama & Y. Igarashi, (1997) Gene cluster for dissimilatory nitrite reductase (*nir*) from *Pseudomonas aeruginosa*: Sequencing and identification of a locus for heme d_1 biosynthesis. *J Bacteriol* **179**: 235-242.

Kelley, L.A. & M.J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371.

Kervinen, J., E.K. Jaffe, F. Stauffer, R. Neier, A. Wlodawer & A. Zdanov, (2001) Mechanistic basis for suicide inactivation of porphobilinogen synthase by 4,7-dioxosebacic acid, an inhibitor that shows dramatic species selectivity. *Biochemistry* **40**: 8227-8236.

Kikuchi, G., A. Kumar, P. Talmage & D. Shemin, (1958) The enzymatic synthesis of delta-aminolevulinic acid. *J Biol Chem* **233**: 1214-1219.

Koch, M., C. Breithaupt, R. Kiefersauer, J. Freigang, R. Huber & A. Messerschmidt, (2004) Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. *Embo J* **23**: 1720-1728.

Kolling, R. & H. Lother, (1985) AsnC: an autogenously regulated activator of asparagine synthetase A transcription in Escherichia coli. *J Bacteriol* **164**: 310-315.

Kranz, R.G., C. Richard-Fogal, J.S. Taylor & E.R. Frawley, (2009) Cytochrome *c* biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. *Microbiol Mol Biol Rev* **73**: 510-528, Table of Contents.

Kuhner, M., K. Haufschildt, A. Neumann, S. Storbeck, J. Streif & G. Layer, (2014) The alternative route to heme in the methanogenic archaeon *Methanosarcina barkeri*. *Archaea* **2014**: 327637.

Laemmli, U.K., (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680-685.

Lash, T.D., (2005) The enigma of coproporphyrinogen oxidase: how does this unusual enzyme carry out oxidative decarboxylations to afford vinyl groups? *Bioorg Med Chem Lett* **15**: 4506-4509.

Layer, G., J. Moser, D.W. Heinz, D. Jahn & W.D. Schubert, (2003) Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *Embo J* **22**: 6214-6224.

Layer, G., D.W. Heinz, D. Jahn & W.D. Schubert, (2004) Structure and function of radical SAM enzymes. *Curr Opin Chem Biol* **8**: 468-476.

Layer, G., K. Grage, T. Teschner, V. Schunemann, D. Breckau, A. Masoumi, M. Jahn, P. Heathcote, A.X. Trautwein & D. Jahn, (2005) Radical S-adenosylmethionine enzyme coproporphyrinogen III oxidase HemN: functional features of the [4Fe-4S] cluster and the two bound S-adenosyl-L-methionines. *J Biol Chem* **280**: 29038-29046.

Layer, G., A.J. Pierik, M. Trost, S.E. Rigby, H.K. Leech, K. Grage, D. Breckau, I. Astner, L. Jansch, P. Heathcote, M.J. Warren, D.W. Heinz & D. Jahn, (2006) The substrate radical of *Escherichia coli* oxygen-independent coproporphyrinogen III oxidase HemN. *J Biol Chem* **281**: 15727-15734.

Layer, G., J. Reichelt, D. Jahn & D.W. Heinz, (2010) Structure and function of enzymes in heme biosynthesis. *Protein Sci* **19**: 1137-1161.

Lee, D.S., E. Flachsova, M. Bodnarova, B. Demeler, P. Martasek & C.S. Raman, (2005) Structural basis of hereditary coproporphyria. *Proc Natl Acad Sci U S A* **102**: 14232-14237.

Lewis, C.A., Jr. & R. Wolfenden, (2008) Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes. *Proc Natl Acad Sci U S A* **105**: 17328-17333.

Lobo, S.A.L., A.A. Brindley, C.V. Romao, H.K. Leech, M.J. Warren & L.M. Saraiva, (2008) Two distinct roles for two functional cobaltochelatases (CbiK) in *Desulfovibrio vulgaris* Hildenborough. *Biochemistry* **47**: 5851-5857.

Lobo, S.A.L., A. Brindley, M.J. Warren & L.M. Saraiva, (2009) Functional characterization of the early steps of tetrapyrrole biosynthesis and modification in *Desulfovibrio vulgaris* Hildenborough. *Biochem J* **420**: 317-325.

Lobo, S.A.L., M.J. Warren & L.M. Saraiva, (2012) Chapter Seven - Sulfate-Reducing Bacteria Reveal a New Branch of Tetrapyrrole Metabolism. In: Adv Microb Physiol. K.P. Robert (ed). Academic Press, pp. 267-295.

Lobo, S.A.L., A.D. Lawrence, C.V. Romao, M.J. Warren, M. Teixeira & L.M. Saraiva, (2014) Characterisation of *Desulfovibrio vulgaris* haem *b* synthase, a radical SAM family member. *Biochim Biophys Acta* **1844**: 1238-1247.

Lockhart, P.J., A.W.D. Larkum, M.A. Steel, P.J. Waddell & D. Penny, (1996) Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. *Proc Natl Acad Sci U S A* **93**: 1930-1934.

Louie, G.V., P.D. Brownlie, R. Lambert, J.B. Cooper, T.L. Blundell, S.P. Wood, M.J. Warren, S.C. Woodcock & P.M. Jordan, (1992) Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature* **359**: 33-39.

Lovell, S.C., I.W. Davis, W.B. Arendall, 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson & D.C. Richardson, (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**: 437-450.

Lubben, M. & K. Morand, (1994) Novel prenylated hemes as cofactors of cytochrome oxidases. Archaea have modified hemes *a* and *o*. *J Biol Chem* **269**: 21473-21479.

Luer, C., S. Schauer, K. Mobius, J. Schulze, W.D. Schubert, D.W. Heinz, D. Jahn & J. Moser, (2005) Complex formation between glutamyl-tRNA reductase and glutamate-1-semialdehyde 2,1-aminomutase in *Escherichia coli* during the initial reactions of porphyrin biosynthesis. *J Biol Chem* **280**: 18568-18572.

Luo, J. & C.K. Lim, (1993) Order of uroporphyrinogen III decarboxylation on incubation of porphobilinogen and uroporphyrinogen III with erythrocyte uroporphyrinogen decarboxylase. *Biochem J* **289 (Pt 2)**: 529-532.

Macieira, S., B.M. Martins & R. Huber, (2003) Oxygen-dependent coproporphyrinogen-III oxidase from *Escherichia coli*: one-step purification and biochemical characterisation. *FEMS Microbiol Lett* **226**: 31-37.

Madhusudhan, K.T., N. Huang & J.R. Sokatch, (1995) Characterization of BkdR-DNA binding in the expression of the bkd operon of *Pseudomonas putida*. *J Bacteriol* **177**: 636-641.

Madhusudhan, K.T., N. Huang, E.H. Braswell & J.R. Sokatch, (1997) Binding of L-branched-chain amino acids causes a conformational change in BkdR. *J Bacteriol* **179**: 276-279.

Maeder, D.L., I. Anderson, T.S. Brettin, D.C. Bruce, P. Gilna, C.S. Han, A. Lapidus, W.W. Metcalf, E. Saunders, R. Tapia & K.R. Sowers, (2006) The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J Bacteriol* **188**: 7922-7931.

Matsunaga, I. & Y. Shiro, (2004) Peroxide-utilizing biocatalysts: structural and functional diversity of heme-containing enzymes. *Curr Opin Chem Biol* **8**: 127-132.

McCoy, A.J., R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni & R.J. Read, (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674.

McGoldrick, H.M., C.A. Roessner, E. Raux, A.D. Lawrence, K.J. McLean, A.W. Munro, S. Santabarbara, S.E. Rigby, P. Heathcote, A.I. Scott & M.J. Warren, (2005) Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. *J Biol Chem* **280**: 1086-1094.

McNicholas, S., E. Potterton, K.S. Wilson & M.E. Noble, (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**: 386-394.

McPherson, A., Jr., (1976) Crystallization of proteins from polyethylene glycol. *J Biol Chem* **251**: 6300-6303.

Medlock, A.E. & H.A. Dailey, (1996) Human coproporphyrinogen oxidase is not a metalloprotein. *J Biol Chem* **271**: 32507-32510.

Mendel, R.R., A.G. Smith, A. Marquet & M.J. Warren, (2007) Metal and cofactor insertion. *Nat Prod Rep* **24**: 963-971.

Miles, E.W., S. Rhee & D.R. Davies, (1999) The molecular basis of substrate channeling. *J Biol Chem* **274**: 12193-12196.

Mobius, K., R. Arias-Cartin, D. Breckau, A.L. Hannig, K. Riedmann, R. Biedendieck, S. Schroder, D. Becher, A. Magalon, J. Moser, M. Jahn & D. Jahn, (2010) Heme biosynthesis is coupled to electron transport chains for energy generation. *Proc Natl Acad Sci U S A* **107**: 10436-10441.

Moser, J., S. Lorenz, C. Hubschwerlen, A. Rompf & D. Jahn, (1999) *Methanopyrus kandleri* glutamyl-tRNA reductase. *J Biol Chem* **274**: 30679-30685.

Muir, H.M. & A. Neuberger, (1950) The biogenesis of porphyrins. 2. The origin of the methyne carbon atoms. *Biochem J* **47**: 97-104.

Munro, A.W., M.A. Noble, L. Robledo, S.N. Daff & S.K. Chapman, (2001) Determination of the Redox Properties of Human NADPH-Cytochrome P450 Reductase. *Biochemistry* **40**: 1956-1963.

Munro, A.W., H.M. Girvan & K.J. McLean, (2007) Variations on a (t)heme--novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat Prod Rep* **24**: 585-609.

Murshudov, G.N., A.A. Vagin & E.J. Dodson, (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **53**: 240-255.

Muyzer, G. & A.J. Stams, (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nat Rev Microbiol* **6**: 441-454.

Nandi, D.L., (1978) Delta-aminolevulinic acid synthase of *Rhodopseudomonas spheroides*. Binding of pyridoxal phosphate to the enzyme. *Arch Biochem Biophys* **188**: 266-271.

Nicke, T., T. Schnitzer, K. Munch, J. Adamczack, K. Haufschildt, S. Buchmeier, M. Kucklick, U. Felgentrager, L. Jansch, K. Riedel & G. Layer, (2013) Maturation of the cytochrome cd_1 nitrite reductase NirS from *Pseudomonas aeruginosa* requires transient interactions between the three proteins NirS, NirN and NirF. *Biosci Rep* **33**.

Nogaj, L.A. & S.I. Beale, (2005) Physical and kinetic interactions between glutamyl-tRNA reductase and glutamate-1-semialdehyde aminotransferase of *Chlamydomonas reinhardtii*. *J Biol Chem* **280**: 24301-24307.

O'Brian, M.R. & L. Thony-Meyer, (2002) Biochemistry, regulation and genomics of haem biosynthesis in prokaryotes. *Adv Microb Physiol* **46**: 257-318.

Ost, T.W., C.S. Miles, A.W. Munro, J. Murdoch, G.A. Reid & S.K. Chapman, (2001) Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3. *Biochemistry* **40**: 13421-13429.

Palmedo, G., P. Seither, H. Korner, J.C. Matthews, R.S. Burkhalter, R. Timkovich & W.G. Zumft, (1995) Resolution of the *nirD* locus for heme d_1 synthesis of cytochrome cd_1 (respiratory nitrite reductase) from *Pseudomonas stutzeri*. *Eur J Biochem* **232**: 737-746.

Panek, H. & M.R. O'Brian, (2002) A whole genome view of prokaryotic haem biosynthesis. *Microbiology* **148**: 2273-2282.

Phillips, J.D., F.G. Whitby, J.P. Kushner & C.P. Hill, (2003) Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase. *Embo J* **22**: 6225-6233.

Phillips, J.D., F.G. Whitby, C.A. Warby, P. Labbe, C. Yang, J.W. Pflugrath, J.D. Ferrara, H. Robinson, J.P. Kushner & C.P. Hill, (2004) Crystal structure of the oxygen-dependant coproporphyrinogen oxidase (Hem13p) of *Saccharomyces cerevisiae*. *J Biol Chem* **279**: 38960-38968.

Phillips, J.D., C.A. Warby, F.G. Whitby, J.P. Kushner & C.P. Hill, (2009) Substrate Shuttling between Active Sites of Uroporphyrinogen Decarboxylase Is Not Required to Generate Coproporphyrinogen. *J Mol Biol* **389**: 306-314.

Puustinen, A. & M. Wikstrom, (1991) The heme groups of cytochrome *o* from *Escherichia coli*. *Proc Natl Acad Sci U S A* **88**: 6122-6126.

Radin, N.S., D. Rittenberg & D. Shemin, (1950) The role of acetic acid in the biosynthesis of heme. *J Biol Chem* **184**: 755-767.

Randau, L., S. Schauer, A. Ambrogelly, J.C. Salazar, J. Moser, S. Sekine, S. Yokoyama, D. Soll & D. Jahn, (2004) tRNA recognition by glutamyl-tRNA reductase. *J Biol Chem* **279**: 34931-34937.

Raux, E., T. McVeigh, S.E. Peters, T. Leustek & M.J. Warren, (1999a) The role of *Saccharomyces cerevisiae* Met1p and Met8p in sirohaem and cobalamin biosynthesis. *Biochem J* **338**: 701-708.

Raux, E., H.L. Schubert, J.M. Roper, K.S. Wilson & M.J. Warren, (1999b) Vitamin B12: Insights into Biosynthesis's Mount Improbable. *Bioorganic Chemistry* **27**: 100-118.

Raux, E., H.K. Leech, R. Beck, H.L. Schubert, P.J. Santander, C.A. Roessner, A.I. Scott, J.H. Martens, D. Jahn, C. Thermes, A. Rambach & M.J. Warren, (2003) Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium. Biochem J.* **370**: 505-516.

Ren, J., S. Sainsbury, S.E. Combs, R.G. Capper, P.W. Jordan, N.S. Berrow, D.K. Stammers, N.J. Saunders & R.J. Owens, (2007) The structure and transcriptional analysis of a global regulator from Neisseria meningitidis. *J Biol Chem* **282**: 14655-14664.

Richardson, D.J. & N.J. Watmough, (1999) Inorganic nitrogen metabolism in bacteria. *Curr Opin Chem Biol* **3**: 207-219.

Richardson, D.J., (2000) Bacterial respiration: a flexible process for a changing environment. *Microbiology* **146 (Pt 3)**: 551-571.

Risatti, J.B., W.C. Capman & D.A. Stahl, (1994) Community structure of a microbial mat: the phylogenetic dimension. *Proc Natl Acad Sci U S A* **91**: 10173-10177.

Rodgers, K.R., (1999) Heme-based sensors in biological systems. *Curr Opin Chem Biol* **3**: 158-167.

Romao, C.V., R. Louro, R. Timkovich, M. Lubben, M.Y. Liu, J. LeGall, A.V. Xavier & M. Teixeira, (2000) Iron-coproporphyrin III is a natural cofactor in bacterioferritin from the anaerobic bacterium *Desulfovibrio desulfuricans*. *FEBS Lett* **480**: 213-216.

Romao, C.V., D. Ladakis, S.A. Lobo, M.A. Carrondo, A.A. Brindley, E. Deery, P.M. Matias, R.W. Pickersgill, L.M. Saraiva & M.J. Warren, (2011) Evolution in a family of chelatases facilitated by the introduction of active site asymmetry and protein oligomerization. *Proc Natl Acad Sci U S A* **108**: 97-102.

Roth, J.R., J.G. Lawrence & T.A. Bobik, (1996) Cobalamin (coenzyme B12): synthesis and biological significance. *Annu Rev Microbiol* **50**: 137-181.

Rudiger, W., (1997) Chlorophyll metabolism: From outer space down to the molecular level. *Phytochemistry* **46**: 1151-1167.

Sambrook, J., E.F. Fritsch & T. Maniatis, (1989) *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press New York.

Sasarman, A., J. Letowski, G. Czaika, V. Ramirez, M.A. Nead, J.M. Jacobs & R. Morais, (1993) Nucleotide-Sequence of the *hemG* Gene Involved in the Protoporphyrinogen Oxidase Activity of *Escherichia coli* K12. *Can J Microbiol* **39**: 1155-1161.

Schafer, G., W. Purschke & C.L. Schmidt, (1996) On the origin of respiration: electron transport proteins from archaea to man. *FEMS Microbiol Rev* **18**: 173-188.

Schauer, S., S. Chaturvedi, L. Randau, J. Moser, M. Kitabatake, S. Lorenz, E. Verkamp, W.D. Schubert, T. Nakayashiki, M. Murai, K. Wall, H.U. Thomann, D.W. Heinz, H. Inokuchi, D. Soll & D. Jahn, (2002) *Escherichia coli* glutamyl-tRNA reductase. Trapping the thioester intermediate. *J Biol Chem* **277**: 48657-48663.

Schulze, J.O., W.D. Schubert, J. Moser, D. Jahn & D.W. Heinz, (2006) Evolutionary relationship between initial enzymes of tetrapyrrole biosynthesis. *J Mol Biol* **358**: 1212-1220.

Scott, A.I., (1990) Mechanistic and Evolutionary Aspects of Vitamin-B_{12} Biosynthesis. *Acc Chem Res* **23**: 308-317.

Senior, N.M., K. Brocklehurst, J.B. Cooper, S.P. Wood, P. Erskine, P.M. Shoolingin-Jordan, P.G. Thomas & M.J. Warren, (1996) Comparative studies on the 5-aminolaevulinic acid dehydratases from *Pisum sativum*, *Escherichia coli* and *Saccharomyces cerevisiae*. *Biochem J* **320 (Pt 2)**: 401-412.

Shemin, D. & D. Rittenberg, (1946) The biological utilization of glycine for the synthesis of the protoporphyrin of hemoglobin. *J Biol Chem* **166**: 621-625.

Shemin, D. & S. Kumin, (1952) The mechanism of porphyrin formation; the formation of a succinyl intermediate from succinate. *J Biol Chem* **198**: 827-837.

Shen, Y., R. Buick & D.E. Canfield, (2001) Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**: 77-81.

Shoolingin-Jordan, P.M., (1995) Porphobilinogen deaminase and uroporphyrinogen III synthase: structure, molecular biology, and mechanism. *J Bioenerg Biomembr* **27**: 181-195.

Shoolingin-Jordan, P.M., P. Spencer, M. Sarwar, P.E. Erskine, K.M. Cheung, J.B. Cooper & E.B. Norton, (2002) 5-Aminolaevulinic acid dehydratase: metals, mutants and mechanism. In: Colloquium on Tetrapyrroles - Their Life, Birth and Death Heriot-Watt, Edinburgh, Scotland: Biochem Soc T, pp. 584-590.

Silva, P.J. & M.J. Ramos, (2008) A comparative density-functional study of the reaction mechanism of the O_2-dependent coproporphyrinogen III oxidase. *Bioorg Med Chem* **16**: 2726-2733.

Silva, P.J., C. Schulz, D. Jahn, M. Jahn & M.J. Ramos, (2010) A tale of two acids: when arginine is a more appropriate acid than H3O+. *J Phys Chem B* **114**: 8994-9001.

Smith, A.G. & J.E. Francis, (1979) Decarboxylation of porphyrinogens by rat liver uroporphyrinogen decarboxylase. *Biochem J* **183**: 455-458.

Smith, M.A., B. Grimm, C.G. Kannangara & D. von Wettstein, (1991) Spectral kinetics of glutamate-1-semialdehyde aminomutase of *Synechococcus*. *Proc Natl Acad Sci U S A* **88**: 9775-9779.

Sofia, H.J., G. Chen, B.G. Hetzler, J.F. Reyes-Spindola & N.E. Miller, (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **29**: 1097-1106.

Spivey, A.C., A. Capretta, C.S. Frampton, F.J. Leeper & A.R. Battersby, (1996) Biosynthesis of porphyrins and related macrocycles. Part 45. Determination by a novel X-ray method of the absolute configuration of the spiro lactam which inhibits uroporphyrinogen III synthase (cosynthetase). *J. Chem. Soc., Perkin Trans. 1*: 2091-2102.

Stephenson, J.R., J.A. Stacey, J.B. Morgenthaler, J.A. Friesen, T.D. Lash & M.A. Jones, (2007) Role of aspartate 400, arginine 262, and arginine 401 in the catalytic mechanism of human coproporphyrinogen oxidase. *Protein Sci* **16**: 401-410.

Storbeck, S., S. Rolfes, E. Raux-Deery, M.J. Warren, D. Jahn & G. Layer, (2010) A novel pathway for the biosynthesis of heme in Archaea: genome-based bioinformatic predictions and experimental evidence. *Archaea* **2010**: 175050.

Stroupe, M.E., H.K. Leech, D.S. Daniels, M.J. Warren & E.D. Getzoff, (2003) CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol* **10**: 1064-1073.

Sun, P.D., S. Radaev & M. Kattah, (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr D Biol Crystallogr* **58**: 1092-1098.

Susanti, D. & B. Mukhopadhyay, (2012) An Intertwined Evolutionary History of Methanogenic Archaea and Sulfate Reduction. *PLoS One* **7**.

Suzuki, M., T. Hirai, H. Arai, M. Ishii & Y. Igarashi, (2006) Purification, characterization, and gene cloning of thermophilic cytochrome cd_1 nitrite reductase from *Hydrogenobacter thermophilus* TK-6. *J Biosci and Bioeng* **101**: 391-397.

Thauer, R.K. & L.G. Bonacker, (1994) Biosynthesis of coenzyme F_{430}, a nickel porphinoid involved in methanogenesis. In: Biosynthesis of the Tetrapyrrole Pigments. D.J.A. Chadwick, K (ed). Chichester, England: John Wiley & Sons Ltd, pp. 210-222; discussion 222-217.

Thauer, R.K., (1998) Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture. *Microbiology* **144 (Pt 9)**: 2377-2406.

Thauer, R.K., E. Stackebrandt & W.A. Hamilton, (2007) Energy metabolism phylogenetic diversity of sulphate-reducing bacteria. In: Sulphate-Reducing Bacteria: Environmental and Engineered Systems. L.A.H. Barton, WA (ed). New York: Cambridge University Press, pp. 1-38.

Thauer, R.K., A.K. Kaster, H. Seedorf, W. Buckel & R. Hedderich, (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**: 579-591.

Thaw, P., S.E. Sedelnikova, T. Muranova, S. Wiese, S. Ayora, J.C. Alonso, A.B. Brinkman, J. Akerboom, J. van der Oost & J.B. Rafferty, (2006) Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res* **34**: 1439-1449.

Tice, M.M. & D.R. Lowe, (2006) The origin of carbonaceous matter in pre-3.0 Ga greenstone terrains: A review and new evidence from the 3.42 Ga Buck Reef Chert. *Earth-Science Reviews* **76**: 259-300.

Trumpower, B.L., (2002) A concerted, alternating sites mechanism of ubiquinol oxidation by the dimeric cytochrome bc(1) complex. *Biochim Biophys Acta* **1555**: 166-173.

Voordouw, G., (2008) Impact of nitrate on the sulfur cycle in oil fields. In: Microbial Sulfur Metabolism. C.a.F. Dahl, C.G. (ed). Springer, pp. 296-302.

Warren, M.J. & P.M. Jordan, (1988) Investigation into the nature of substrate binding to the dipyrromethane cofactor of *Escherichia coli* porphobilinogen deaminase. *Biochemistry* **27**: 9020-9030.

Warren, M.J. & A.I. Scott, (1990) Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors. *Trends Biochem Sci* **15**: 486-491.

Warren, M.J., E.L. Bolt, C.A. Roessner, A.I. Scott, J.B. Spencer & S.C. Woodcock, (1994) Gene dissection demonstrates that the *Escherichia coli cysG* gene encodes a multifunctional protein. *Biochem J* **302 (Pt 3)**: 837-844.

Warren, M.J., E. Raux, H.L. Schubert & J.C. Escalante-Semerena, (2002) The biosynthesis of adenosylcobalamin (vitamin B_{12}). *Nat Prod Rep* **19**: 390-412.

Warren, M.J., A.G. Smith, E. Deery & R.-S. Rose, (2009) Biosynthesis of Siroheme and Coenzyme F_{430}. In: Tetrapyrroles: Birth, Life and Death. Springer New York, pp. 344-351.

Weinstein, J.D. & S.I. Beale, (1983) Separate physiological roles and subcellular compartments for two tetrapyrrole biosynthetic pathways in *Euglena gracilis*. *J Biol Chem* **258**: 6799-6807.

Whitby, F.G., (1998) Crystal structure of human uroporphyrinogen decarboxylase. *Embo J* **17**: 2463-2471.

Whitby, F.G., J.D. Phillips, J.P. Kushner & C.P. Hill, (1998) Crystal structure of human uroporphyrinogen decarboxylase. *Embo J* **17**: 2463-2471.

Wilks, A. & K.A. Burkhard, (2007) Heme and virulence: how bacterial pathogens regulate, transport and utilize heme. *Nat Prod Rep* **24**: 511-522.

Willins, D.A., C.W. Ryan, J.V. Platko & J.M. Calvo, (1991) Characterization of Lrp, and *Escherichia coli* regulatory protein that mediates a global response to leucine. *J Biol Chem* **266**: 10768-10774.

Winn, M.D., C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin & K.S. Wilson, (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**: 235-242.

Woodcock, S.C. & P.M. Jordan, (1994) Evidence for participation of aspartate-84 as a catalytic group at the active site of porphobilinogen deaminase obtained by site-directed mutagenesis of the *hemC* gene from *Escherichia coli*. *Biochemistry* **33**: 2688-2695.

Wu, C.K., H.A. Dailey, J.P. Rose, A. Burden, V.M. Sellers & B.C. Wang, (2001) The 2.0 A structure of human ferrochelatase, the terminal enzyme of heme biosynthesis. *Nat Struct Biol* **8**: 156-160.

Xiong, J., C.E. Bauer & A. Pancholy, (2007) Insight into the haem d_1 biosynthesis pathway in heliobacteria through bioinformatics analysis. *Microbiology* **153**: 3548-3562.

Yapbondoc, F., L.L. Bondoc, R. Timkovich, D.C. Baker & A. Hebbler, (1990) C-methylation occurs during the biosynthesis of heme d_1. *J Biol Chem* **265**: 13498-13500.

Ycas, M., (1974) On earlier states of the biochemical system. *J Theor Biol* **44**: 145-160.

Zajicek, R.S., S. Bali, S. Arnold, A.A. Brindley, M.J. Warren & S.J. Ferguson, (2009) d_1 haem biogenesis - assessing the roles of three *nir* gene products. *Febs J* **276**: 6399-6411.

Appendices

Appendix I – Thermal Shift Assay Buffers

Screen 1

All buffers were pH adjusted using glacial acetic acid or sodium hydroxide prior to use. All conditions were used at a final concentration of 100 mM containing 100 mM NaCl.

Buffer	pH	Melting temperature (°C)
Tri-sodium citrate	4.2	25.2
Tri-sodium citrate	4.6	34.8
Tri-sodium citrate	5	50.0
Tri-sodium citrate	5.4	54.6
Tri-sodium citrate	5.8	50.8
Tri-sodium citrate	6.2	55.8
MES	5.7	47.4
MES	5.9	48.4
MES	6.4	53.0
MES	6.8	53.2
Sodium cacodylate	6.5	54.6
HEPES	6.8	52.2
HEPES	7.2	56
HEPES	7.6	57
HEPES	8.0	56.8
Tris	7.2	57.2
Tris	7.6	54.6
Tris	8.0	55.6
Tris	8.4	54.6
Tris	8.8	54.8
Tris	9.0	54.4
Glycine	9.4	54.8
Glycine	9.8	53.2

Screen 2

Screen 2 was performed using 100 mM HEPES, pH 8.0 (adjusted using glacial acetic acid), with the following additives:

Additive(s)	Concentration (mM)	Melting temperature (°C)
NaCl	10	58.2
NaCl	100	58.6
NaCl	250	58.8
NaCl	500	42.8
NaCl / β-mercaptoethanol	100 / 2.5	58
NaCl / β-mercaptoethanol	100 / 5	58.6
NaCl / dithiothreitol	100 / 2.5	57.6
NaCl / dithiothreitol	100 / 5	58.4
NaCl / glycerol	100 / 5% (v/v)	58
NaCl / glycerol	100 / 10% (v/v)	58.6
KCl	100	58

Appendix II – Crystallisation Screens

Structure screen 1™

Condition	Salt	Precipitant(s)	Buffer
1	0.02 M Calcium chloride dihydrate	30% (v/v) 2-methyl-2,4 pentanediol	0.1 M Na Acetate trihydrate pH 4.6
2	0.2 M Ammonium acetate	30% (w/v) PEG 4000	0.1 M Na Acetate trihydrate pH 4.6
3	0.2 M Ammonium sulphate	25% (w/v) PEG 4000	0.1 M Na acetate trihydrate pH 4.6
4		2.0 M Sodium formate	0.1 M Na acetate trihydrate pH 4.6
5		2.0 M Ammonium sulphate	0.1 M Na acetate trihydrate pH 4.6
6		8% (w/v) PEG 4000	0.1 M Na acetate trihydrate pH 4.6
7	0.2 M Ammonium acetate	30% (w/v) PEG 4000	0.1 M tri-sodium citrate dihydrate pH 5.6
8	0.2 M Ammonium acetate	30% (v/v) 2-methyl-2,4-pentanediol	0.1 M tri-sodium citrate dihydrate pH 5.6
9		20% (w/v) 2-propanol, 20% (w/v) PEG 4000	0.1 M tri-Sodium citrate dihydrate pH 5.6
10		1.0M Ammonium dihydrogen phosphate	0.1 M Na Citrate pH 5.6
11	0.2 M Calcium chloride dihydrate	20% (v/v) 2-propanol	0.1 M Na acetate trihydrate pH 4.6
12		1.4 M Na acetate trihydrate	0.1 M Na Cacodylate pH 6.5
13	0.2 M tri-sodium citrate dihydrate	30% (v/v) 2-propanol	0.1 M Na Cacodylate pH 6.5
14	0.2 M Ammoniium	30% (w/v) PEG 8000	0.1 M Na Cacodylate pH

	sulphate		6.5
15	0.2 M Magnesium acetate tetrahydrate	20% PEG 8000	0.1 M Na Cacodylate pH 6.5
16	0.2 M Magnesium acetate tetrahydrate	30% (v/v) 2-methyl-2,4-pentanediol	0.1 M Na Cacodylate pH 6.5
17		1.0 M Sodium acetate trihydrate	0.1 M Imidazole pH 6.5
18	0.2 M Sodium acetate trihydrate	30% (w/v) PEG 8000	0.1 M Na Cacodylate pH 6.5
19	0.2 M Zinc acetate dihydrate	18% (w/v) PEG 8000	0.1 M Na Cacodylate pH 6.5
20	0.2 M Calcium acetate hydrate	18% (w/v) PEG 8000	0.1 M Na Cacodylate pH 6.5
21	0.2 M tri-sodium citrate dihydrate	30% (v/v) 2-methyl-2,4-pentanediol	0.1 M Na HEPES pH 7.5
22	0.2 M Magnesium chloride hexahydrate	30% (v/v) 2-propanol	0.1 M Na HEPES pH 7.5
23	0.2 M Calcium chloride dihydrate	28% (v/v) PEG 400	0.1 M Na HEPES pH 7.5
24	0.2 M Magnesium chloride hexahydrate	30% (v/v) PEG 400	0.1 M Na HEPES pH 7.5
25	0.2 M tri-sodium citrate dihydrate	20% (v/v) 2-propanol	0.1 M Na HEPES pH 7.5
26		0.8 M K, Na tartrate tetrahydrate	0.1 M Na HEPES pH 7.5
27		1.5 M Lithium sulphate monohydrate	0.1 M Na HEPES pH 7.5
28		0.8 M Na dihydrogen phosphate	0.1 M Na HEPES pH 7.5
			0.8 M K dihydrogen phosphate monohydrate
29		1.4 M tri-Sodium citrate dihydrate	0.1 M Na HEPES pH 7.5
30		2% (v/v) PEG 400, 2.0M Ammonium sulphate	0.1 M Na HEPES pH 7.5
31		10% (v/v) 2-propanol, 20% (w/v) PEG 4000	0.1 M Na HEPES pH 7.5
32		2.0 M Ammonium sulphate	0.1 M Tris HCl pH 8.5
33	0.2 M Magnesium chloride hexahydrate	30% (w/v) PEG 4000	0.1 M Tris HCl pH 8.5
34	0.2 M tri-sodum citrate dihydrate	30% (v/v) PEG 400	0.1 M Tris HCl pH 8.5
35	0.2 M Lithium sulphate monohydrate	30% (w/v) PEG 4000	0.1 M Tris HCl pH 8.5
36	0.2 M Ammonium acetate	30% (v/v) 2-propanol	0.1 M Tris HCl pH 8.5
37	0.2 M Sodium acetate trihydrate	30% (w/v) PEG 4000	0.1 M Tris HCl pH 8.5
38		8% (w/v) PEG 8000	0.1 M Tris HCl pH 8.5
39		2.0 M Ammonium dihydrogen phosphate	0.1 M Tris HCl pH 8.5
40		0.4 M K, Na Tartrate tetrahydrate	

41		0.4 M Ammonium dihydrogen phosphate	
42	0.2 M Ammonium sulphate	30% (w/v) PEG 8000	
43	0.2 M Ammonium sulphate	30% (w/v) PEG 4000	
44		2.0 M Ammonium sulphate	
45		4.0M Sodium formate	
46	0.05 M Potassium dihydrogen phosphate	20% (w/v) PEG 8000	
47		30% (w/v) PEG 1500	
48		0.2M Magnesium formate	
49	1.0 M Lithium sulphate monohydrate	2% (w/v) PEG 8000	
50	0.5 M Lithium sulphate monohydrate	15% (w/v) PEG 8000	

Structure screen 2™

Condition	Salt/Additive	Precipitant(s)	Buffer
1	0.1 M Sodium chloride	30% (w/v) PEG monomethylether 550	0.1 M Bicine pH 9.0
2		2.0 M Magnesium chloride hexahydrate	0.1 M Bicine pH 9.0
3	2% (w/v) Dioxane	10% (w/v) PEG 20,000	0.1 M Bicine pH 9.0
4	0.2 M Magnesium chloride hexahydrate	3.4 M 1,6 Hexanediol	0.1 M Tris pH 8.5
5		25% (v/v) tert-butanol	0.1 M Tris pH 8.5
6	0.01 M Nickel chloride hexahydrate	1.0 M Lithium sulphate	0.1 M Tris pH 8.5
7	1.5 M Ammonium sulphate	12% (v/v) Glycerol	0.1 M Tris pH 8.5
8	0.2 M Ammonium phosphate monobasic	50% (v/v) MPD	0.1 M Tris pH 8.5
9		20% (v/v) Ethanol	0.1 M Tris pH 8.5
10	0.01 M Nickel chloride hexahydrate	20% (w/v) PEG monomethylether 2000	0.1 M Tris pH 8.5
11	0.5 M Ammonium sulphate	30% (v/v) MPD	0.1 M HEPES pH 7.5
12		10% (w/v) PEG 6000, 5% (v/v) MPD	0.1 M HEPES pH 7.5
13		20% (v/v) Jeffamine M-600	0.1 M HEPES pH 7.5
14	0.1 M Sodium chloride	1.6 M Ammonium sulphate	0.1 M HEPES pH 7.5
15		2.0 M Ammonium formate	0.1 M HEPES pH 7.5

16	0.05 M Cadmium sulphate octahydrate	1.0 M Sodium acetate	0.1 M HEPES pH 7.5
17		70% (v/v) MPD	0.1 M HEPES pH 7.5
18		4.3 M Sodium chloride	0.1 M HEPES pH 7.5
19		10% (w/v) PEG 8000, 8% (v/v) Ethylene glycol	0.1 M HEPES pH 7.5
20		1.6 M Magnesium sulphate heptahydrate	0.1 M MES pH 6.5
21	0.1 M Na phosphate monobasic	2.0 M Sodium Chloride	0.1 M MES pH 6.5
	0.1 M K phosphate monobasic		
22		12% (w/v) PEG 20,000	0.1 M MES pH 6.5
23	1.6 M Ammonium sulphate	10% (v/v) Dioxane	0.1 M MES pH 6.5
24	0.05 M Cesium chloride	30% (v/v) Jeffamine® M-600	0.1 M MES pH 6.5
25	0.01 M Cobalt chloride hexahydrate	1.8 M Ammonium sulphate	0.1 M MES pH 6.5
26	0.2 M Ammonium sulphate	30% (w/v) PEG monomethylether 5000	0.1 M MES pH 6.5
27	0.01 M Zinc sulphate heptahydrate	25% (v/v) PEG monomethylether 550	0.1 M MES pH 6.5
28		20% (w/v) PEG 10,000	0.1 M HEPES pH 7.5
29	0.2 M K/Na Tartrate	2.0 M Ammonium sulphate	0.1 M Sodium citrate pH 5.6
30	0.5 M Ammonium sulphate	1.0 M Lithium sulphate	0.1 M Sodium citrate pH 5.6
31	0.5 M Sodium chloride	4% (v/v) polyethyleneimine	0.1 M Sodium citrate pH 5.6
32		35% (v/v) tert-butanol	0.1 M Sodium citrate pH 5.6
33	0.01 M Ferric chloride hexahydrate	10% (v/v) Jeffamine® M-600	0.1 M Sodium citrate pH 5.6
34	0.01 M Manganese chloride tetrahydrate	2.5 M 1,6 Hexanediol	0.1 M Sodium citrate pH 5.6
35		2.0 M Sodium chloride	0.1 M Sodium acetate pH 4.6
36	0.2 M Sodium Chloride	30% (v/v) MPD	0.1 M Sodium acetate pH 4.6
37	0.01 M Cobalt Chloride hexahydrate	1.0M 1,6 Hexanediol	0.1 M Sodium acetate pH 4.6
38	0.1 M Cadmium chloride	30% (v/v) PEG 400	0.1 M Sodium acetate pH 4.6
39	0.2 M Ammonium sulphate	30% (w/v) PEG monomethylether 2000	0.1 M Sodium acetate pH 4.6
40	2.0 M Sodium chloride	10% (w/v) PEG 6000	
41	0.01 M Cetyl trimethyl ammoniumbromide	0.5 M Sodium chloride 0.1 M Magnesium chloride hexahydrate	
42		25% (v/v) Ethylene	

		glycol	
43		35% (v/v) Dioxane	
44	2.0 M Ammonium sulphate	5% (v/v) Isopropanol	
45		1.0 M Imidazole pH 7.0	
46		10% (w/v) PEG 1000, 10% (w/v) PEG 8000	
47	1.5 M Sodium chloride	10% (v/v) Ethanol	
48		1.6 M Sodium citrate pH 6.5	
49	15% (w/v) polyvinylpyrolidone		
50	2.0 M Urea		

Clear Strategy Screen I™

Clear Strategy Screen I™ is formulated such that a range of buffers (see below) can be added to each salt/precipitant solution to expand the range of conditions screened. Typically, 0.1 mL of buffer solution is added to 0.9 mL salt/precipitant solution to produce the final screening condition.

Condition	**Salt**	**Precipitant(s)**
1	0.3 M sodium acetate	25% (w/v) PEG 2000 MME
2	0.2 M lithium sulphate	25% (w/v) PEG 2000 MME
3	0.2 M magnesium chloride	25% (w/v) PEG 2000 MME
4	0.2 M potassium bromide	25% (w/v) PEG 2000 MME
5	0.2 M potassium thiocyanate	25% (w/v) PEG 2000 MME
6	0.8 M sodium formate	25% (w/v) PEG 2000 MME
7	0.3 M sodium acetate	15% (w/v) PEG 4000
8	0.2 M lithium sulphate	15% (w/v) PEG 4000
9	0.2 M magnesium chloride	15% (w/v) PEG 4000
10	0.2 M potassium bromide	15% (w/v) PEG 4000
11	0.2 M potassium thiocyanate	15% (w/v) PEG 4000
12	0.8 M sodium formate	15% (w/v) PEG 4000
13	0.3 M sodium acetate	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
14	0.2 M lithium sulphate	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
15	0.2 M magnesium chloride	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
16	0.2 M potassium bromide	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
17	0.2 M potassium thiocyanate	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
18	0.8 M sodium formate	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
19	0.3 M sodium acetate	8% (w/v) PEG 20000 + 8% (w/v) PEG 550 MME
20	0.2 M lithium sulphate	8% (w/v) PEG 20000 + 8% (w/v) PEG 550 MME
21	0.2 M magnesium chloride	8% (w/v) PEG 20000 + 8% (w/v) PEG 550 MME
22	0.2 M potassium bromide	8% (w/v) PEG 20000 + 8% (w/v) PEG 550

		MME
23	0.2 M potassium thiocyanate	8% (w/v) PEG 20000 + 8% (w/v) PEG 550 MME
24	0.8 M sodium formate	8% (w/v) PEG 20000 + 8% (w/v) PEG 550 MME

Buffers:

Sodium acetate, pH 4.5

Sodium acetate, pH 5.5

Sodium cacodylate, pH 6.5

Tris, pH 7.5

Tris, pH 8.5

All buffers are titrated to the specified pH using glacial acetic acid.

Clear Strategy Screen II™

Clear Strategy Screen II™ is formulated in an identical manner to Clear Strategy Screen I™ using the same buffers and volumes as above.

Condition	Salt	Preciptant(s)
1	1.5 M ammonium sulphate	
2	0.8 M lithium sulphate	
3	2 M sodium formate	
4	0.5 M potassium dihydrogen phosphate	
5	0.2 M calcium acetate	25% (w/v) PEG 2000 MME
6	0.2 M calcium acetate	15% (w/v) PEG 4000
7	2.7 M ammonium sulphate	
8	1.8 M lithium sulphate	
9	4 M sodium formate	
10	1 M potassium dihydrogen phosphate	
11	0.2 M calcium acetate	10% (w/v) PEG 8000 + 10% (w/v) PEG 1000
12	0.2 M calcium acetate	8% (w/v) PEG 20000 + 8% (v/v) PEG 550 MME
13		40% (v/v) MPD
14		40% (v/v) 1,4-Butanediol
15	0.005 M cadmium chloride	20% (w/v) PEG 4000
16	0.15 M potassium thiocyanate	20% (v/v) PEG 550 MME
17	0.15 M potassium thiocyanate	20% (v/v) PEG 600
18	0.15 M potassium thiocyanate	20% (w/v) PEG 1500
19		35% (v/v) 2-Propanol
20		30% (v/v) Jeffamine® M-600
21	0.005 M nickel chloride	20% (w/v) PEG 4000

22	0.15 M potassium thiocyanate	18% (w/v) PEG 3350
23	0.15 M potassium thiocyanate	18% (w/v) PEG 5000 MME
24	0.15 M potassium thiocyanate	15% (w/v) PEG 6000

MIDAS™

Condition	Salt/Additive	Precipitant(s)	Buffer
1		50% (v/v) polypropylene glycol 400 + 5% (v/v) dimethyl sulphoxide	0.1 M HEPES pH 6.0
2		12% (w/v) polyvinylpyrrolidone	0.1 M MES pH 5.5
3		45% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M HEPES pH 6.5
4		14% (v/v) poly(acrylic acid-*co*-maelic acid) solution	
5	0.5 M ammonium phosphate monobasic	12.5% (w/v) poly(acrylic acid sodium salt) 2100	
6		19% (v/v) poly(acrylic acid-*co*-maelic acid) solution	0.1 M Tris pH 8.5
7		10% (v/v) polypropylene glycol 400	
8		5% (w/v) poly(acrylic acid sodium salt) 2100	
9		25% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M MES pH 6.0
10	0.1 M sodium sulphate	24% (w/v) polyvinylpyrrolidone	
11	0.2 M calcium chloride dihydrate	35% (v/v) pentaerythritol ethoxylate (15/4 EO/OH)	0.1 M HEPES pH 6.5
12		35% (v/v) polypropylene glycol 400	0.1 M K/Na phosphate pH 7.0
13	0.2 M sodium chloride	20% (v/v) Jeffamine® D-2000 + 10% (v/v) Jeffamine® M-2005	0.1 M MES pH 5.5
14	0.2 M sodium thiocyanate	15% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M HEPES pH 7.0
15	0.2 M potassium acetate	5% (w/v) polyvinyl alcohol + 10% (v/v) Jeffamine® T-403	0.1 M HEPES pH 7.0
16	0.2 M sodium chloride	45% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M MES pH 6.0
17		8% (w/v) polyvinyl alcohol + 10% (v/v) 1- propanol	0.1 M HEPES pH 7.0
18	0.1 M lithium sulphate	30% (w/v) polyvinylpyrrolidone	0.1 M HEPES pH

			7.0
19		40% (v/v) polypropylene glycol 400	0.2 M Imidazole pH 7.0
20	0.06 M lithium sulphate	8% (v/v) poly(acrylic acid-*co*-maelic acid) solution + 3% (v/v) pentaerythritol ethoxylate (3/4 EO/OH)	0.1 M HEPES pH 7.5
21	0.1 M sodium chloride	35% (v/v) Jeffamine® SD-2001	0.1 M Tris pH 8.0
22		30% (v/v) Jeffamine® M-600 + 10% (v/v) dimethyl sulphoxide	
23		20% (v/v) polypropylene glycol 400 + 10% (v/v) 1-propanol	
24		28% (v/v) poly(acrylic acid-*co*-maelic acid) solution	0.1 M HEPES pH 6.5
25		15% (v/v) Jeffamine® ED-2003 + 10% (v/v) ethanol	
26	0.2 M sodium chloride	30% (v/v) Jeffamine® ED-2003	0.1 M MES pH 6.0
27	0.1 M sodium malonate dibasic monohydrate	25% (v/v) Jeffamine® SD-2001	0.1 M MES pH 5.5
28	0.2 M sodium chloride	15% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M MES pH 6.0
29	0.2 M magnesium chloride hexahydrate	35% (v/v) pentaerythritol ethoxylate (3/4 EO/OH)	
30		40% (v/v) pentaerythritol propoxylate (5/4 PO/OH) + 15% (v/v) ethanol	
31		50% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M Tris pH 8.0
32	0.2 M sodium chloride	12.5% (w/v) polyvinylpyrrolidone + 10% (w/v) PEG 4000	0.1 M Tris pH 8.0
33	0.1 M sodium chloride	25% (v/v) pentaerythritol propoxylate (5/4 PO/OH) + 10% (v/v) dimethyl sulphoxide	
34	0.2 M ammonium sulphate	35% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M HEPES pH 7.5
35	0.1 M magnesium formate dihydrate	30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH)	0.1 M Tris pH 8.5
36	0.2 M potassium acetate	24% (v/v) poly(acrylic acid-*co*-maleic acid) solution	
37		60% (v/v) polypropylene glycol 400	0.1 M Tris pH 8.0
38		30% (v/v) pentaerythritol ethoxylate (15/4 EO/OH) + 6% (w/v) polyvinylpyrrolidone	0.1 M HEPES pH 7.5

39		45% (v/v) polypropylene glycol 400 + 10% (v/v) ethanol	
40		10% (v/v) pentaerythritol ethoxylate (3/4 EO/OH) + 10% (v/v) 1-butanol	
41		12.5% (w/v) poly(acrylic acid sodium salt) 2100 + 6% (v/v) Jeffamine® SD-2001	0.1 M HEPES pH 7.0
42		6% (w/v) polyvinylpyrrolidone	0.1 M HEPES pH 6.5
43		20% (v/v) Jeffamine® ED-2003	0.1 M HEPES pH 6.5
44		20% (v/v) glycerol ethoxylate + 10% (v/v) tetrahydrofuran	0.1 M Tris pH 8.0
45		25% (v/v) Jeffamine® D-2000	0.2 M Imidazole pH 7.0
46	0.2 M potassium chloride	30% (v/v) Jeffamine® SD-2001	0.1 M HEPES pH 6.5
47	0.1 M sodium chloride	30% (v/v) polypropylene glycol 400	
48		20% (v/v) Jeffamine® SD-2001 + 15% (v/v) 1-propanol	
49	0.2 M lithium sulphate	25% (v/v) Jeffamine® T-403	0.1 M Tris pH 8.0
50	0.2 M potassium acetate	35% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	
51	0.2 M potassium chloride	20% (v/v) pentaerythritol ethoxylate (15/4 EO/OH)	0.1 M Glycine pH 9.5
52	0.2 M sodium thiocyanate	40% (v/v) pentaerythritol propoxylate (5/4 PO/OH)	0.1 M HEPES pH 7.0
53	0.2 M potassium chloride	15% (v/v) Jeffamine® T-403 + 15% (v/v) Jeffamine® ED-2003	0.1 M HEPES pH 6.5
54	0.2 M potassium acetate	15% (v/v) pentaerythritol ethoxylate (15/4 EO/OH) + 3% (v/v) Jeffamine® T-403	0.1 M MES pH 6.0
55	0.1 M sodium malonate dibasic monohydrate	30% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M HEPES pH 7.0
56		10% (v/v) Jeffamine® D-2000 + 10% (v/v) Jeffamine® M-2005 + 10% (v/v) ethanol	
57	0.1 M lithium sulphate	25% (v/v) Jeffamine® ED-2003	0.1 M Tris pH 8.0
58		10% (v/v) Jeffamine® T-403 10% (v/v) Jeffamine® ED-2003	0.1 M Tris pH 8.0
59	0.1 M lithium	25% (w/v) poly(acrylic acid sodium salt)	0.1 M

	sulphate	2100	HEPES pH 6.5
60	0.2 M magnesium chloride hexahydrate	15% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M HEPES pH 7.5
61		40% (v/v) Jeffamine® D-2000	0.1 M HEPES pH 6.5
62	0.5 M sodium chloride	10% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M Tris pH 8.0
63		14% (v/v) Jeffamine® ED-900 + 11% (v/v) Jeffamine® SD-2001	0.1 M K/Na phosphate pH 7.0
64	0.2 M sodium chloride	20% (w/v) poly(acrylic acid sodium salt) 2100	0.1 M BICINE pH 9.0
65	0.2 M sodium malonate dibasic monohydrate	20% (v/v) Jeffamine® D-2000	0.1 M MES pH 5.5
66	0.2 M potassium chloride	30% (v/v) Jeffamine® M-2070	0.1 M Tris pH 8.0
67		20% (v/v) Jeffamine® M-2070 + 20% (v/v) dimethyl sulphoxide	
68	0.2 M magnesium chloride hexahydrate	40% (v/v) pentaerythritol propoxylate (17/8 PO/OH)	0.1 M MES pH 5.5
69		20% (w/v) poly(acrylic acid sodium salt) 5100	0.1 M Tris pH 8.0
70		28% (v/v) polyethyleneimine	0.1 M HEPES pH 7.0
71	0.1 M ammonium formate	20% (v/v) SOKALAN® CP 7	0.1 M HEPES pH 7.0
72	0.2 M sodium sulphate	20% (w/v) SOKALAN® HP 56	0.1 M Tris pH 8.0
73	0.1 M potassium chloride	25% (v/v) SOKALAN® CP 7	0.1 M HEPES pH 7.0
74	0.3 M ammonium formate	20% (v/v) SOKALAN® CP 5	0.1 M HEPES pH 7.0
75		40% (v/v) glycerol ethoxylate	
76		30% (v/v) glycerol ethoxylate	0.1 M Tris pH 8.5
77		55% (v/v) polypropylene glycol 400	
78	0.2 M lithium citrate tribasic tetrahydrate	35% (v/v) glycerol ethoxylate	

79	0.2 M ammonium acetate	30% (v/v) glycerol ethoxylate	0.1 M MES pH 6.5
80		20% (w/v) SOKALAN® CP 42 + 5% (v/v) methanol	0.1 M Tris pH 8.0
81		25% (w/v) SOKALAN® CP 42 + 10% (v/v) tetrahydrofuran	0.1 M Tris pH 7.0
82	0.1 M lithium acetate dihydrate	20% (w/v) SOKALAN® CP 42	0.1 M Bis-Tris pH 6.0
83		10% (v/v) Jeffamine® M-2005	0.2 M HEPES pH 6.5
84		15% (v/v) SOKALAN® CP 5	0.1 M Bis-Tris pH 6.0
85		25% (w/v) SOKALAN® CP 42	0.1 M Bis-Tris pH 6.0
86		35% (v/v) Jeffamine® D-2000	
87		20% (v/v) glycerol ethoxylate + 3% (v/v) polyethyleneimine	0.1 M Tris pH 8.5
88	0.2 M ammonium chloride	25% (v/v) glycerol ethoxylate	0.1 M HEPES pH 7.5
89		10% (w/v) SOKALAN® CP 42	0.1 M Tris pH 8.5
90		30% (w/v) poly(acrylic acid sodium salt) 5100 + 10% (v/v) ethanol	0.1 M MES pH 6.0
91	0. 2 M potassium citrate tribasic monohydrate	15% (w/v) SOKALAN® CP 42	
92		30% (w/v) SOKALAN® CP 42	0.1 M Tris pH 8.5
93	0.2 M ammonium acetate	25% (w/v) SOKALAN® HP 56	0.1 M HEPES pH 7.0
94		25% (v/v) SOKALAN® CP 5	0.1 M Tris pH 8.5
95	0.2 M ammonium formate	10% (w/v) polyvinylpyrrolidone + 20% (w/v) PEG 4000	
96		15% (w/v) polyvinylpyrrolidone + 25% (w/v) PEG 5000 MME	0.1 M Tris pH 8.0

Hampton Additive Screen

Standard screens were performed with 0.2 µL additive combined with 1 µL protein and 0.8 µL well condition, with the exception of ethanol, isopropanol and methanol. Ethanol, isopropanol and methanol were added to the well solution (200 µL additive : 800 µL well solution) prior to drop mixing. However, after initial hits with $BaCl_2$, conditions were mixed with 200 µL $BaCl_2$: 800 µL well solution prior to drop mixing.

Additive	Concentration
$BaCl_2$	100 mM
$CdCl_2$	100 mM
$CaCl_2$	100 mM
$CuCl_2$	100 mM
$MgCl_2$	100 mM
$MnCl_2$	100 mM
$SrCl_2$	100 mM
YCl_3	100 mM
$ZnCl_2$	100 mM
Ethylene glycol	30% (v/v)
Glycerol (anhydrous)	30% (v/v)
1,6-hexanediol	30% (v/v)
MPD	30% (v/v)
PEG 400	50% (v/v)
Trimethylamine HCl	100 mM
Guanidine HCl	1 M
Urea	100 mM
1,2,3-heptanetriol	15% (v/v)
Benzamidine HCl	20% (v/v)
Dioxane	30% (v/v)
Ethanol	30% (v/v)
Isopropanol	30% (v/v)
Methanol	30% (v/v)

Appendix III – Heavy Atom Soak Solutions

Crystals were soaked in 1.8 µL stabilising solution [30% (v/v) PEG 4000, 0.2 M ammonium acetate, 0.01 M $BaCl_2$, 0.1 M sodium acetate, pH 5.0] containing 0.2 µL heavy atom solution for 10-120 min.

Heavy Atom Compound	Stock Concentration (mM)
Gold (III) chloride	100
Mercury (II) acetate	100
Methylmercury (II) chloride	100