The University Of Sheffield.

Structural and Functional Studies of the Light-Dependent Protochlorophyllide Oxidoreductase Enzyme

David Robert Armstrong

Department of Molecular Biology and Biotechnology

A thesis submitted for the degree of Doctor of Philosophy

September 2014

Abstract

The light-dependent enzyme protochlorophyllide oxidoreductase (POR) is a key enzyme in the chlorophyll biosynthesis pathway, catalysing the reduction of the C17 - C18 bond in protochlorophyllide (Pchlide) to form chlorophyllide (Chlide). This reaction involves the light-induced transfer of a hydride from the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor, followed by proton transfer from a catalytic tyrosine residue. Much work has been done to elucidate the catalytic mechanism of POR, however little is known about the protein structure. POR isoforms in plants are also notable as components of the prolamellar bodies (PLBs), large paracrystalline structures that are precursors to the thylakoid membranes in mature chloroplasts.

Bioinformatics studies have identified a number of proteins, related to POR, which contain similar structural features, leading to the production of a structural model for POR. A unique loop region of POR was shown by EPR to be mobile, with point mutations within this region causing a reduction in enzymatic activity. Production of a ^{2}H, ^{13}C, ^{15}N-labelled sample of POR for NMR studies has enabled significant advancement in the understanding of the protein structure. This includes the calculation of backbone torsion angles for the majority of the protein, in addition to the identification of multiple dynamic regions of the protein.

The protocol for purification of Pchlide, the substrate for POR, has been significantly improved, providing high quality pigment for study of the POR ternary complex. Various biophysical techniques have been used to study the macromolecular structure of these complexes, indicating the formation of large aggregates of the cyanobacterial enzyme induced by substrate binding, similar to PLBs. This has also led to the identification of ring structures, composed of 5 and 6 monomers of POR, which are likely to be the primary components of the cyanobacterial POR structures.

Acknowledgements

Firstly, I would like to thank my two supervisors Mike and Neil for their endless support and expertise, which has helped me tremendously throughout my project. I have been privileged to have two supervisors with such a vast combined knowledge of all things biochemical, and I have attempted to absorb as much of that as possible! I have also been extremely lucky to have formed a number of excellent collaborations, gaining experience in many different biophysical techniques, including: NMR with Hans Wienk in Utrecht; AUC with Steve Harding and David Morou-Besong in Nottingham; EPR with Steve Rigby, and NMR with Matt Cliff, both in Manchester; and DLS at Malvern Instruments. I am also grateful to the BBSRC for funding my research (and my Friday night pub trips!).

From the Hunter group, I must especially thank Mo for all her help and preventing me, literally, from burning down the lab. Notable mentions also go to Andy Proudfoot for laying the foundations for my project, Sarah for her collaboration on pigment work, and Cvet and Qian for their wizardry on AFM and EM, respectively. I am so glad to have made so many friends within this lab: Dave and Lizzy, who drew out the board game geek in me; Jimbo, for doing all my running at football; Gooch, in particular for winding up Ken; Ken, for being so enjoyable to wind up (top red!); and Craig, for his many philosophical 'would you rather' questions.

From the NMR group, special thanks go to Andrea, who has helped me immensely and has shown a lot of patience with me over the years. All of the NMR group have given me invaluable help during lab meetings and carried me through to the weekend with Friday cake! A shout out goes to all my excellent drinking buddies, the various Kelly lab members, those from other regions of Firth Court, and the friends I've had since undergrad. I have had some great housemates, particularly Dean, Mike and Gav, and felt safe with Ellie's Neighbourhood Watch alerts. I'm also indebted to Intra Mural for many hours of cheap football.

My best times have always come with two pairs of brilliant friends. In Sheffield, Jack and Ash, who have always been there, from the sweaty Corp nights through to the drug checks with German police. In St Albans, Paul and Mark, who never let me forget where home is - I could never stay North for more than a few weeks without coming back to see you guys.

Finally, I am so grateful for the support and love of my family. My Nana has always shown how proud she is of me and sent me back up North with many a cake! Bex, you're the best big sister anyone could hope for. Mostly though, this is for my parents, who are two of the kindest people and have always pushed me to do my best. I love you both very much.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

List of Abbreviations

ACN	Acetonitrile
ALA	δ-aminolaevulinic acid
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
Bchl	Bacteriochlorophyll
βOG	n-octyl-β-D-glucopyranoside
CBR	Carbonyl reductase
Chl	Chlorophyll
Chlide	Chlorophyllide
CMC	Critical micelle concentration
CSA	Chemical shift anisotropy
DAD	Diode array detector
DLS	Dynamic light scattering
DPOR	Light-independent protochlorophyllide oxidoreductase
DTT	Dithiothreitol
DV	Divinyl
DVR	Divinyl reductase
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
Et_2O	Diethyl Ether
HPLC	High performance liquid chromatography
HSQC	Homonuclear single quantum coherence spectrum
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KIE	Kinetic Isotope effect

LHPP	Light-harvesting Pchlide-POR complex
Mg Cyclase	Mg protoporphyrin IX monomethyl ester cyclase
MgPME	Mg protoporphyrin IX monomethyl ester
MTSL	*S*-[1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl]methyl methanesulphonothioate
MV	Monovinyl
$NADP^+$	Oxidised Nicotinamide adenine dinucleotide phosophate
NADPH	Reduced Nicotinamide adenine dinucleotide phosophate
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect
NUS	Non-uniform sampling
Pchlide	Protochlorophyllide
PDB	Protein Data Bank
PLBs	Prolamellar bodies
POR	Protochlorophyllide oxidoreductase
PRE	Paramagnetic relaxation enhancement
PTs	Prothylakoids
R. capsulatus	*Rhodobacter capsulatus*
R. sphaeroides	*Rhodobacter Sphaeroides*
RDC	Residual dipolar coupling
SAM	S-adenosyl-L-methionine
SDR	Short chain dehydrogenase/reductases
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SIE	Solvent Isotope effect
VR	Vestitone reductase
T. elongatus	*Thermosynechococcus elongatus* BP-1

$\mathbf{T_1}$	Longitudinal relaxation
$\mathbf{T_2}$	Transverse relaxation
$\boldsymbol{\tau_c}$	Correlation time
TROSY	Transverse relaxation optimised spectroscopy

Chapter 1

1. Introduction

1.1 Oxygenic photosynthesis

Photosynthesis is the process of converting the light energy from the sun into chemical energy, which subsequently sustains almost all the life on earth; chemolithoautotrophs being the only exception (Falkowski et al., 2000). The process of photosynthesis causes the fixation of carbon dioxide to carbohydrates as a store of energy. Additionally, this process reduces the levels of atmospheric carbon dioxide and increases the levels of oxygen, essential to the development of aerobic life on earth (Buick, 2008). Plants and cyanobacteria utilise oxygenic photosynthesis in order to convert solar energy to a chemical form, through electrical energy, in the reaction:

$$CO_2 + H_2O \xrightarrow{\text{Light}} [CH_2O]_n + H_2O + O_2$$

Equation 1.1 The general equation for oxygenic photosynthesis in photoautotrophs.

This is the reduction of a carbon dioxide using the electrons provided by the water molecules, producing carbohydrate. Molecular oxygen is produced as the oxidation product of this process, hence the term oxygenic photosynthesis. There are a number of other bacteria that carry out photosynthesis but with an alternative electron donor to water, for example molecular hydrogen, succinate and hydrogen sulphide, termed anoxygenic photosynthesis.

The process of photosynthesis can be split into two stages, light-dependent (photochemical) and light-independent. The light-dependent step involves the excitation of the electrons within photosynthetic pigments by photons, elevating these electrons to a higher energy state. The return of these excited electrons to a lower energy state can then provide the energy to drive production of ATP and NAD(P)H. These 'high energy' molecules can then be used in the light-independent reactions which cause the fixation of carbon dioxide to carbohydrates.

The capture of photons of light energy requires specific pigments, molecules which absorb light at various different wavelengths, specific to each pigment. Chlorophyll (Chl) is the most abundant pigment on Earth and is estimated to be synthesised, and degraded, at a rate of

approximately 10^9 tonnes per year. Other pigments that absorb light, such as carotenoids and phycobilins, are often tuned to absorb light at very specific wavelengths by the proteins they interact with, in order to utilise the maximum amount of available energy. Generally, the light at either end of the spectrum is most utilised, i.e. the high abundance yet low energy red light and the low abundance but high energy near-UV light.

Current worldwide population growth and the increase in global energy demands related to this have led to the requirement of alternative energy sources. This has thrust photosynthesis into increasingly high scientific prominence due to the great possibilities of harnessing energy from the sun using microorganisms. Of the current available solar energy available to earth, cyanobacteria harvest just 0.2 - 0.3 % (Waterbury et al., 1979); this is equivalent to approximately 25 times the current global energy consumption by humans. If we could effectively harvest even a small proportion of this energy then it would go a long way to providing a sustainable global energy source for our developing society.

1.2 Photosynthetic Bacteria

Due to the rapid growth, relatively low complexity, and ease of genetic manipulation, photosynthetic bacteria have been widely used as model systems to study photosynthesis. Photosynthetic bacteria are generally found in freshwater or marine environments and can be classified in 5 groups:

Green bacteria, further subdivided into sulphur (*Chlorobi*) and non-sulphur (*Chloroflexi*) forms, the former utilising sulphide or thiosulphide compounds as electron donors while the latter employs organic compounds for this purpose (Bryant and Frigaard, 2006).

Heliobacteria, obligate anaerobes that utilise organic carbon compounds as electron donors (Bryant and Frigaard, 2006).

Purple Sulphur Bacteria (*Chromatiales*), which utilise either sulphur or H_2 as electron donors, though little is known about these organisms (Bryant and Frigaard, 2006).

Purple Non-Sulphur Bacteria (*Rhodospirillaceae*) require simple organic compounds as electron donors. Many species have had their genomes sequenced, including *Rhodobacter sphaeroides* and *Rhodobacter capsulatus* (Bryant and Frigaard, 2006).

Cyanobacteria are oxygenic phototrophs with a number of well characterised species, including *Synechocystis*, *Synechococcus* and *Prochlorococcus* (Bryant and Frigaard, 2006).

The differing metabolic requirements of these bacteria mean that each has developed photosynthetic apparatus that is uniquely tailored to optimise the harvesting of light energy.

1.2.1 Cyanobacteria

Cyanobacteria are also referred to as "blue-green algae" and comprise a group containing over 1500 species. *Synechocystis sp.* PCC6803 (from here on *Synechocystis*) is a cyanobacterium capable of both phototrophic growth, when grown in the light, and heterotrophic growth, when grown in the dark with a carbon source such as glucose. *Synechocystis*, along with other Cyanobacteria, has been studied intensively as a model organism for photosynthesis. This is due to the highly similar organisation of photosynthetic apparatus to that of plants, the simplified genetic material compared to plants and the ability to perform oxygenic photosynthesis unlike other photosynthetic bacteria. The genome of *Synechocystis* was one of the first to be sequenced and it was the first phototrophic organism to be mapped (Kaneko and Tabata, 1997).

Thermosynechococcus elongatus BP-1 (from here on *T. elongatus*) is another cyanobacterium that is capable of both phototrophic and heterotrophic growth. Originally isolated from a hot spring near Beppu in Japan, it is a unicellular rod-shaped cyanobacterium with an optimum growth temperature of 55 °C and a fully sequenced genome (Nakamura et al., 2002). *T. elongatus* has long been used as a model organism for photosynthesis, for many of the same reasons as *Synechocystis* but additionally due to its thermostable nature, and thus the advantages for structural studies that this thermostability provides. Using gene products from this organism, successful attempts to obtain structural information of proteins and protein complexes have been documented (Ferreira et al., 2004; Jordan et al., 2001; Zouni et al., 2001).

1.3 Structure of chlorophylls

Tetrapyrroles are composed of four pyrrole rings (C_4H_4NH), connected to each other by methine bridges from position 2 of one ring to position 5 of the next. Porphyrins are cyclic tetrapyrroles and include haem, chlorophyll and bacteriochlorophyll, although pyrroles can also be linked linearly, for example in phycobilins and bilanes. The general structure of all cyclic tetrapyrroles is similar, with the major difference being in the identity of the central coordinating divalent metal ion, chelated into the structure at various important branch points of the tetrapyrrole biosynthetic pathways.

Chlorophylls are part of the group of pigments called Chlorins, the most abundant of which are Chl *a* (Fig. 1.1) and Chl *b*; bacteriochlorophylls (Bchl) (Fig.1.1) are the chief pigments utilised by phototrophic bacteria. These are modified tetrapyrroles, containing a fifth isocyclic ring and a long chain esterifying alcohol at position C17. There are various derivatives of Chl with differing side chain groups but all are characterised by the coordination of a central Mg^{2+} ion, in contrast to the coordination of Fe^{2+} by haem (Fig. 1.1) and sirohaem, and Co^{2+} by coenzyme B12. Both *Synechocystis* and *T. elongatus* utilise solely Chl *a* for energy transfer.

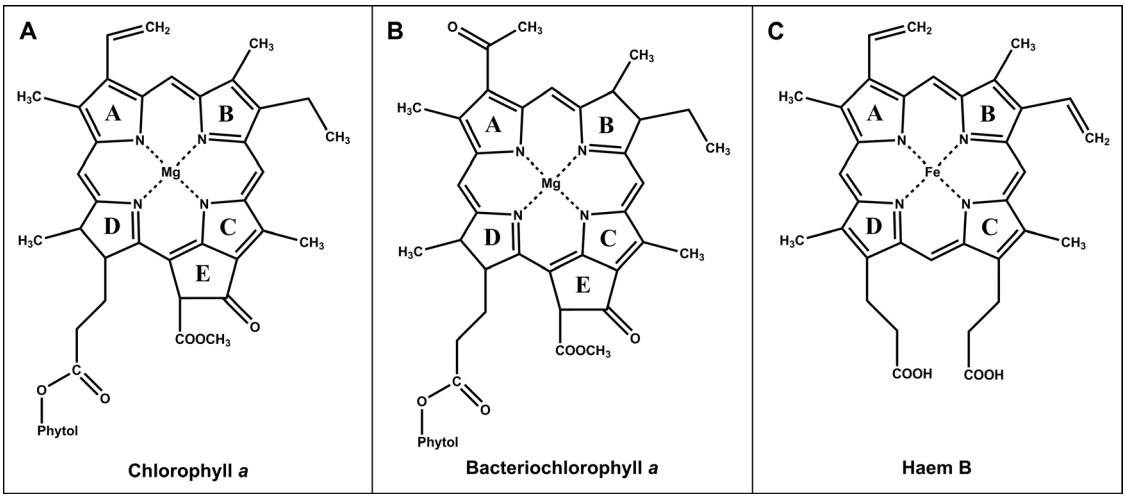

Figure 1.1 Molecular structures of notable porphyrin molecules: (A) Chlorophyll *a*; (B) Bacteriochlorophyll *a*; (C) Haem B.

1.4 Common steps in tetrapyrrole biosynthesis

1.4.1 Formation of δ-aminolaevulinic acid

The formation of δ-aminolaevulinic acid (ALA) is the first committed step of tetrapyrrole biosynthesis, with ALA the common precursor for all tetrapyrrole biosynthesis. It is produced via two distinct pathways: the Shemin pathway (Fig. 1.2), involving the condensation of Succinyl-CoA and Glycine by the enzyme δ-aminolevulinate synthase, and the C_5 pathway (Fig. 1.3), a three step process converting the precursor glutamate to ALA via a tRNA linked intermediate. The formation of ALA can be considered as the first regulatory point of the chlorophyll biosynthesis pathway.

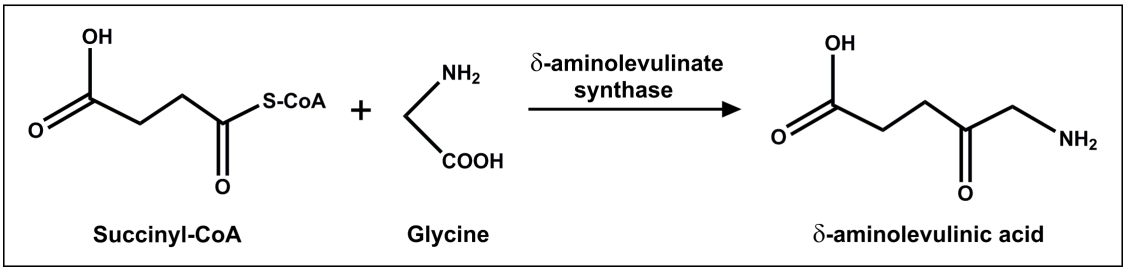

Figure 1.2 Conversion of Succinyl-CoA to ALA, by the enzyme δ-aminolevulinate synthase, via the Shemin pathway.

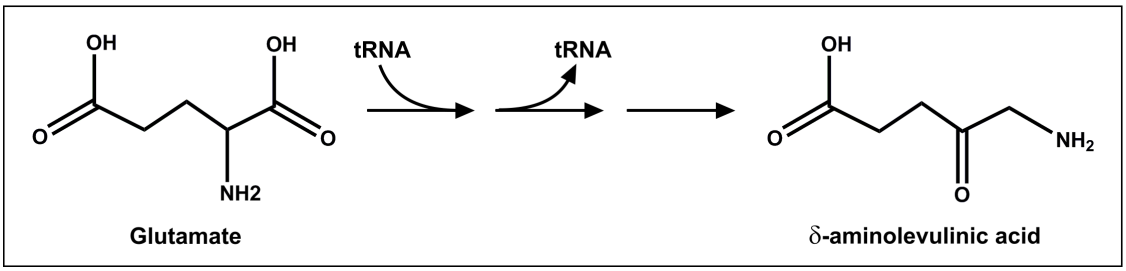

Figure 1.3 Conversion of glutamate to ALA, along the C_5 pathway, via a tRNA-bound intermediate.

1.4.2 δ-aminolaevulinic acid to porphobilinogen

The first reaction of the common tetrapyrrole biosynthetic pathway involves the asymmetric condensation of two molecules of δ-aminolevulinic acid to form porphobilinogen (Fig. 1.4), catalysed by δ-aminolevulinic acid dehydratase (Granick, 1954; Schmid and Shemin, 1955; Gibson et al., 1955). In various organisms the enzyme has been shown to have a relatively high homology, indicating that the catalytic mechanism is likely to be identical throughout all organisms (Willows and Kriegel, 2009a). All characterised δ-aminolevulinic acid dehydratase isoforms have been identified as metalloenzymes, binding either Mg^{2+} or Zn^{2+} ions, which are shown to be essential for catalysis (Jaffe et al., 1992); though their exact function is unknown.

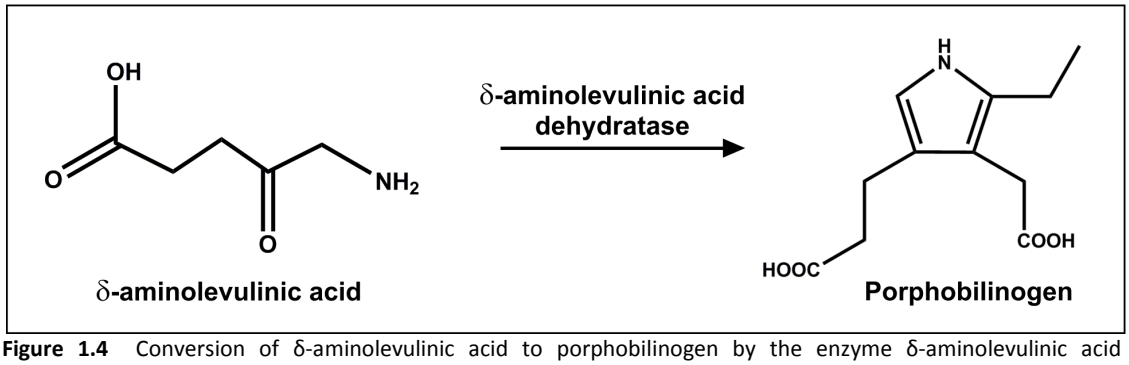

Figure 1.4 Conversion of δ-aminolevulinic acid to porphobilinogen by the enzyme δ-aminolevulinic acid dehydratase.

1.4.3 Porphobilinogen to hydroxymethylbilane

The next step of the pathway involves the condensation of four molecules of porphobilinogen by the enzyme hydroxymethylbilane synthase to form hydroxymethylbilane (Fig. 1.5), a highly unstable linear tetrapyrrole intermediate (Burton et al., 1979; Jordan et al., 1979). The hydroxymethylbilane molecule has been shown to assemble in an ordered fashion, starting with ring A and continuing clockwise through rings B, C and D, respectively (Battersby et al., 1979; Jordan and Seehra, 1979). The crystal structure of hydroxymethylbilane synthase has been solved, indicating three domains attached via linkers, suggesting a level of protein flexibility (Jordan and Warren, 1987); and an active site cleft large enough to enclose the full product (Louie et al., 1992, 1996).

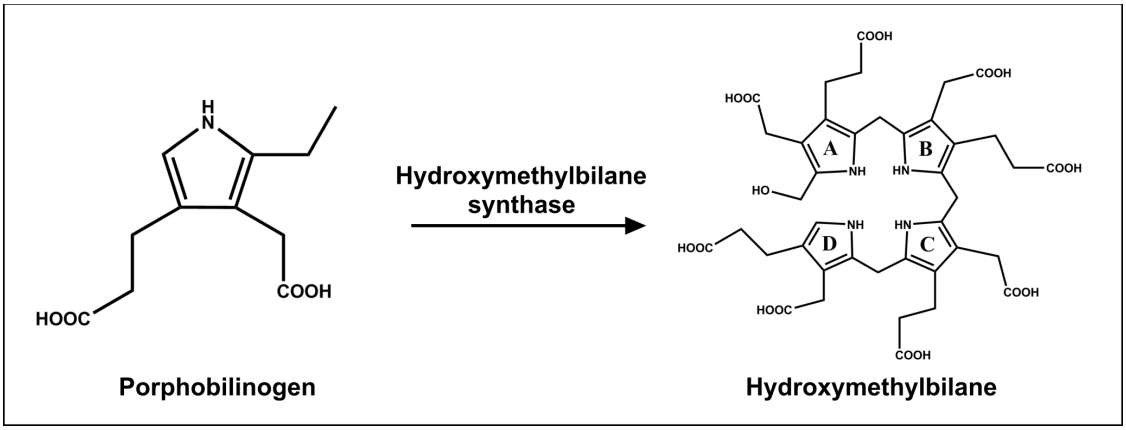

Figure 1.5 Assembly of porphobilinogen molecules into hydroxymethylbilane by the enzyme hydroxymethylbilane synthase.

1.4.4 Hydroxymethylbilane to uroporphyrinogen III

Hydroxymethylbilane is cyclised to form the first macrocyclic tetrapyrrole of the pathway, uroporphyrinogen III, by the enzyme uroporphyrinogen III synthase (Fig. 1.6). In the absence of the enzyme, hydroxymethylbilane is rapidly converted, non-catalytically, to the biologically inactive isomer uroporphyrinogen I (Battersby et al., 1979). The catalysed reaction requires the D ring of hydroxymethylbilane to be inverted, forming uroporphyrinogen III via a spiro-pyrrolenine intermediate (Crockett et al., 1991; Stark et al., 1993; Mathews et al., 2001).

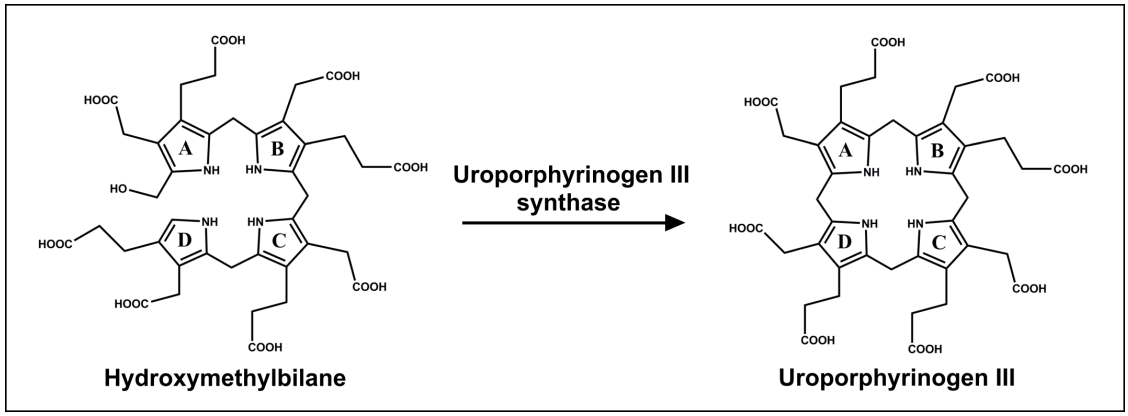

Figure 1.6 Closure of the porphyrin ring by Uroporphyrinogen III synthase, converting hydroxymethylbilane to uroporphyrinogen III.

1.4.5 Uroporphyrinogen III to coproporphyrinogen III

The next step of tetrapyrrole biosynthesis is the first branchpoint of the pathway as the sirohaem and vitamin B_{12} branches split from that of Chl, Bchl and haem. The Chl branch is continued by the conversion of uroporphyrinogen III to coproporphyrinogen, catalysed by uroporphyrinogen III decarboxylase (Fig. 1.7). The reaction involves the decarboxylation of the four acetate groups in a stepwise manner, thought to occur clockwise from ring D until all the groups are decarboxylated to methyl groups (Luo and Lim, 1993).

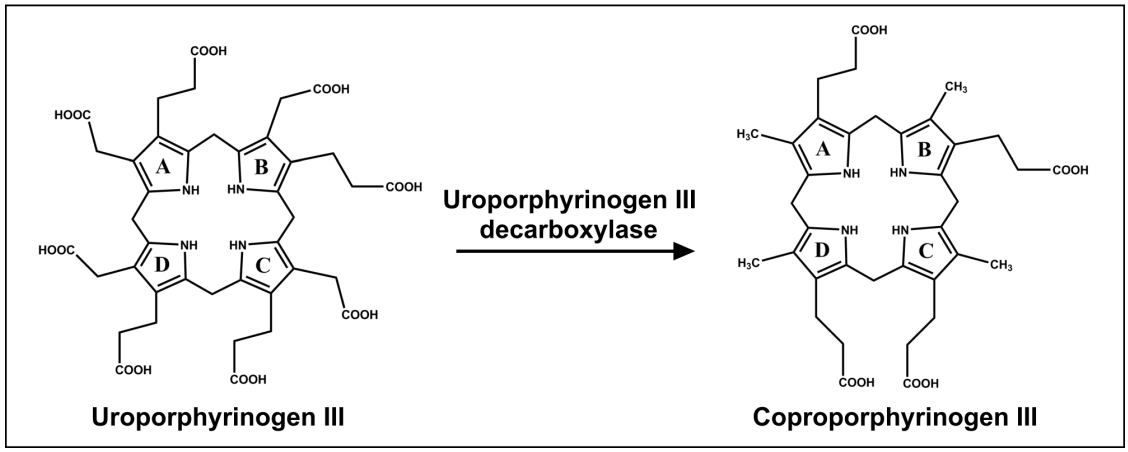

Figure 1.7 Stepwise decarboxylation of the 2', 7', 12' and 18 acetate chains of uroporphyrinogen III to form coproporphyrinogen III, catalysed by Uroporphyrinogen III decarboxylase.

1.4.6 Coproporphyrinogen III to protoporphyrinogen IX

The enzyme Coproporphyrinogen oxidase catalyses the conversion of Coproporphyrinogen III to protoporphyrinogen IX (Fig. 1.8). The propionate groups of ring A and B of coproporphyrinogen undergo oxidative decarboxylation to form vinyl groups, with studies suggesting that the vinyl group of ring A is formed first (Cavaleiro et al., 1974; Elder et al.,

1978). There are two forms of the enzyme, an oxygenic form and an anaerobic form, which uses a terminal electron acceptor other than oxygen (Tait, 1969, 1972; Breckau et al., 2003).

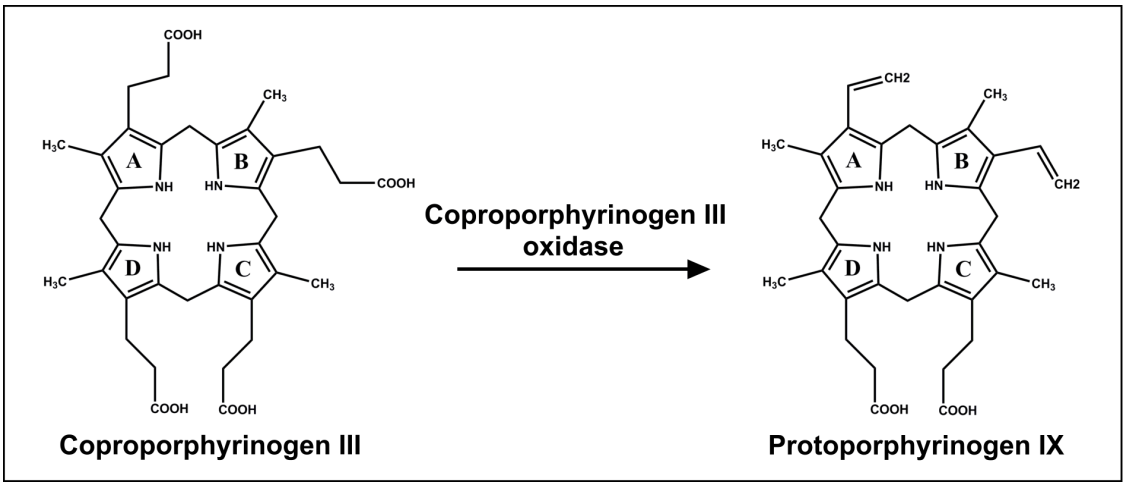

Figure 1.8 Oxidation of the 3' and 8' propionate chains of coproporphyrinogen III to vinyl groups to form protoporphyrinogen IX; catalysed by the enzyme coproporphyrinogen III oxidase.

1.4.7 Protoporphyrinogen IX to protoporphyrin IX

This step is the final common reaction between the haem and Chl pathways, with protoporphyrinogen oxidase catalysing the six electron oxidation of protoporphyrinogen IX to protoporphyrin IX (Fig. 1.9). There are two unrelated forms of protoporphyrin oxidase both requiring a flavin cofactor, the oxygen-dependent form (Siepker et al., 1987; Camadro et al., 1994), and the oxygen-independent form, which uses either nitrate or fumarate as the terminal electron acceptor (Jacobs and Jacobs, 1975, 1976; Siepker et al., 1987). There have also been studies which indicate that the oxidation can occur spontaneously without the need for enzymatic catalysis (Porra and Falk, 1964; Sano and Granick, 1961; Porra and Falk, 1961; Jacobs and Jacobs, 1975, 1976). This step leads to the formation of a conjugated electron system around the whole porphyrin ring, essential to the electronic properties of the tetrapyrrole molecules.

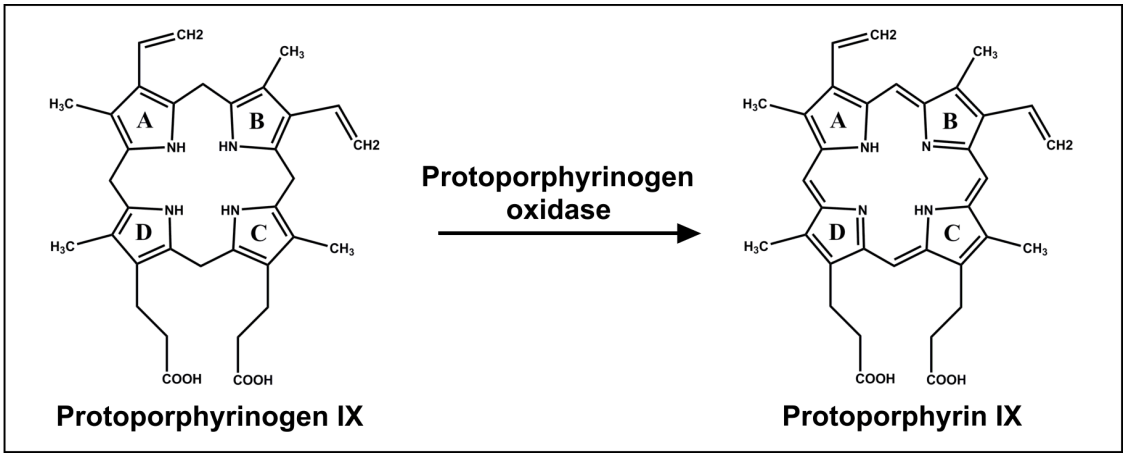

Figure 1.9 Oxidation of protoporphyrinogen IX to form the conjugated ring structure of protopophyrin IX, by the enzyme Protoporphyrinogen oxidase.

1.5 The haem and chlorophyll branch point

The direction the pathway takes from this point is determined by which ion is chelated into the ring structure. The chelation of Fe^{2+} by the enzyme ferrochelatase leads to the formation of protohaem or haem B (Fig. 1.10), the last reaction in the haem biosynthesis pathway. If the chelated metal is an Mg^{2+} ion, catalysed by the enzyme Magnesium chelatase, then the pathway is committed to the biosynthesis of chlorophyll. Both protohaem and chlorophyll are essential in photosynthetic cells but are required in varying quantities therefore these two reactions must be highly regulated, and the substrate protoporphyrin IX carefully shuttled to the correct locations.

There are a number of proposed control mechanisms for regulating flux through each pathway. The redox states in the chloroplasts, indicating the level of photosynthetic activity, may regulate the branchpoint enzymes; with the ChlI subunit of the magnesium chelatase (MgCh) identified as a thioredoxin-target protein. MgCh activity is also likely to be regulated by the ATP/ADP ratio and the concentration of free Mg^{2+}: ATP is required for activation of MgCh and also inhibits ferrochelatase activity; Mg^{2+} concentration implicated in the activation or inactivation, as well as localisation, of MgCh (Tanaka and Tanaka, 2007).

1.5.1 Ferrochelatase

The insertion of a Fe^{2+} ion into protoporphyrin IX is energetically favourable under standard conditions, however cannot occur without the presence of the ferrochelatase enzyme. The reaction is therefore ATP-independent and there have been studies that indicate that the presence of ATP actually inhibits the reaction (Cornah et al., 2002). The enzyme is a membrane-associated protein, found as a homodimer in eukaryotes but as a monomer in

bacteria (Wu et al., 2001; Grzybowska et al., 2002). The eukaryotic and bacterial forms also differ in the presence and absence, respectively, of iron-sulphur [2Fe-2S] clusters, however both appear to share the same method of chelation (Dailey and Dailey, 2002).

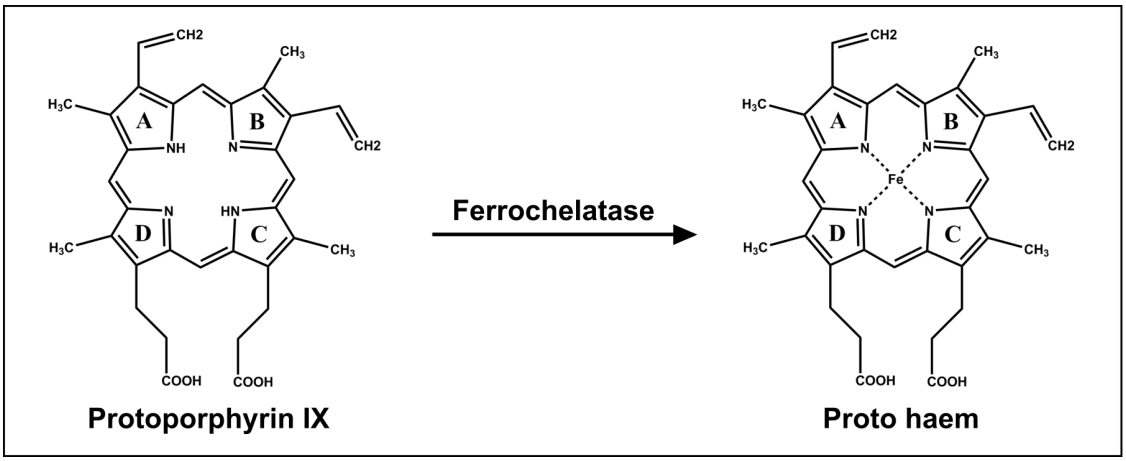

Figure 1.10 Coordination of iron into the protoporphyrin IX molecule by Ferrochelatase, to form proto haem.

1.6 Protoporphyrin IX to Mg protoporphyrin IX

The step which commits the tetrapyrrole to the chlorophyll biosynthesis pathway is the chelation of Mg^{2+} by the enzyme magnesium chelatase (Fig. 1.11) (Walker and Willows, 1997; Reid and Hunter, 2002). Unlike Fe^{2+} chelation, the process of Mg^{2+} chelation is energetically unfavourable, due to the shell of water molecules surrounding the ion which must be removed. This energy intensive removal of water molecules means that the enzyme catalysed reaction requires 15 molecules of ATP per Mg^{2+} to proceed (Gibson et al., 1995; Jensen et al., 1996; Walker and Willows, 1997).

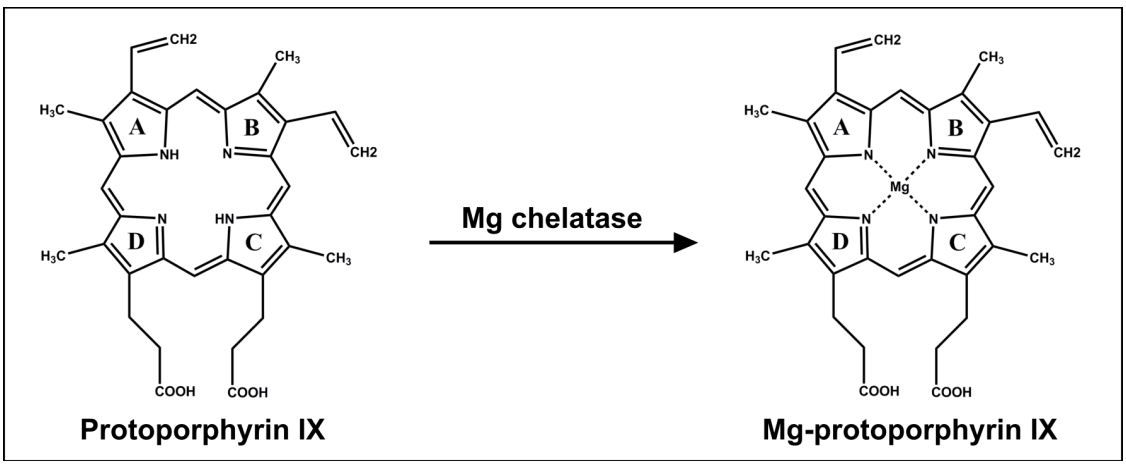

Figure 1.11 Coordination of magnesium into the protoporphyrin IX molecule, by Mg chelatase, to form Mg-protoporphyrin IX. This commits the pathway to the formation of chlorophyll.

The magnesium chelatase enzyme has been found to be composed of 3 essential subunits, designated I, D and H, products of the genes *bch/chl I, bch/chl D* and *bch/chl H*, respectively, with each well conserved between purple bacteria and cyanobacteria (Coomber et al., 1990; Jensen et al., 1996; Suzuki et al., 1997). The smaller I and D subunits, with molecular weights across organisms of 60 - 87 and 37 – 46 kDa respectively, have been shown to contain several ATP binding sites, and thus have been placed in the AAA superfamily, members of which are known to form nucleotide dependent ring structures (Willows and Kriegel, 2009b). The N-terminus of the D subunit has around 40 % identity to the I subunit, thus it has been proposed that it may have evolved as a gene duplication of *bch/chl I* followed by a gene fusion event (Jensen et al., 1996). The H subunit is the largest of the three at 120 - 155 kDa, and is the subunit responsible for porphyrin binding.

Both I and D subunits have been shown to individually form hexameric rings in the presence of Mg-ATP, though hexamers of D require I to be present in order to maintain their stability (Lake et al., 2004; Willows et al., 2004). Formation of the I-D complex inhibits the ATPase activity of I (Lake et al., 2004), though this induces the association of the H subunit and the I-D complex, stimulating ATPase activity and resulting in Mg^{2+} chelation (Jensen et al., 1998). The reaction is regulated by the presence of the genome uncoupled (*gun*) protein, Gun4, which restores activity at low Mg^{2+} or protoporphyrin IX levels (Reid and Hunter, 2004; Davison et al., 2005). Gun4 has been shown to bind both the magnesium chelatase H subunit, and Mg protoporphyrin IX, the latter more tightly; therefore it has been suggested it may play a role in chaperoning the product to the next enzyme in the pathway (Larkin et al., 2003).

1.7 Mg protoporphyrin IX to Mg protoporphyrin IX monomethyl ester

The enzyme Mg-protoporphyrin IX methyltransferase (ChlM/BchM) catalyses the transfer of a methyl group from the cofactor *S*-adenosyl-*L*-methionine (SAM) to the propionate chain on ring C of Mg-protoporphyrin IX (Fig. 1.12) (Jensen et al., 1999). It is thought that this methyl group protects the propionate group from spontaneous decarboxylation prior to the formation of ring E in the next step of the pathway (Beale, 1999). It is likely that this is another regulatory step in the Chl biosynthesis pathway, due to the effect of a ChlM/BchM knockout on expression of proteins further down the pathway; suggesting that the substrate Mg-protoporphyrin IX acts as a negative effector for the pathway (Pontier et al., 2007). There have also been studies that indicate that the H subunit of magnesium chelatase enhances the activity of ChlM/BchM, implying that Mg-protoporphyrin IX may be channelled directly from

magnesium chelatase to Mg-protoporphyrin IX methyltransferase (Hinchigeri et al., 1997; Masuda, 2008).

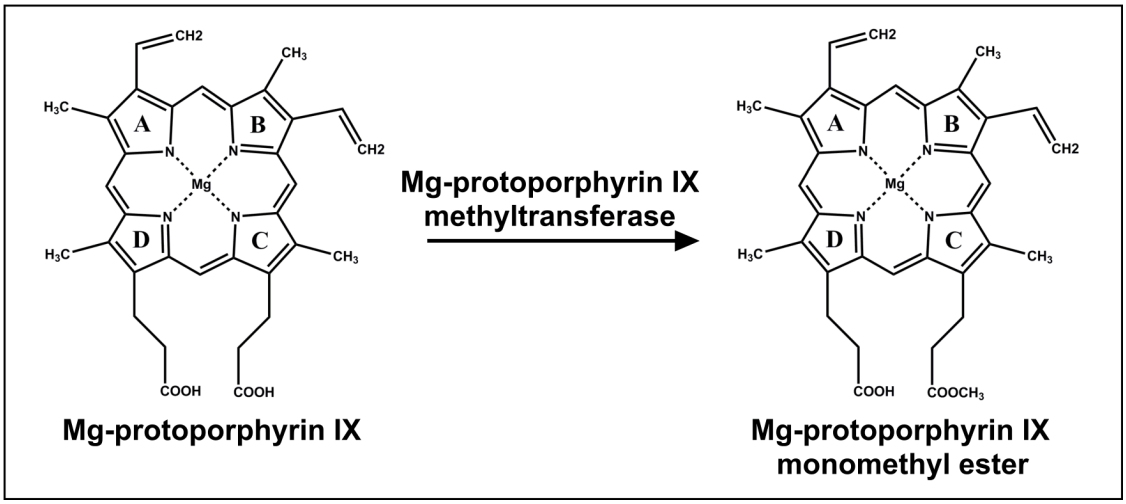

Figure 1.12 Esterification of the C13 chain of Mg-protoporphyrin IX, by Mg-protoporphyrin IX methyltransferase, to form the monomethyl ester.

1.8 Formation of the isocyclic ring

The formation of the isocyclic ring E of chlorophyll is catalysed by the enzyme Mg-protoporphyrin IX monomethyl ester cyclase (Fig. 1.13), converting Mg-protoporphyrin IX monomethyl ester to divinyl protochlorophyllide *a* (Suzuki et al., 1997). This is a multi step reaction that initially involves oxidation of the first carbon atom of the propionate side chain, incorporating oxygen to form a ketone group, followed by cyclisation of the propionate chain to the meso carbon, thought to occur via a free radical intermediate. The formation of ring E leads to a colour change of the pigment from red to green, due to the absorption in the red part of the visible spectrum.

There have been two types of cyclase identified, aerobic and anaerobic, which incorporate molecular oxygen and atomic oxygen, from water, respectively. There have been two genes discovered in cyanobacterial species, *bchE* and *acsF*, that are thought to encode the anaerobic and aerobic cyclase respectively. Though no structures exist for either protein, from sequence analysis and structure predictions it is expected that the two forms share very little structural homology.

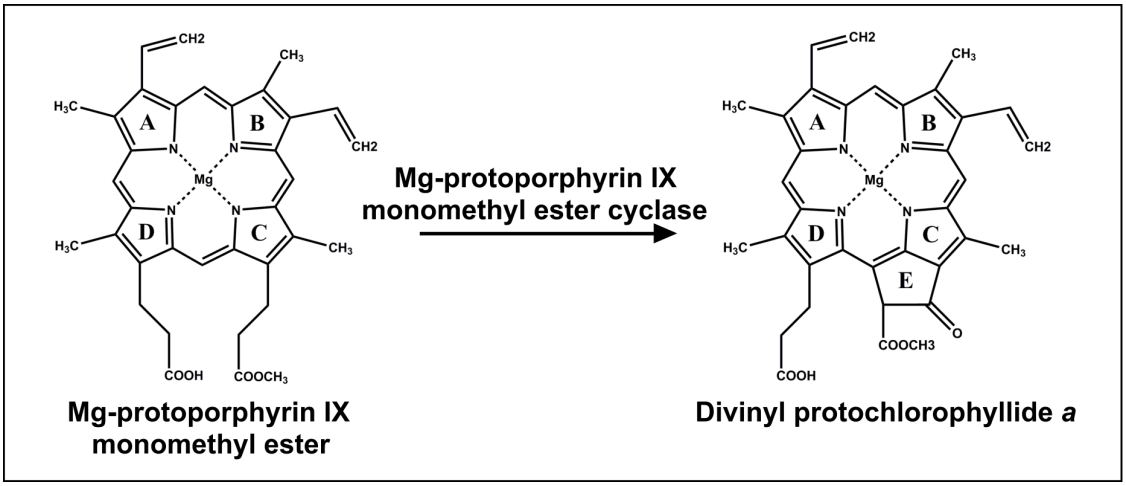

Figure 1.13 Cyclisation of the C13 chain of Mg-protoporphyrin IX monomethyl ester, along with the oxidation at position C13[1], to form divinyl Protochlorophyllide *a*. This reaction is catalysed by the multisubunit enzyme Mg-protoporphyrin IX monomethyl ester cyclase.

The *bchE* gene contains a CXXXCXXC motif forming a 4Fe-4S cluster, a feature exhibited by all free radical SAM enzymes (Sofia et al., 2001). It has therefore been suggested that the ring formation in the anaerobic cyclase may occur after the generation of a free radical on the propionate side chain (Gough et al., 2000; Sofia et al., 2001). Analysis of the *acsF* gene has indicated that it contains a conserved putative bi-nuclear-iron-cluster motif. The proposed mechanism for the aerobic cyclase is that it is the meso proton of Mg-protoporphyrin IX monomethyl ester that is radicalised, to form a 13^1-radical. A further electron is then thought to be removed to produce a 13^1-cation which can then react with the side chain to form the isocyclic ring (Pinta et al., 2002).

Recently, Hollingshead *et al.* have discovered another protein in *Synechocystis* which is important for correct function of the cyclase. The gene products of *Synechocystis* genes Sll1214 and Sll1874, homologous to *acsF*, were found to associate with the gene product of the *ycf54* gene. Mutation of this gene resulted in significantly impaired growth and an accumulation of chlorophyll biosynthesis intermediates prior to the cyclase gene. The authors suggest that Ycf54 may have a role in the assembly of the cyclase complex or even a larger association with the enzymes preceding this step (Hollingshead et al., 2012).

1.9 Protochlorophyllide to Chlorophyllide

The enzyme Protochlorophyllide oxidoreductase catalyses the reduction of the C17-C18 double bond in protochlorophyllide to a carbon-carbon single bond, to produce chlorophyllide (Lebedev and Timko, 1998); this reaction can occur before or after the divinyl reductase (DVR) reaction (detailed 1.10), with both monovinyl and divinyl isoforms of Pchlide

found to be equally suitable substrates for POR (Heyes et al., 2006). It was originally observed that dark-grown angiosperms produced etiolated yellow leaves, as they failed to produce chlorophyll, indicating a light-dependent component to the biosynthesis pathway. This led to the discovery of the light-dependent NADPH:protochlorophyllide oxidoreductase (POR) enzyme which was shown to require light to catalyse the reduction of the C17-C18 bond (Griffiths, 1978). It has also been shown that organisms that produce bacteriochlorophyll do not require light for its synthesis, leading to the discovery of the light-independent, or dark-operative, protochlorophyllide oxidoreductase (DPOR) (Bauer et al., 1993). The presence of a light-dependent POR in the anoxygenic bacterium *Dinoroseobacter shibae* has recently been reported, possibly obtained from cyanobacteria (Kaschner et al., 2014).

1.9.1 Light-independent protochlorophyllide oxidoreductase (DPOR)

The chlorophyll biosynthesis pathway, utilised in higher plants and cyanobacteria such as *T. elongatus*, and the bacteriochlorophyll biosynthesis pathway, found in purple bacteria (eg. *R. Sphaeroides*), green sulphur bacteria and green non-sulphur bacteria, initially proceed via almost identical steps (Chew and Bryant, 2007). These pathways, however, diverge at the point of Pchlide reduction, with each employing unique reactions; while chlorophyll biosynthesis proceeds primarily via the light-dependent reduction of Pchlide by LPOR, the bacteriochlorophyll biosynthesis pathway utilises a light-independent POR (DPOR); (Lebedev and Timko, 1998). These two forms of POR are structurally unrelated and proceed via two very distinct mechanisms, indicating the independent evolution of these two proteins (Masuda and Takamiya, 2004; Muraki et al., 2010).

1.9.1.1 DPOR subunits and homology to nitrogenase

DPOR is composed of three subunits, encoded in *R. sphaeroides* by the *bchL, bchN* and *bchB* genes; some oxygenic Chl *a*-containing organisms also contain a DPOR enzyme, encoded by the *chlL, chlN* and *chlB* genes. These three subunits of DPOR are homologs of the component subunits of the nitrogenase enzyme subunits *nifH, nifD* and *nifK*, respectively (Burke et al., 1993). There are two components of DPOR: L-protein, a dimer of BchL which forms the reductase component, and NB-protein, a heterotetramer of BchN and BchB forming the catalytic component; these are again homologous to the Fe and MoFe proteins of nitrogenase, respectively (Fujita and Bauer, 2000).

1.9.1.2 Structure of L-protein

The structure of the BchL dimer, comprising the L-protein component of DPOR, has been determined to a resolution of 1.6 Å (Sarma et al., 2008). The L-protein from *R. sphaeroides* was crystallised with MgADP bound; the MgATP binding site, along with the [4Fe-4S] cluster, are highly conserved with those of the nitrogenase enzyme. Despite the high structural similarity of the L-protein with the Fe protein, it cannot functionally replace this nitrogenase component when involved in *in vitro* assays with the MoFe component of nitrogenase. The L-protein component does not associate with the MoFe component, likely due to distinct charge distributions between the docking sites of L-protein and Fe protein; thus L-protein cannot facilitate electron transfer between itself and the MoFe protein (Sarma et al., 2008).

1.9.1.3 Structure of NB-protein

The crystal structure of the catalytic component of DPOR, from *R. capsulatus*, was determined for both the Pchlide-bound and free forms of NB-protein (Fig. 1.14); with resolutions of 2.3 Å and 2.6 Å, respectively (Muraki et al., 2010). The BchB and BchN proteins form a heterodimer, which further associates to form a $BchB_2$-$BchN_2$ heterotetramer. At the interface of each BchB-BchN dimer is a [4Fe-4S] cluster (the NB-cluster); coordinated by 3 cysteine residues, from BchN, and an aspartate residue, from BchB. The cysteine residues were found to be essential for the formation of the complex; however the aspartate residue appeared to have no role in complex formation, though it is important for enzymatic activity (Muraki et al., 2010).

Unsurprisingly, the Pchlide molecule in the substrate-bound structure is positioned within a cavity composed of hydrophobic residues (Fig. 1.15). The side chains of two residues from the BchN protein, Trp 387 and Phe 393, are reoriented significantly upon binding of the Pchlide molecule, forming part of this hydrophobic cavity. The central region, from residues 405 to 409, of the C-terminal α-helix of BchB becomes partly unwound upon substrate binding; this enables the interaction of Met 408, Gly 409 and Leu 410 of this region with the Pchlide molecule, with this region acting as a 'lid' for the binding pocket. This binding cavity distorts the C17 propionate of the Pchlide molecule, causing it to position itself almost perpendicular to the plane of the porphyrin ring. This enables the propionate of the Pchlide to act as its own proton donor to the C18 position; while Asp 274, from BchB, appears most likely to act as proton donor to the C17 position (Muraki et al., 2010). Interestingly, the C-terminal domain of the neighbouring BchB-BchN heterodimer also forms part of the binding pocket, in a 'domain-swapping' structure (Reinbothe et al., 2010).

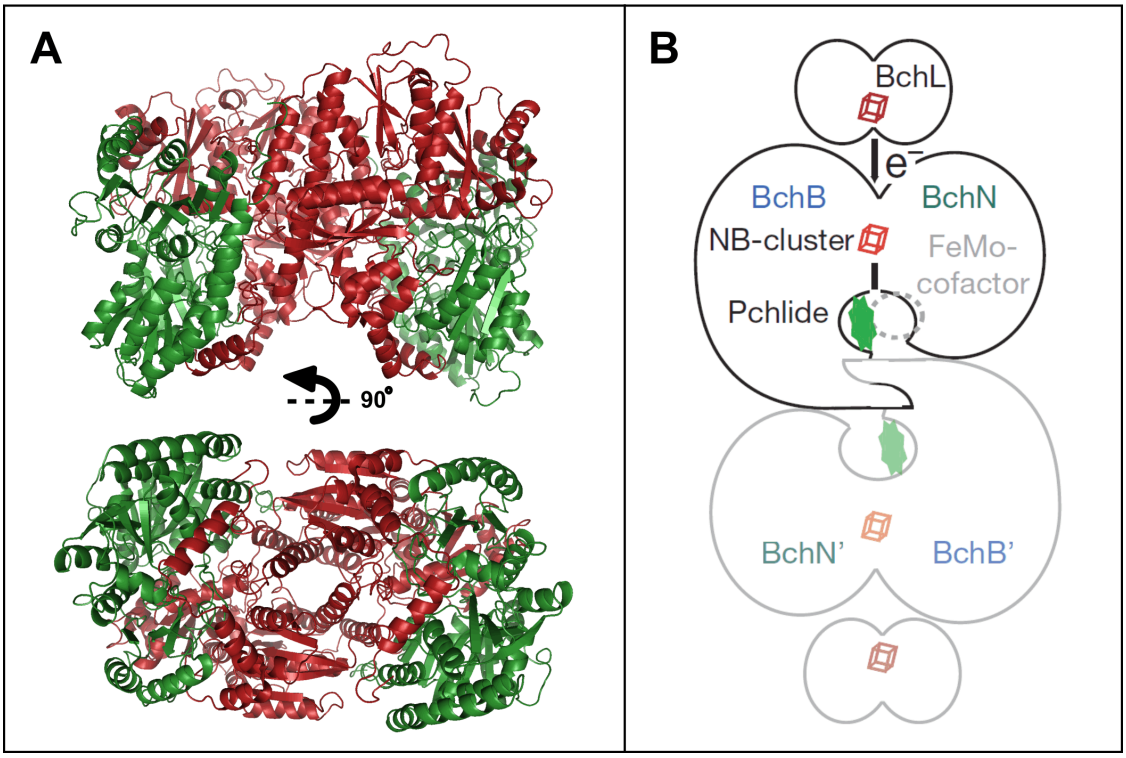

Figure 1.14 (A) Structure of DPOR NB-protein tetramer from two views. BchB subunits are coloured red, BchN subunits are coloured green. (B) Schematic model indicating how L and NB-proteins could associate. Two NB proteins are likely to associate via contacts between BchB subunits. L-proteins associate at either side of this complex, with electrons shuttling towards the centre of the complex where the Pchlide molecules are located (Muraki et al., 2010).

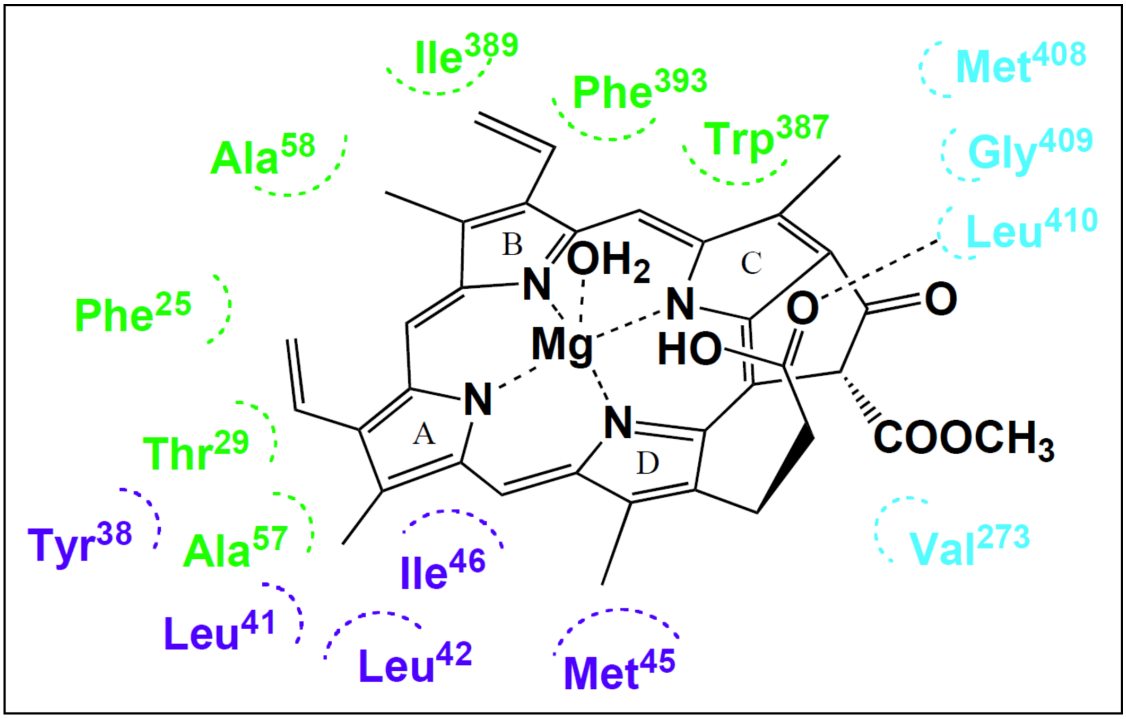

Figure 1.15 Schematic diagram indicating the residues in the DPOR complex that form the hydrophobic pocket for Pchlide binding. Residues from BchN and BchB are shown in green and blue, respectively, and residues from the exchanging BchB (BchB') in light blue (Muraki et al., 2010).

1.9.1.4 Redox function of DPOR

Though the donation of protons to the C17 and C18 positions of the Pchlide is likely to stem from the protein side-chains and substrate, this does not incorporate an electron transfer, of which two are required. It is likely that the DPOR electron transfer proceeds as it does in the nitrogenase enzyme, with electrons transferred from the L-protein [4Fe-4S] cluster, to the NB-cluster, and finally to the substrate; this is an ATP-dependent process. The reaction mechanism proposed by Muraki *et al.* involves the initial transfer of a single electron to the substrate; this is followed by the transfer of protons from the Asp 274 side-chain and the propionate side-chain, either simultaneously or successively. This produces a cation radical of the pigment, to which a second electron can be transferred by the same route described above (Muraki et al., 2010; Nomata et al., 2014).

1.9.2 Light-dependent protochlorophyllide oxidoreductase

The focus of this thesis is the enzyme protochlorophyllide oxidoreductase (POR) which catalyses the light dependent conversion of Pchlide to Chlide. The reduction of the C17-C18 double bond does not require ATP, but instead involves the absorption of light by the pigment, making it one of only two enzymes which require light for catalysis; the other being DNA photolyase which requires light to catalyses the repair of pyrimidine dimers in DNA. The POR reaction is likely to be a key regulatory step in chlorophyll biosynthesis and the subsequent assembly of the photosynthetic apparatus, with its ability to respond to changes in light intensity, an obvious control mechanism.

Due to the light dependent mechanism of the POR reaction, it has become an attractive model to studying this type of reaction mechanism, due to the ability to initiate the reaction with a flash of light. The hydride and proton transfer in the reduction reaction is characteristic of nearly all reductase reactions, therefore the study of these mechanisms of transfer is of great interest. By preparing samples in the dark it is possible to form the active complex and initiate any experiments, such as ultrafast spectroscopic analysis, from the very start of catalysis; this has enabled new insights into the mechanism which are not possible with conventional enzymes.

1.9.3 POR in plants

Mature chloroplasts develop in two different pathways, depending on conditions: in the light, chloroplasts develop from colourless proplastids; while, in the absence of light, proplastids first pass through an etioplast stage prior to chloroplast formation (Solymosi and Schoefs, 2008). Like chloroplasts, etioplasts are also pigmented; however the pigment composition is different, with etioplasts being a yellowish colour due to the absence of Chl, instead they contain a relatively small amount of Pchlide and a significant amount of carotenoids. Etioplasts comprise an internal membrane system, named the prolamellar body (PLB); this is a paracrystalline array, from which the prothylakoids (PTs) eventually develop, and is known to be partly composed of POR in complex with its substrates (Gunning, 2001).

There is controversy over whether this etioplast stage is a relevant stage in biology; the development of etioplasts occurs only after a number of days of growth in the dark, which would not be the case in usual plant conditions. There is evidence that this is indeed a required control mechanism, with plant buds exhibiting this developmental stage while closed, before opening to the light and enabling chloroplast formation. It is likely that this stage enables the accumulation of lipids required for the mature thylakoids, as lipid biosynthesis is not suppressed. The accumulation of POR and Pchlide complexes allows the production of chlorophyll rapidly in the presence of light, enabling the photosynthetic apparatus to be formed (Solymosi and Schoefs, 2010).

Two isoforms of POR in plants, PORA and PORB, are known to be involved in the formation of PLBs; both are expressed only in the dark, while PORA is rapidly degraded upon exposure to light. PORA and PORB form light-harvesting POR Pchlide (LHPP) complexes, with each isoform present in different amounts and bound a different substrate variant; five PORA proteins, each bound to a Pchlide *b* molecule, to one PORB protein bound to Pchlide *a*. Deletion of the loop region in PORA and PORB (detailed 1.9.3.3), was shown to disrupt the formation of LHPP complexes, suggesting this region has a key role in the formation of these oligomers (Reinbothe et al., 1999). These LHPP complexes form part of the structure of the PLBs, with various lipids also involved in the overall PLB structure (Gunning, 2001).

It has also been observed that Pchlide *b* is not turned over upon irradiation, despite its photoactivation; it instead appears to have a photoprotective role, transferring light energy to non-activated Pchlide *a* molecules in addition to functioning as a light scavenger, in order to prevent the production of reactive oxygen species (Reinbothe et al., 2003a). The turnover of Pchlide *a* upon irradiation subsequently leads to the breakdown of the PLB structures, which subsequently go on to develop into the PTs of the mature chloroplast (Reinbothe et al.,

1999; Grzyb et al., 2013). This process is a clear regulatory step in chlorophyll biosynthesis in plants, preventing the accumulation of chlorophyll when it is not required, yet enabling its rapid production by an efficient light harvesting structure.

The structure of PLBs has been thoroughly researched, and they appear to be quite widespread in their structure; though the most common structures appear to be zincblende and wurtzite crystal lattices (Ikeda, 1968; Gunning, 2001). These structures are composed of tubules of lipid bilayer in tetrahedral units, assembled into large paracrystalline arrays; these structures enable maximum exposure of membrane on the stromal side, the surface where POR is located (Kovacevic et al., 2007). More open structures have been shown to be composed of a combination of pentagonal and hexagonal tubule arrangements (Gunning, 2001). Recently, AFM has enabled these structures to be imaged under native conditions; though at a low resolution, the PLB structures could be observed and seen to dissipate upon irradiation (Grzyb et al., 2013).

1.9.4 SDR Superfamily of enzymes

POR is a member of the SDR (short-chain dehydrogenases/reductases) superfamily, a large superfamily of proteins which has been extended by the reclassification of previously established families of dehydrogenases/reductases; a large number of these gene products are found within almost every sequenced genome (Jornvall et al., 1999). There have been at least 140 distinct enzymes known to be classified in the SDR superfamily, though with over 20,000 sequence depositions represented, there is likely to be a significant amount more. Structures for well over 200 SDR enzymes have been deposited to the PDB, from each subsuperfamily: classical, extended, intermediate, divergent and complex. A schematic of POR secondary structure, based on the features of the SDR superfamily, is shown in Fig. 1.16. The majority of substrate functional groups converted by SDR enzymes are hydroxyl and carbonyl groups, though others catalyse reduction of C=C and C=N double bonds (Kavanagh et al., 2008).

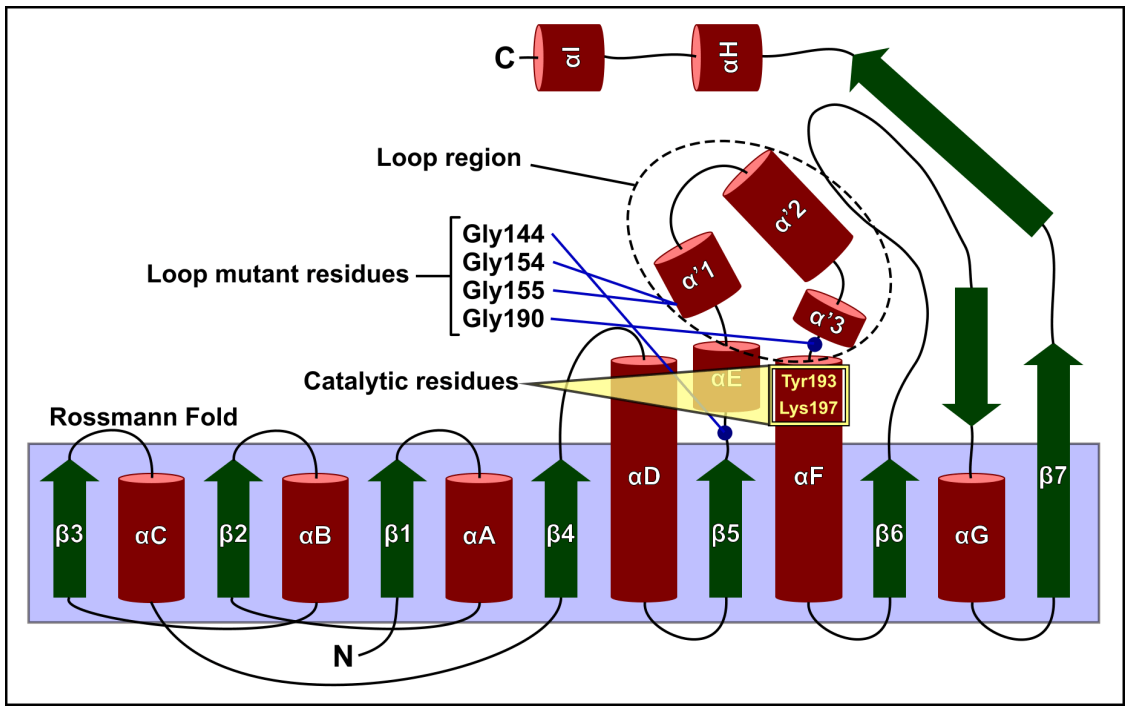

Figure 1.16 Schematic domain diagram of POR, indicating the key structural features and residues. The α-helices are shown as red cylinders, β-strands as green arrows and loops as black lines. The catalytic tyrosine and lysine residues are highlighted in yellow. The Rossmann fold is indicated by a blue box and is composed of 7 β-strands forming a sheet with 3 α-helices either side. The unique loop of POR is highlighted by a dotted ring; this contains the glycine residues analysed in chapter 4, with their positions in the loop indicated by blue lines. The D and F helices are typically the dimerisation helices in the SDR superfamily.

1.9.4.1 Rossmann Fold

The only completely unifying feature of the SDR superfamily is the presence of the Rossmann fold, a binding scaffold for NAD(P) nucleotide cofactors; this region is highly conserved throughout the superfamily. The fold is composed of a parallel β-sheet, in a twisted conformation, flanked by α-helices; approximately 3-4 α-helices are found on either side of the β-sheet, which is composed of 6-7 β-sheets in a 3-2-1-4-5-6-7 topology. There is a longer loop found between strands 3 and 4, which contains a Gly-rich motif; this is the region where the pyrophosphate of the cofactor is accommodated. The presence of an acidic or basic residue, 20 residues C-terminally of the Gly-rich sequence, determines specificity for NAD(H) or NADP(H), respectively (Kavanagh et al., 2008).

1.9.4.2 Catalysis

The SDR reactions proceed with the cofactor binding first and being released last, termed a 'bi-bi' mechanism (Sahni-Arya et al., 2007). It is the 4-*pro S* hydride which is transferred in SDR catalysis, in contrast to the 4-pro R hydride in medium chain dehydrogenases/reductases (MDRs); this mechanism is facilitated by the binding of the cofactor in an extended

conformation (Kavanagh et al., 2008). Most SDRs also exhibit a catalytic centre containing a tyrosine residue, which acts as an acid/base catalyst; the presence of an adjacent lysine acts to lower the pK_a of the tyrosine. A highly conserved asparagine residue produces a characteristic kink within helix αD; additionally its backbone carbonyl group coordinates a water molecule to a close proximity with the aforementioned lysine, forming a proton relay from bulk solvent to the catalytic tyrosine (Koumanov et al., 2003).

1.9.4.3 Oligomerisation

Most proteins within the SDR superfamily form oligomeric complexes, mainly homodimers or homotetramers (Kavanagh et al., 2008). The dimerisation face for many of these structures involves the association of hydrophobic residues along the exposed faces of helices αD and αF, forming a four-helix bundle at the homodimeric interface. Monomeric SDRs such as carbonyl reductase (CBR), however, contain an insert of approximately 20 residues prior to the catalytic tyrosine; this insert forms an α-helical segment which is positioned across the hydrophobic face, preventing the homodimeric association at this interface (Ghosh et al., 2001); a similar insert is also present in POR, suggesting that this structural feature could be involved in the regulation of POR oligomerisation.

1.9.5 POR homology model

A homology model of *Synechocystis* POR has been proposed (Fig. 1.17), based on its homology to the SDR superfamily of proteins and tryptophan fluorescence experiments (Townley et al., 2001). This is one of only two structural models of POR to act as a basis for the interpretation of various experiments pertaining to POR catalytic function (Heyes and Hunter, 2005), the other a model of the photoprotective PORA protein from barley (Buhr et al., 2008). Three proteins from the SDR superfamily, with structures in the PDB, were identified by the authors as suitable for POR homology modelling: the short-chain alcohol dehydrogenase from *Mus musculus* (1cyd); the 3α, 20β-hydroxysteroid dehydrogenase from *Streptomyces hydrogenans* (2hsd); and the 7α-hydroxysteroid dehydrogenase from *E. coli* (1ahi), which was chosen as the final structural template for POR due to the similar shapes of their substrates (Townley et al., 2001).

Modelling of the N-terminal Rossmann fold region, the binding site for the NADPH cofactor, was relatively straightforward due to the high homology of this region within the SDR superfamily; the position of the NADPH cofactor could also be confidently docked, making

protein contacts at the well-conserved GxxxGxG NADPH-binding motif. The 33-residue insert of POR (detailed 1.9.3.3) is modelled in a structure described by the authors as a helix-turn-strand, positioned over the Pchlide binding site, preventing solvent exposure of the pigment; the authors, however, do highlight the evidence to suggest the involvement of this loop in intermolecular interactions. Other work, which suggested that POR is able to function as a monomer, led them to assume a monomeric model of POR; thus the structure is built around the minimisation of both external hydrophobic residues and exposed Pchlide (Townley et al., 2001). This is despite the clear evidence to support the formation of multimeric complexes of POR upon substrate binding, likely to be driven largely by hydrophobic interactions (Reinbothe et al., 1999, 2003a).

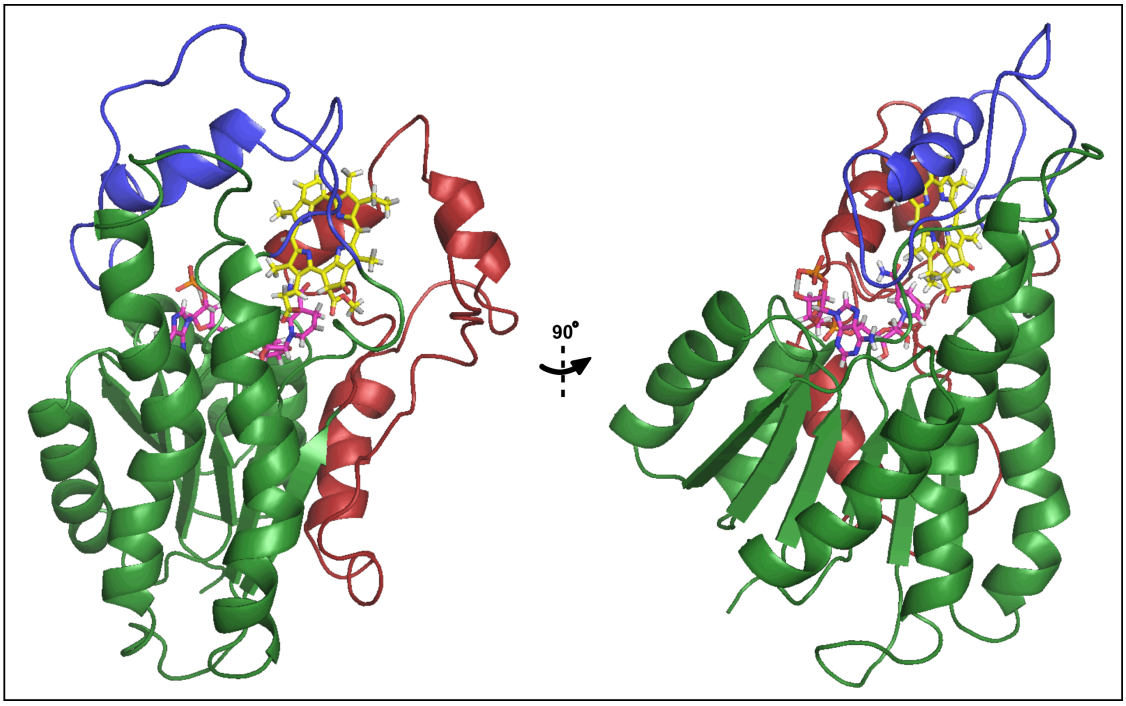

Figure 1.17 Homology model of POR proposed by Townley *et al.* N-terminus is coloured green, C-terminus red, and insert, with 'helix-turn-strand' structure, blue. The POR insert is modelled as a helix-turn-strand, positioned over the Pchlide molecule, enclosing it within the structure (Townley et al., 2001).

1.9.6 Biochemistry of POR

1.9.6.1 Substrate binding

Pchlide has been shown to be unable to bind to the 'free' POR enzyme, with NADPH cofactor required to be bound to POR in order to facilitate a conformation that is favourable for pigment binding (Heyes et al., 2008). Binding of NADPH to POR has been shown to involve at least three kinetic events, indicating a number of conformational changes involved in this binding event: the first kinetic phase has a large anisotropic signal and is inhibited by ADP,

suggesting this involves binding of the ADP moiety within the active site; the second phase also involves a significant anisotropic effect, and is likely to involve the positioning of the nicotinamide portion of the NADPH molecule; the final step is suggested to involve optimal alignment of the cofactor for hydride transfer (Heyes et al., 2008). The subsequent binding of the Pchlide substrate is similarly intricate, with a relatively fast first phase involving formation of an initial complex, with slower conformational changes following this in order to form the final, photoactive complex; these slower conformational rearrangements, responsible for the red-shift in the Pchlide absorbance maximum, appear to be the rate-limiting step in the POR reaction (Heyes et al., 2008).

1.9.6.2 Hydride transfer

The reduction of the C17-C18 bond of Pchlide by the light-dependent POR, to form Chlide, requires the addition of 2 protons and 2 electrons to the pigment (Fig. 1.18). POR requires NADPH for its reaction and has been identified in complex with both Pchlide and NADPH in membranes of etioplasts grown in the dark (Griffiths, 1978). Formation of the excited state of Pchlide, after photon absorption, facilitates the transfer of a hydride from NADPH to the substrate. NMR studies using NADPH isomers, labelled with ^{3}H at either the 4*R* or 4*S* position, indicated that the hydride is transferred from the *pro-S* face of the nicotinamide ring to the C17 position of the porphyrin ring (Valera et al., 1987; Begley and Young, 1989).

1.9.6.3 Proton transfer

There are two well conserved residues in POR, a tyrosine and a lysine residue, which are thought to be involved in the proton transfer reaction; mutation of either residue reduces the activity of the enzyme to negligible rates. It has been suggested that formation of the hydride transfer intermediate facilitates the transfer of a proton from the phenol of the conserved tyrosine to the C18 position of the porphyrin; it is expected that the proximity of the conserved lysine residues acts to lower the effective isoelectric point (pK_a) of the tyrosine to allow proton transfer to occur (Wilks and Timko, 1995).

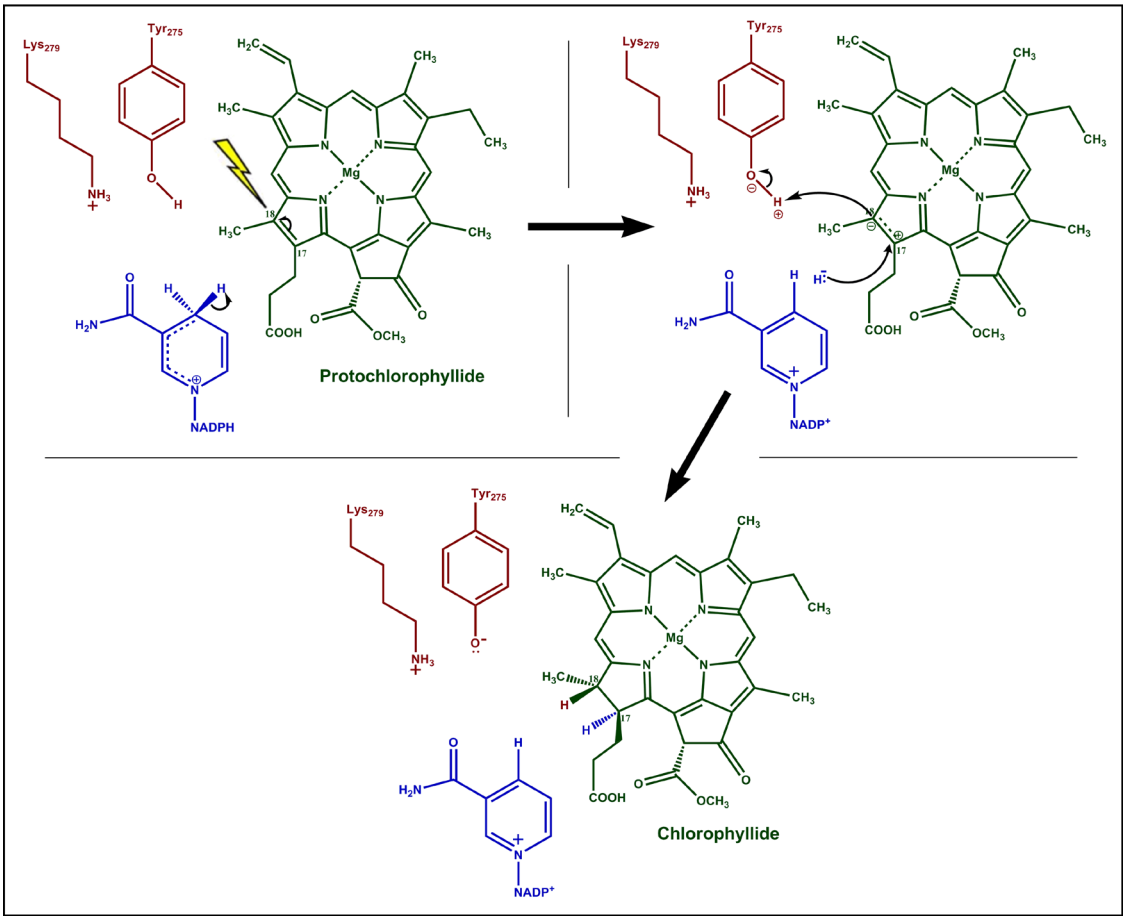

Figure 1.18 Proposed reaction mechanism for POR. The pigments, cofactors and enzyme residues are coloured green, blue and red, respectively. Curly arrows denote movement of electrons, with full arrow heads indicating movement of an electron pair and half arrow heads indicating direction of electron movement to give partial charge character. Irradiation of the molecule leads to the production of an excited state intermediate and a charge separation across the C17-C18 double bond, facilitating the transfer of the *pro-S* hydride of NADPH to the C17 position of Pchlide. This enables the subsequent transfer of a proton from the catalytic tyrosine to the C18 position of Pchlide, to form Chlide (Valera et al., 1987; Begley and Young, 1989; Wilks and Timko, 1995).

Recent work has indicated that, although the activity of POR is greatly reduced, proton donation can still occur in the absence of these residues. This work also exhibited alternative functions for the tyrosine and lysine residues; both appeared to be involved in Pchlide binding in some form, as well as a possible role in stabilisation of the hydride intermediate (Menon et al., 2009). Lysine mutations, in particular, had a significant effect upon Pchlide binding; in conjunction to other findings, indicating that porphyrin variants with alternative C17 side chains were not accepted as POR substrates (Klement et al., 1999), this suggests a possible role of lysine for coordination of the C17 carboxyl group (Lebedev et al., 2001).

1.9.6.5 Excited state processes

The light-dependency of POR makes it an ideal model enzyme with which to study high speed reaction dynamics; substrates can be bound in the dark and the reaction followed from the

exact moment of illumination. This has enabled intermediates in the reaction to be monitored, by both fluorescence and absorbance, using low temperature spectroscopic techniques (Heyes and Hunter, 2004). It was identified that, even at temperatures around 120 K, at which protein motions should not occur, Pchlide was converted to an intermediate. It was determined that this was the light-dependent step and occurred without any significant protein side-chain motions; the remainder of the reaction is composed of so-called 'dark' steps and could be observed as the temperature was increased (Heyes et al., 2002, 2003; Heyes and Hunter, 2004).

The initial light-dependent photochemical step leads to the conversion of Pchlide to a non-fluorescent intermediate. This light step was initially characterised as the conversion of bound Pchlide, with absorbance at 642 nm and fluorescence at 644 nm, to a non-fluorescent intermediate, with absorbance at 696 nm, denoted A_{696}. Formation of this intermediate occurred only with NADPH bound, not $NADP^+$, and the change of oxidation of NADPH has been followed, indicating the conversion to $NADP^+$ in this step (Heyes et al., 2006); this, along with the non-fluorescent nature of the intermediate, initially suggested this to be an ion-radical complex (Heyes et al., 2002); though the presence of a functional radical species has since been questioned, with the A_{696} intermediate considered a non-radical species with considerable charge transfer character (Heyes et al., 2006).

Further investigation into the light-driven step has led to the proposal of an I675* intermediate, named due to its absorbance at 675 nm; it has been suggested that this may be an activated Pchlide form, initiated by a first photon, with catalysis proceeding via a second photon (Sytina et al., 2008). Kinetic isotope effects indicate a proton contribution to this intermediate, whether by multiple hydrogen bonding to the pigment or by transfer of a proton; it has been speculated that a proton could be transferred to the C18 position, or possibly to the C13 position to form an enolate intermediate (Sytina et al., 2012).

1.9.6.4 Product release

More light has been shed on the 'dark' steps of the reaction, with a thorough catalytic cycle proposed based on low temperature spectroscopic experiments, using both mesophilic and thermophilic POR variants (Heyes and Hunter, 2004) (Fig. 1.19); the substrate binding and release steps were also studied at higher temperatures, using the thermophilic *T. elongatus* POR isoform, confirming this catalytic cycle at physiological temperatures (Heyes et al., 2007). These experiments monitored each 'dark' step subsequent to the non-fluorescent A_{696} intermediate, with absorbance and fluorescence monitored as the temperature was gradually

increased, to allow further steps to proceed. There were four 'dark' steps identified, which involved the following processes in this order: the conversion of A_{696} to Chlide; the release of $NADP^+$; the binding of NADPH; and finally the release of the Chlide product. It is unknown whether this is a functional, cofactor-mediated product release mechanism; it is conceivable that this mechanism could ensure retention of the Chlide molecule until its delivery to the next enzyme on the pathway.

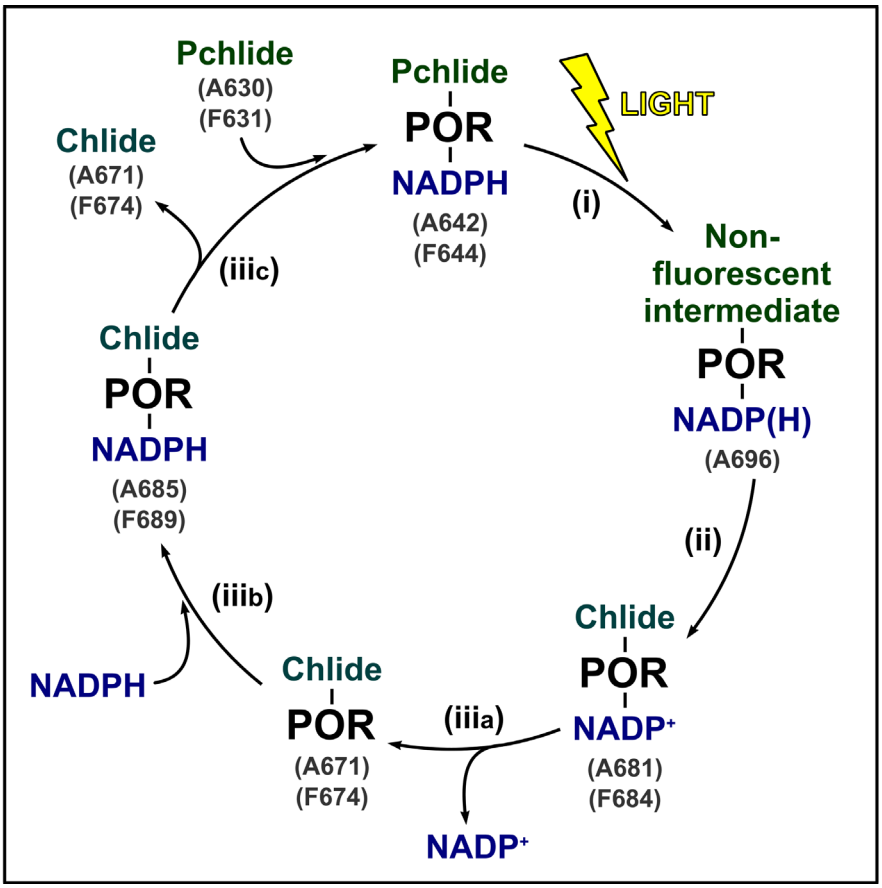

Figure 1.19 POR catalytic cycle, indicating the absorbance and fluorescence maxima of each intermediate. (i) Initially, Pchlide, in complex with POR and NADPH, is converted by light to a non-fluorescent intermediate. (ii) The 'dark' reactions lead to the product formation and oxidation of NADPH. (iii$_a$) $NADP^+$ is then released and (iii$_b$) NADPH subsequently binds. (iii$_c$) This facilitates the release of Chlide, with Pchlide now able to rebind POR. Figure adapted from (Heyes and Hunter, 2004).

1.9.6.6 Role of conserved cysteine residues

There are 3 cysteine residues that are completely conserved throughout all POR proteins, and are thought to be located in, or near, the active site (Townley et al., 2001); these are Cys-37, Cys-199 and Cys-226 in the *T. elongatus* POR variant. The use of thiol modification reagents has indicated the importance of Cys residues in the binding of substrates and catalysis (Heyes et al., 2000). Mutations of Cys-199 and Cys-226 equivalents in PORA and PORB of barley arrested Chlide production in addition to causing accumulation of singlet oxygen; this

suggests that the cysteine residues are important in the correct coordination of Pchlide to enable energy transfer throughout POR proteins (Buhr et al., 2008). Mutation of Cys-37 was found to have a weakening effect upon NADPH binding, however it is not clear whether this is due to a direct interaction or a change in orientation of the neighbouring Arg-38 residue, proposed to coordinate the 2'-ribose-phosphate (Menon et al., 2010).

Spectroscopic analysis of Cys-226 mutants indicated that this residue is important for enzyme catalysis; it was observed that C226S variants exhibited a 14-fold reduction in enzyme activity, while C37S and C199S had negligible effects upon activity. The binding of Pchlide was also weakened 4-fold in the C226S variants, suggesting involvement of this cysteine residue in substrate binding. Additionally, this mutant has an effect on the solvent isotope effect (SIE) but not on the kinetic isotope effect (KIE), suggesting that the proton transfer step is limiting in this case (Menon et al., 2010); further evidence from low temperature spectroscopy experiments indicate that the proton transfer step in the C226S variant can occur without the same levels of solvent contribution as observed for the WT enzyme.

1.9.6.7 Quantum tunnelling

Work by Heyes et al. has indicated that both the hydride and proton transfer steps, in the POR reaction, occur via nuclear quantum tunnelling (Heyes et al., 2009, 2011); this was inferred from the various temperature effects upon both proton and hydride transfer rates, and the values for the SIE and KIE. The perception is that the electronic configuration upon absorption of a photon leads to a Franck-Condon transition that can be considered a 'tunnelling-ready configuration'; hydrogen tunnelling from this state enables the remainder of the reaction, from the excited Pchlide state onwards, to proceed via an endothermic reaction (Heyes et al., 2009).

1.9.7 Comparing LPOR and DPOR

Despite the apparently distinct structures between LPOR and DPOR, there are a number of details pointing to similarity in the way that their structure relates to their function. Cyanobacteria contain both LPOR and DPOR and thus are ideal model organisms to study how these two variants of POR coexist, and whether they share any physiological attributes. It has been found that knockouts of LPOR have a greater sensitivity to light (Fujita et al., 1998); this suggests a possible photoprotective role for LPOR towards DPOR, as observed

with PORA and PORB proteins within LHPP complexes. The prospect that LPOR and DPOR could interact has to be considered; as does a possible function of the L-protein of DPOR as an oxygen sensor, in order to regulate the production of singlet oxygen species due to photoexcitation of Pchlide.

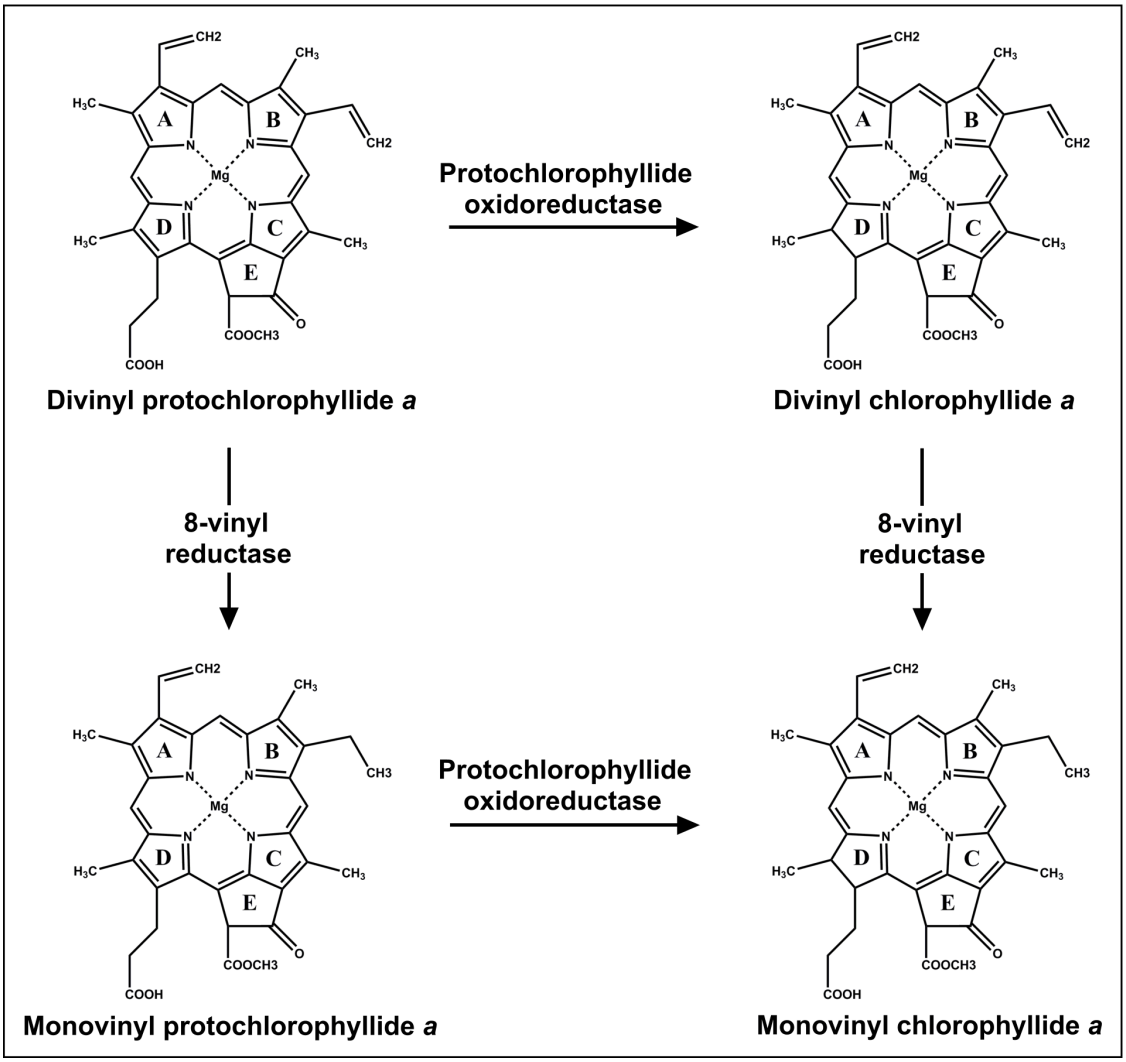

Figure 1.20 The reduction of Pchlide to Chlide and reduction of the 8-vinyl group has been proposed to occur in either order (Heyes et al., 2006). POR can function on either monovinyl (MV) or divinyl (DV) Pchlide, while 8-vinyl reductase has been observed to function on either DV Pchlide or DV Chlide in various organisms.

1.10 Reduction of the C8-vinyl group

Most (bacterio)chlorophylls have an ethyl group at the C8 position of the tetrapyrrole ring, produced by the reduction of the C8-vinyl group by the a C8-vinyl reductase (8VR) (Fig. 1.20). There are two unrelated classes of 8VR known to exist, BciA, from plants (Nagata et al., 2005) and purple bacteria, for example *R. sphaeroides* (Canniffe et al., 2013), and BciB, from cyanobacteria such as *Synechocystis* sp. (Ito et al., 2008; Islam et al., 2008). In the original analysis of tetrapyrrole biosynthesis by Granick, the pathway was assumed to be linear, with

8VR being placed between cyclase and POR, converting DV Pchlide to MV Pchlide. Subsequent studies indicated that the 8VR enzyme acts on multiple substrates within the tetrapyrrole biosynthesis pathway, therefore this enzyme cannot be described as part of a linear pathway (Rebeiz et al., 2007; Nagata et al., 2007). Both BciA and BciB have now been shown to act preferentially on DV Chlide, in both *R. sphaeroides* and *Synechocystis* sp. PCC6803 (Canniffe et al., 2014).

1.11 Addition of the phytol tail

The final step of chlorophyll synthesis involves the attachment of the phytol tail to the propionate chain of ring D of the Chlide porphyrin (Beale, 1999). This step is catalysed by the enzyme chlorophyll synthase (Fig. 1.21) and has been observed to function on different variations of substrates, depending on the organism; both MV and DV Chlide *a* can be utilised in cucumber, while in *Synechocystis* both Chlide *a* and *b*, but not bacteriochlorophyllide, can be turned over (Oster et al., 1997; Adra and Rebeiz, 1998). The phytol tail, once attached, composes around 30 % of the chlorophyll molecule and significantly increases the hydrophobicity of the pigment (Scheer, 1991).

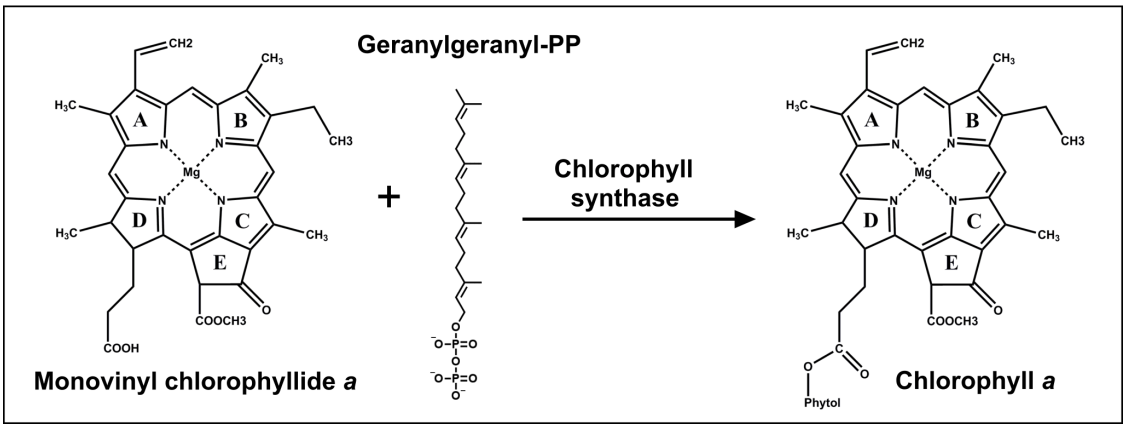

Figure 1.21 Addition of the phytyl tail to Chlide to form mature chlorophyll. The reaction involves the precursor Geranylgeranyl disphosphate, with the geranylgeranyl chain undergoing subsequent reductions to form the phytol chain.

The chlorophyll synthase reaction requires both NADPH and ATP, and is composed of two stages: addition of a geranylgeraniol moiety via a geranylgeranyl diphosphate precursor; this is followed by three systematic reductions of three of the four C=C double bonds of the alcohol chain (Schoch, 1978; Benz et al., 1980). It is not known whether one or two enzymes are involved in this reaction; two genes have been found to be essential: *chl/bch P* and *chl/bch G*. Mutation of *bch P* has been shown to produce chlorophyll with an attached

geranylgeraniol moiety (Addlesee et al., 1996; Addlesee and Hunter, 1999); this has lead to suggestions that the gene products of *chl/bch P* and *chl/bch G* are the reductase and synthase, respectively (Oster et al., 1997; Addlesee et al., 2000). An alternative process exists in spinach, where the enzyme geranylgeranyl-PP reductase converts geranylgeranyl-PP to phytyl-PP, which is subsequently attached to the chlorophyll molecule (Soll et al., 1983; Rüdiger, 1987).

1.12 NMR of large proteins

NMR is an extremely valuable tool for studying structures of monomeric proteins; despite X-ray crystallography being more suited to high resolution structure determination, NMR studies, used concomitantly or alternatively, can give greater insight into the organisation of mobile and unstructured regions. Historically, NMR has been limited to smaller proteins (up to 25 kDa) for two reasons: loss of sensitivity, due to the increased relaxation rate, and spectral complexity. These problems with NMR studies of larger proteins are beginning to be addressed by a number of new experimental techniques, which will be discussed below.

1.12.1 Isotopic labelling

The use of labelled isotopes of various atoms has long been used in protein NMR in order to observe signals throughout proteins; ^{13}C and ^{15}N labelled protein can be used to observe these atoms in a number of different experiments, with the associated chemical shifts exhibiting further information about the atomic environment than proton shifts alone. Deuteration of proteins, the replacement of protons (^{1}H) with deuterons (^{2}H), yields an extension of relaxation times, in turn causing a narrowing in linewidths of the proton resonances; the relaxation rate of ^{13}C atoms is also reduced as ^{1}H-^{13}C dipole interactions are the dominant relaxation mechanism for carbon atoms in the protein (LeMaster, 1990). The application of ^{2}H, ^{13}C and ^{15}N labelling together, along with the reprotonation of backbone amide groups, enables sequential NMR assignment of residues in much larger proteins; the reduction of relaxation rates allows the use of 3-dimensional, or even 4-dimensional, experiments that are required for sequential assignment along the protein backbone.

1.12.2 TROSY experiments

The use of TROSY (transverse relaxation optimised spectroscopy) techniques, when combined with isotopic labelling, combats the problems with transverse relaxation; this has enabled the study of much larger proteins by NMR. These experiments suppress transverse relaxation between ^{15}N and ^{1}H atoms; this is achieved by interference between the dipole-dipole interaction and chemical shift anisotropy (CSA). Sensitivity is therefore improved during ^{15}N evolution periods, due to greater nitrogen relaxation times, as well as during detection on the proton, due to increased proton relaxation time (Salzmann et al., 1999). TROSY techniques can therefore be implemented on both protonated and deuterated samples. There are now a number of different experiments that have been adapted to TROSY techniques, enabling the application of a number of techniques to larger proteins (Frueh et al., 2013).

1.12.3 Structural restraints

Though many of the problems with relaxation and signal intensities can be overcome by deuteration of proteins, this causes problems when it comes to identifying structural restraints, required to determine accurate protein structures. Deuteration of all carbon bound protons in the structure prevents observation of side chain signals; these are key signals in NOESY experiments, which indicate distances of atoms through-space, allowing structures to be built. There are, however, a number of techniques that can be applied in conjunction with deuteration, in order to obtain structural restraint data for large proteins.

Residual dipolar coupling (RDC) experiments involve the use of partially aligned samples; these measurements allow the determination of bond orientations and thus determination of backbone torsion angles (Bhattacharya et al., 2010). A more direct method of introducing structural restraints is selective methyl labelling, involving the selective proton labelling of various methyl groups in the protein, via the introduction of specific precursors (Tugarinov et al., 2003); therefore NOESY experiments can be carried out to calculate distances between methyl groups for structure determination. Paramagnetic relaxation enhancement (PRE) involves the introduction of one or more paramagnetic labels onto a specific part of the protein; the label has a broadening effect upon local signals, dependent on distance, and can therefore be used to determine distances of various residues to that location, providing an alternative structural restraint (Frueh et al., 2013).

1.12.4 Non-uniform sampling (NUS)

Acquisition of NMR data in three, four or even five dimensions requires the collection of massive amounts of NMR data, due to the exponential increase in data points required for each dimension; this can lead to extremely long acquisition times, particularly for large proteins where more increments are required for optimal resolution. A technique called non-uniform sampling (NUS) can be applied to these experiments; this involves the collection of a much smaller number of data points in the indirect dimensions, therefore dramatically reducing acquisition times from days to hours (Barna et al., 1987). This makes these multidimensional experiments much more feasible for large proteins, which are likely to have a limited stability at acquisition temperatures.

Chapter 2

2. Materials and Methods

2.1 Materials

All chemicals were purchased from Sigma or Fisher, unless stated otherwise. SDS and Native PAGE buffers were acquired from Life Technologies. Chromatographic resins were obtained from GE Healthcare Life Sciences.

2.2 Standard buffers, reagents and media

All growth media and buffers were prepared as described in (Sambrook et al., 1989) and sterilised by autoclaving (at 126 °C), unless stated otherwise. All media and buffers were prepared with distilled water, further purified by a Milli-Q system. All water soluble antibiotics and small reagent volumes were filter sterilised using a 0.2 µM syringe filter.

2.3 Bacterial strains and plasmids

Bacterial strains and plasmids used are detailed in tables 2.1 and 2.2, respectively. *E. coli* was grown on Luria-Bertani (LB) growth medium (Sambrook et al., 1989) with suitable antibiotics: kanamycin (detailed A2.1); chloramphenicol (detailed A2.2). *R. sphaeroides* was grown on M22 media (detailed A1.9).

All labelled *E. coli* growths were grown in M9 media (detailed A1.3) containing ^{15}N labelled ammonium sulphate and ^{13}C labelled glucose, where appropriate, for labelling schemes. For deuterated growths, deuterium oxide (D_2O) was used for all buffer components, rather than Milli-Q water, unless stated otherwise. Deuterated growths also required fully deuterated glucose as a carbon source. Antibiotics were added at the concentrations detailed above for LB media.

2.4 Production of competent *E. coli* cells

E. coli cells were plated out onto LB agar containing the appropriate antibiotic selection. A few colonies were taken and used to inoculate 10 ml of LB media. This was then incubated at 37 °C for 5 hours, with agitation, and placed on ice for 30 min before being centrifuged at

3,500 rpm for 10 min to pellet the cells. The cells were then resuspended in 5 ml of 0.1 M $CaCl_2$ and left on ice for 20 min before being centrifuged at 3,500 rpm for 10 min at 4 °C. The resulting cell pellets were then resuspended in 2 ml of ice cold 0.1 M $CaCl_2$ and left on ice for 20 min before being pelleted as before. Finally these cells were resuspended in 500 µl of ice cold 0.1 M $CaCl_2$, 15 % glycerol solution before aliquoting into 50 µl samples and storing at -80 °C.

2.5 Transformation of *E. coli* cells

A 50 µl aliquot of competent cells, prepared as above, was thawed on ice. 1 µl of POR plasmid DNA was added to the cells and the sample incubated on ice for 15 min. The sample was then heat shocked (42 °C) for 50 s and subsequently incubated on ice for a further 2 min. 200 µl of LB media was then added and the sample incubated at 37 °C with agitation for 60 min. 200 µl of the sample was plated out onto an LB agar plate with the appropriate antibiotic selection. Plates were incubated overnight at 37 °C and then stored at 4 °C.

2.6 Over-expression of pET9His T.POR in *E. coli* without induction

2.6.1 Growth of starter cultures

A single transformed colony was picked from the plate using a sterile loop, and used to inoculate 50 ml of LB media in a 250 ml conical flask, with the appropriate antibiotic selection. This culture was incubated overnight at 37 °C with agitation.

2.6.2 Large Scale Culture Growths

10 ml of starter culture (above) was used to inoculate 500 ml of LB media in a 2 L flask, with the appropriate antibiotic selection. This culture was incubated overnight at 30 °C with limited agitation.

2.7 Over-expression of pET9His T.POR in *E. coli* by IPTG induction

2.7.1 Growth of starter cultures

A single transformed colony was picked from the plate using a sterile loop, and used to inoculate 50 ml of desired media in a 250 ml conical flask, with the appropriate antibiotic selection. This culture was incubated overnight at 37 °C with agitation.

2.7.2 Large Scale Culture Growths

10 ml of starter culture (above) was used to inoculate 500 ml of desired media in a 2 L flask, with the appropriate antibiotic selection. This culture was incubated at 37 °C with agitation.

2.7.3 Measuring the Growth of *E. Coli*

1 ml of *E. Coli* growth was removed from the flask under sterile conditions and transferred to a 1.5 ml cuvette. This was placed in a Cary 50 spectrophotometer and the absorbance at 600 nm wavelength measured, using sterile media as a blank. The cell growth samples were diluted as required to ensure that an accurate reading, ie. 0.6 absorbance units or below, could be measured.

2.7.4 Protein Induction

Once the cell cultures had reached an OD^{600} of 1.2, IPTG (Isopropyl β-D-1-thiogalactopyranoside) was added to a final concentration of 0.4 mM (detailed A2.4). The cell cultures were then incubated at 20 °C with limited agitation. The cultures were induced for: 3 hours for LB media, overnight for ^{15}N-labelled M9 media, and for 48 hours for deuterated M9 media.

2.8 Harvesting *E. coli* cells

Cell cultures were divided amongst 1 L plastic Nalgene centrifuge flasks and centrifuged at 4,000 rpm in a Beckmann centrifuge at 4 °C for 25 min. The cell pellet was resuspended in minimal Low Salt Buffer (detailed A3.10), transferred to sterile universal tubes and stored at -20 °C.

2.9 Fractionation of *E. coli* cells

Cell pellets were defrosted and an equal volume of Low Salt Buffer added. This was placed into a round bottomed glass sonicating vessel and sonicated on ice for 24 cycles (15 s sonication, 30 s rest). The resulting broken cell solution was heated at 42 °C for 12 min. This was then transferred into 50 ml Beckmann centrifuge tubes and centrifuged at 18,500 rpm for 25 min at 4 °C.

2.10 Protein purification

2.10.1 Ammonium Sulphate precipitation

The supernatant from the centrifugation step in the fractionation method was taken, and Ammonium Sulphate ($NH_4[SO_4]_2$) added stepwise at a weight of 52.3 g per 100 ml solution, to give an 80 % saturated solution. This was left to mix for 1 hour at room temperature and then centrifuged at 18,500 rpm for 25 min at 4 °C, with the resultant pellet stored at -20 °C until required.

2.10.2 Purification of His-tagged POR on a Nickel column

A 10 ml column was packed with 2.5 g Protino Ni-TED resin and equilibrated with 25 ml LEW Binding Buffer (detailed A3.1). The protein pellet from the Ammonium sulphate precipitation was defrosted and resuspended in LEW Binding Buffer. This was then passed through 0.5 µm and 0.2 µm syringe filters before being loaded onto the column. The column was then washed with first 50 ml LEW Binding Buffer and then 25 ml LEW Wash Buffer (detailed A3.2), before the protein was eluted with LEW Elution Buffer (detailed A3.3). The eluted protein was collected in 2 ml fractions, with each fraction checked for protein content using BioRad reagent. The fractions with high protein concentration were pooled and 1 mM DTT and Roche complete protease inhibitor added.

2.10.3 Purification of His-tagged POR on an SP Sepharose column

Eluted protein from the Nickel column (above) was concentrated to 2.5 ml using a 20 ml Vivaspin concentrator and made up to 10.5 ml with No Salt SP buffer. A glass chromatography column was packed with 30 ml SP sepharose resin (GE Healthcare), attached to an ÄKTA Prime with Low Salt SP Buffer (detailed A3.4) as buffer A and High Salt SP Buffer

(detailed A3.5) as buffer B. The column was equilibrated with 2 column volumes of 90 % buffer A and 10 % buffer B. The protein was injected via a 10 ml loop and the column washed with a further 2 column volumes of 10 % buffer B; during this time 10 ml fractions were collected. A gradient of 100 ml was set up to take the buffer concentration from 10 % buffer B to 45 % buffer B, during which time 5 ml fractions were collected. The column was then washed with 2 column volumes of 100 % buffer B, during which time 10 ml fractions were collected. Fractions containing pure POR were pooled and the protein concentrated to a smaller volume, using a 20 ml Vivaspin concentrator.

2.11 Estimating the Concentration of His_6-POR

A POR sample was diluted in Milli-Q water to a few different concentrations, roughly estimated to be 1 - 10 μg ml^{-1}, to give 800 μl diluted protein samples. 200 μl of BioRad Protein Assay Reagent was added, to these as well as a pure Milli-Q sample (as a blank), the samples vortexed, and incubated at room temperature for 15 min. An absorbance reading at 595 nm was recorded for each sample and, for readings in the 1 - 10 μg ml^{-1} range, the corresponding concentration deduced from a BSA standard concentration curve and adjusted for dilution factors.

2.12 Site-directed mutagenesis

Mutagenesis was carried out on the pET9AHis-POR plasmid, using the QuikChange Site-Directed Mutagenesis Kit from Agilent, as per the manufacturers instructions. Primers were ordered from Sigma, with those used in this study detailed in Table 2.3. Initial G190A, G154/155A and G154/155/190A mutants, along with all mutants for EPR, were prepared by Dr Amanda Brindley, in conjunction with project students.

2.13 Small-scale preparation of plasmid DNA (mini-prep)

A 5 ml culture of *E. coli* was prepared as in 2.6.1. The cells were pelleted at 4,000 rpm at 4 °C for 20 min in a benchtop centrifuge. These cells were treated using a QIAGEN mini-prep kit as per the instructions to give a preparation of plasmid DNA in Milli-Q water. The plasmid DNA was stored at -20 °C.

2.14 DNA sequencing

15 µl of plasmid DNA was placed in an eppendorf and sent to GATC Biotech for sequencing. Both forward and reverse data was obtained, by sequencing from the T7 promoter and terminator, respectively. The sequence data was analysed using the CodonCode Aligner software to ensure plasmid quality.

2.15 Pigment preparation from *R. sphaeroides* Δ*bchJ*

2.15.1 Growth of *R. sphaeroides* starter cultures

A small aliquot of *R. sphaeroides* Δ*bchJ* glycerol stock was streaked out onto M22 agar (detailed A1.6) under sterile conditions and left to incubate for a few days at 30 °C in the dark. Once the cells had grown, a single colony was picked and used to inoculate a 10 ml starter of M22 media (detailed A1.7), in a universal with vitamins added, and left to shake for 2 days at 30 °C in the dark. This was used to inoculate 80 ml of M22 media, with vitamins, in a 125 ml flask, and this left to shake for 2 days at 30 °C in the dark.

2.15.2 Large Scale Growth

2 L conical flasks were filled to 1.5 L with M22 media and sterilised by autoclaving, with vitamins added just prior to inoculation. 20 ml of the 80 ml M22 starters was added to each flask as inoculants and these were left to shake for 2 days in the dark, or until the media appeared green. The media was centrifuged at 4,000 rpm at 4 °C for 30 min, the supernatant retained and kept in the dark.

2.15.3 Pigment Extraction

All the following was carried out under green light. A solvent emulsion of 2 volumes Acetonitrile (ACN) to 1 volume of Diethyl Ether (Et_2O) was produced. This was added to the media at 2 volumes of media to 1 volume solvent plus salt, mixed thoroughly and allowed to settle into 2 phases. The green upper organic solvent phase was isolated and the lower water phase discarded. A rotary evaporator was used to evaporate off the solvent in order to give a highly concentrated sample of Pchlide, which was then aliquotted into eppendorfs, dried in a vacuum centrifuge and stored at -20 °C.

2.15.4 Pigment Purification by high performance liquid chromatography

Analytical HPLC of the extracted Pchlide was performed on an Agilent 1200 high-performance liquid chromatograph using a reverse phase C18 column. Buffer A is detailed in A3.9, with buffer B consisting of 100 % methanol; both were filtered through a 0.22 µM membrane and degassed before use. The column was equilibrated in 65 % Buffer B at a flow rate of 3.5 ml min^{-1}. Pigment was resuspended in 100 % methanol, diluted with equal parts water and 500 µl loaded onto the column using an injection loop. A program was set up to run a linear gradient from 65 % to 75 % buffer B over 35 min. Pigment elution was detected using a multichannel diode array detector (DAD) [Agilent] to record absorbance at 440 nm and 632 nm, and a fluorescence detector (Agilent) using an excitation wavelength of 440 nm and monitoring emission at 632 nm. Fractions were collected throughout this time and those containing pure Pchlide were pooled. A gradient was then run from 75 % to 100 % B over 10 min, 100 % B for 5 min and 65 % B for a final 10 min, in order to wash the column and return it to the binding buffer.

2.16 POR Assays

2.16.1 Concentration of Pchlide

An aliquot of dried Pchlide was re-suspended in 1 ml of methanol. The sample was vortexed, to ensure maximum Pchlide solubility had been achieved, before being centrifuged at 15,000 rpm for 5 min at 4 °C, to remove any insoluble Pchlide. The absorbance of Pchlide was analysed at 630 nm, in 80 % Acetone and 20 % water, using a Cary 50 spectrophotometer. The extinction coefficient of 30.4 mM^{-1} cm^{-1} was used to calculate the concentration of Pchlide present in the sample.

2.16.2 Concentration of NADPH

A 1 ml stock solution of 100 mM NADPH was made up with MilliQ water. The absorbance of NADPH was analysed at 340 nm using a Cary 50 spectrophotometer. The extinction coefficient of 6.2 mM^{-1} cm^{-1} was used to calculate the concentration of NADPH present in the sample.

2.16.3 POR Assay

The following protocol was carried out under low light conditions. A water bath set to 55 °C was attached to the sample compartment of a Cary 50 spectrophotometer. A program was set up in Scanning Kinetics in order to scan an absorbance range of 600 - 750 nm at a rate of 150 nm sec^{-1} for 2 min. A baseline spectrum was collected using assay buffer as a blank. A sample of 1 µM POR, 100 µM NADPH and 20 µM Pchlide, in a total volume of 1 ml with Assay Buffer (detailed A3.7), was incubated for 2 min. The program was then initiated and, after 3 sec, the sample was illuminated with a KL 1500 electronic fibre optic light source with a 400 - 500 nm filter attached. Both the loss of Pchlide and the formation of Chlide could be observed by the respective absorbance decrease at 630 nm and increase at 669 nm. The files were exported in CSV format and further analysis carried out in Microsoft Excel.

2.17 Formation of a POR-Pchlide-NADP(H) Ternary Complex

The following protocol was carried out under low light conditions. A 5 ml sample of 50 µM POR, 5 mM NADP(H), 100 µM Pchlide and 25 mM DTT was made up in Complex Buffer (detailed A3.8). This sample was incubated for 10 min in a 55 °C water bath and centrifuged at 3,750 rpm at 4 °C for 20 min to pellet out any precipitated aggregate. The sample was then made up to 10 ml with low salt buffer and loaded to a SP sepharose column, connected to an ÄKTA Prime and pre-equilibrated with 20 ml Low Salt Buffer. The column was washed with 20 ml Low Salt Buffer before application of a gradient, running from 100 mM to 2 M NaCl over 100 ml, with 5 ml fractions taken at equal intervals. The column was then washed with 20 ml High Salt Buffer.

All buffers were passed through the column at 1.5 ml/min and absorbance of the eluant was recorded using a UV spectrometer, with elution traces compiled by the Primeview software. Fractions which exhibited an absorbance at 280 nm and were also green in colour were collected and pooled.

2.18 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated using precast 16 % SDS PAGE (Invitrogen) in a Laemmli buffer system (Laemmli, 1970). Gels were set up as detailed by the manufacturers' instructions, with 200 ml of 1× MES buffer placed between the gels and 400 ml of 1× MES buffer in the tank surrounding the gels. 20 µl of protein sample was mixed with an equal volume of 2 x sample buffer and between 10 and 30 µl loaded onto the gel, depending on protein concentration;

alongside either pre-stained BioRad SDS broad range markers or Sigma low range molecular markers. The gels were run at 180 mV for 67 min at room temperature and protein bands were visualised using Coomassie Brilliant Blue R250.

2.19 Native Gels

Proteins were separated, based on the Blue Native Polyacrylamide Gel Electrophoresis technique, developed by Schägger (Schägger *et al.* 1991), using precast 4 - 16 % *NativePAGE* NOVEX (Invitrogen) in a Laemmli buffer system (Laemmli, 1970). Gels were set up as detailed by the manufacturers' instructions, with 200 ml of 1 × NativePAGE running buffer containing 1 × NativePAGE blue cathode buffer, placed between the gels and 400 ml of 1 × NativePAGE running buffer in the tank surrounding the gels. 20 µl of protein sample was mixed with an equal volume of 2 x sample buffer and loaded onto the gel alongside 20 µl of NativeMark unstained native protein marker (Invitrogen). The gels were run at 150 mV for 120 min at 4 °C and protein bands were visualised using Coomassie Brilliant Blue R250.

2.20 Labelling POR with MTSL

Protein samples were buffer exchanged into Low Salt Buffer (detailed A3.10), with no DTT present. MTSL label (*S*-[1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl]methyl methanesulphonothioate) was added to the samples in a 10-fold excess to protein, with the samples left to incubate overnight at 4 °C. Samples were buffer exchanged, using a PD10 column, into Low Salt Buffer, with DTT present, in order to remove any excess label.

2.21 NMR Experiments

NMR experiments were conducted using a Bruker Avance DRX 600 instrument equipped with a cryoprobe, unless stated otherwise. Pigment NMR was carried out at an acquisition temperature of 298 K, and protein NMR at 218 K. All data was processed in Felix 2007 using in-house macros.

2.22 Asstools

The chemical shift data of all backbone CO, Cα and Cβ, as well as the data for the preceding residues, for each amide were inputted to the simulated annealing program of the Asstools suite of assignment programs. Asstools works by using the protein's amino acid sequence

and attempting to reorder the spin systems into the correct sequence, based on minimising an energy function. This uses both the pairing up the chemical shift data for the selected residue and its adjacent counterpart, along with the characteristic chemical shift ranges for resonances in particular residue types (Reed *et al.* 2003). 30 iterations were conducted with the programme starting at a different randomly chosen part of the sequence each time. Pairs of spin systems are swapped and ordered according to both their inter- and intra-molecular carbon shifts until the best possible sequence is achieved; the outputted suggested sequence was then manually checked for errors.

2.23 Analytic Ultracentrifugation

Apo protein samples were prepared as detailed in 2.10. Ternary complex samples were prepared as detailed in 2.17. For product samples, NADPH was added at a 100-fold excess, and the samples illuminated as in 2.16.3. Samples were analysed by Dr David Morou-Besong, supervised by Professor Steve Harding, based at the National Centre for Macromolecular Hydrodynamics, University of Nottingham. Samples were run on a Beckmann Ultracentrifuge at 50,000 rpm, analysing the sedimentation velocity of the samples, with data analysed using the SEDFIT-MSTAR software package.

2.24 Electron Paramagnetic Relaxation (EPR)

Protein samples were prepared as detailed in 2.10. $NADP^+$ was added to the samples in 100-fold excess to protect cysteine labelling, where required. Samples were labelled with MTSL (detailed 2.20) and concentrated to a final POR concentration of 100 µM. Complex samples were formed by the addition of Pchlide to a concentration of 60 µM or 150 µM. For product samples, NADPH was added at a 100-fold excess, and the samples illuminated as in 2.16.3. All samples were run by Dr Steve Rigby from the Manchester Institute of Biotechnology at the University of Manchester, using a Bruker ELEXSYS E580 spectrometer operating at X-band (~9.5 GHz).

2.25 Electron Microscopy (EM)

Apo protein samples were prepared as detailed in 2.10. Ternary complex samples were prepared as detailed in 2.17. For product samples, NADPH was added at a 100-fold excess, and the samples illuminated as in 2.16.3. The following was carried out by Dr Pu Qian: samples were diluted appropriately and adsorbed onto freshly glow-discharged carbon-

coated copper grids that were subsequently negatively stained and EM images of each sample were collected

2.26 Atomic Force Microscopy (AFM)

Ternary complex samples were prepared as detailed in 2.17. For product samples, NADPH was added at a 100-fold excess, and the samples illuminated as in 2.16.3. Samples were deposited on charged mica grids and AFM carried out by Dr Cvetelin Vasilev on a Bruker AFM.

2.27 Dynamic Light Scattering (DLS)

Apo protein samples were prepared as detailed in 2.10. Ternary complex samples were prepared as detailed in 2.17. For product samples, NADPH was added at a 100-fold excess, and the samples illuminated as in 2.16.3. The DLS experiments were carried out at Malvern Instruments Ltd. on a Zetasizer Nano ZS instrument, fitted with a 532 nm laser.

2.28 Bioinformatics

In order to perform multiple sequence alignments of proteins with as much control as possible, the Python programming language was used. The Clustal W program (Larkin et al., 2003) was used within these scripts for pairwise alignment of each sequence with that of POR; the resultant alignments could then be further sorted and constrained using locally written scripts. HMMsearch was run, using the HMM sequence for the SDR family, against a database of PDB structures, in order to produce a database of structures to perform structural alignments again. The Rossmann fold region was isolated from the POR ITASSER structural model, with the 'alignto' function in Pymol used to align each of the SDR structures, in turn, to the Rossmann fold region.

Strain	Properties	Source/Reference
***E. coli* BL21 Rosetta™(DE3)pLysS**	*E. coli* B F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE (Cam^R)	Life Technologies
***E. coli* BL21 Star™ (DE3)pLysS One Shot®**	*E. coli* B F-*omp*T *hsd*SB (r_B-, m_B-) *galdcmrne*131 (DE3) pLysS (CamR)	Life Technologies
R. sphaeroides	*ΔbchJ*	Dr D. Canniffe (Canniffe et al., 2013)

Table 2.1 Details of the *E. coli* and *R. sphaeroides* strains used.

Plasmid	Properties	Source/Reference
pET9aHis	pET9a derivative containing the XbaI-NdeI (ribosome binding site plus six histidine tag) fragment of pET14b (Novagen), Nm^R	Dr Lucien Gibson, unpublished data. (Heyes et al., 2000)

Table 2.2 Details of the plasmids used.

Plasmid	Forward Primer	Reverse Primer
pET9aHis	5'-GCGCTTGGTGATTCTCGCTACAGTGAC AGCCAACC-3'	5'-GGTTGGCTGTCACTGTAGCGAGAATCA CCAAGCGC-3'

Table 2.3 Details of the primers used.

Chapter 3

3. Bioinformatic analysis of protochlorophyllide oxidoreductase

3.1 Introduction

As discussed in 1.9.4, POR is a member of the SDR superfamily and has a high sequence homology with the NAD(P)H binding region of the members of this superfamily. Much of the rest of the structure has less sequence homology within the superfamily (Wilks and Timko, 1995), particularly a loop region from approximately residues 154 – 190 which appears quite unique within the superfamily; a similar region has, however, been identified in carbonyl reductase (Baker, 1994), suggesting that there may be other proteins within the superfamily with similar inserts. This loop presents difficulties for sequence alignments of POR with other members of the superfamily. Though alignment is good for the N-terminus, the loop disrupts the alignment for the remainder of the protein, particularly as the C-terminal region is likely to be involved in specific substrate binding and, therefore, more variable throughout the superfamily. This chapter details a thorough bioinformatics analysis of POR, using locally written computer programs in order to sort and align sequences and to structurally align them according to the presence or absence of the insert; the aim was to improve our understanding of the POR structure and develop a model which represents an improvement upon the only previously published homology models (Townley et al., 2001; Buhr et al., 2008). The POR reference sequence used was that of *Thermosynechococcus elongatus*.

3.2 Structure and disorder predictions

One of the problems highlighted with homology analysis of POR is the relatively low sequence homology with other members of the same superfamily. For this reason it was decided to focus bioinformatics work on structural homology; within the superfamily this appears to be much stronger, particularly in the N-terminal, NAD(P)H binding region. Before undertaking any manual techniques, a number of database searches were utilised in order to get an impression of the likely structural composition of the protein and any regions which are be likely to be disordered.

3.2.1 Structural Predictions for POR and the loop region

Structural predictions were run using a number of methods: Jpred (Cole et al., 2008), CFSSP (Chou and Fasman, 1974), APSSP2 (Raghava, 2002) and YASPIN (Lin et al., 2005). Predictions used either: the full length *T. elongatus* POR sequence as a template, as with YASPIN (Fig. 3.1A) and APSSP2 (Fig. 3.1B); or only the previously described unique loop sequence, as with Jpred (Fig. 3.1C) and CFSSP (Fig. 3.1D). The characteristic alternating structure of β-sheets and α-helices in the Rossmann fold NAD(P)H binding region of SDR proteins is generally well predicted by the YASPIN and APSSP2 servers. The C-terminal region of POR, likely to be less conserved to other SDR proteins, is predicted to be predominantly α-helical with the exception of a region from residues 270 – 299. This section of residues is predicted to be mainly random coil, possibly with a small section of β-sheet. The full-length predictions from the YASPIN and APSSP2 methods also give probability values for structure confidence, with the probability score for this section of β-sheet being very low.

All results, for both the full-length and loop sequences, predict that the loop region should consist of an α-helical section, flanked by loop regions. The length and position of the predicted helix vary slightly, however all methods predict residues from Leu168 to Glu172 to be part of the α-helical structure. The CFSSP method predicts a broken helix with Lys173 exhibiting loop character. Both predictions run using the full protein sequence indicate that the final Gly190 residue of the loop is likely to form the first turn of the subsequent helix, containing the catalytic Tyrosine and Lysine. The probability score for the APSSP2 server suggests that the structure given for the loop region is moderately likely, however the YASPIN results give a very low probability for the loop region structure, suggesting the confidence in this structure is actually quite low.

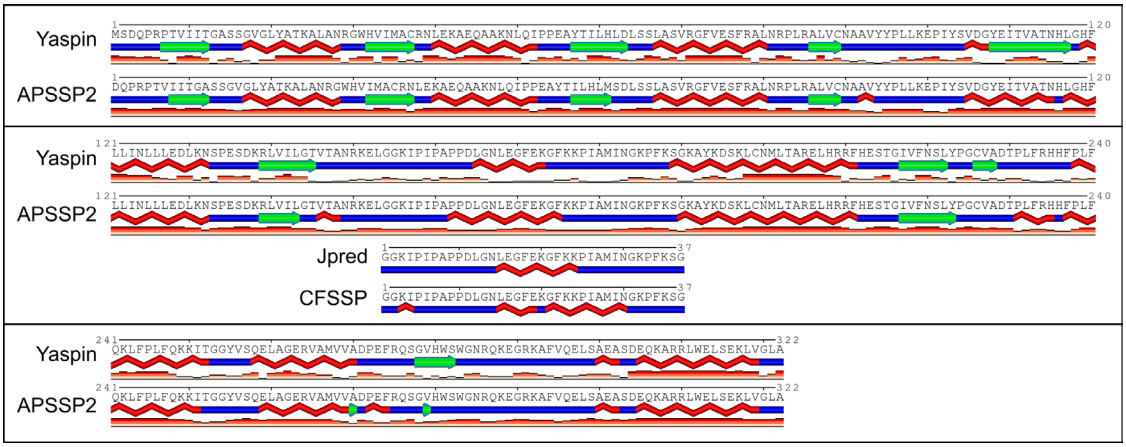

Figure 3.1 Structure predictions for the full length protein, Yaspin and APSSP2, and for the loop region only, Jpred and CFSSP. Red zigzags represent α-helical regions, green arrows represent β-strands, and blue lines indicate loop regions. The orange graph underneath the secondary structure indicates the probability score for that prediction. All predictions show some helical propensity within the loop, with varying helix length, flanked by significant loop regions.

3.2.2 Disorder Predictions for POR

In addition to the structural predictions, it was decided to utilise disorder prediction servers, in order to get an idea of regions which are not conducive to forming clear secondary structural characteristics. This is of benefit in addition to structure predictions as, instead of being suited to identification of secondary structures, they are tailored to identification of regions of disorder; these are therefore more likely to be identified in these predictions. The Genesilico Metadisorder server (Kozlowski and Bujnicki, 2012) was used for the disorder prediction as it gathers data from a large number of different servers in order to give a number of independent results.

All the results show a general trend (Fig. 3.2) with all predictions exhibiting an increase in disorder at the N-terminus and at the region just prior to residue 300; though the disorder tendency varies between servers. Many of the server results also indicate an elevated disorder tendency for residues in the loop region, with a few of these predicting that the loop region is more likely to be disordered than ordered. There are also a few servers which predicted a very low disorder tendency for the loop region, predicting that it would instead have structure; as suggested by the structure prediction results.

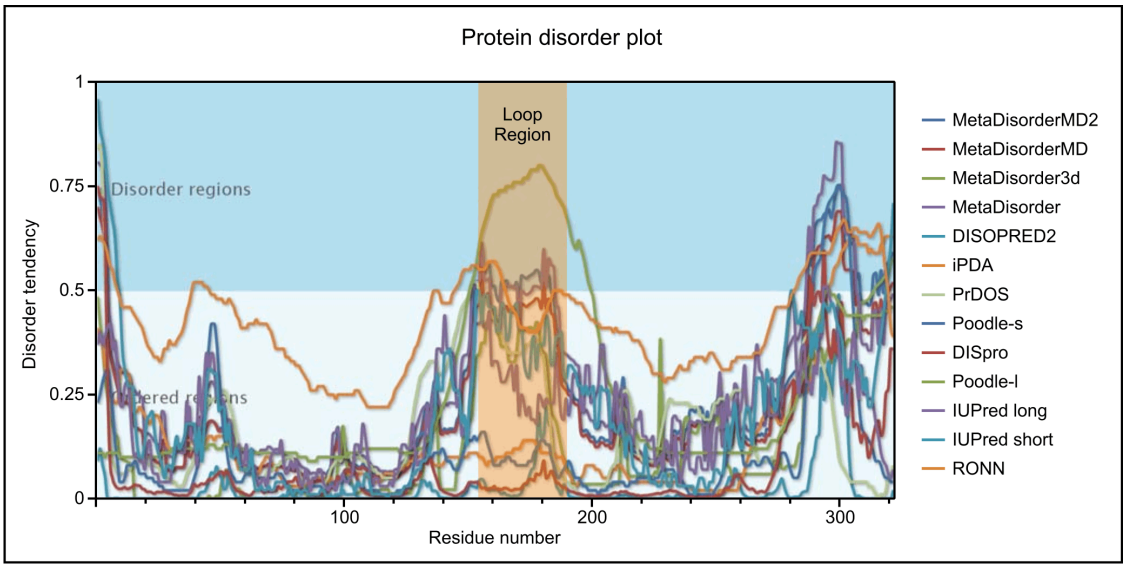

Figure 3.2 Disorder predictions for full length POR using a number of different of servers. Disorder tendency is assigned a value, with 1 being completely disordered and 0 completely ordered. All servers predict a high tendency for disorder at around residue 300. The majority of the servers also predict an elevated likelihood of disorder around the loop region of POR, indicated by the orange shading.

3.3 Homology modelling

The I-TASSER server for protein structure and function prediction (Zhang, 2008) was previously used to produce a homology model for POR (Proudfoot, 2011). This server starts by using template proteins with homologous folds before threading these together to predict a full structure. A number of the template proteins used to create the POR model were from carbonyl reductase structures, many of which have a similar loop region to that of POR. The loop in these proteins is approximately 40 residues in length and lies across two helices, which are commonly the dimerisation site in other members of the SDR superfamily. The loop in the resultant I-TASSER model is therefore heavily based on the carbonyl reductase loop as this was the only identifiable loop structure in common with POR.

3.4 Database searching and Python scripting

A search using Pfam (Finn et al., 2013), a database that groups related proteins into families, was carried out using the *T. elongatus* POR sequence; this identified it as being a member of the superfamily of short chain dehydrogenases (SDRs). The Pfam database also contains a Hidden Markov Model (HMM) for the superfamily, which is a general sequence model for the family that is generated from the probabilities of each residue in the sequence (Krogh et al., 1994). This HMM could then be used as a template to search for proteins from this family, as only those that fit the model would be selected to give a database of SDR proteins. The Swiss-Prot UniRef 100 database (Suzek et al., 2007) was used as a starting point for this, as it is a reduced sized database from the full Uniprot database yet still provides good coverage across all non-identical proteins.

Figure 3.3 Alignment of proteins from the SDR superfamily with the POR sequence. Conserved residue types coloured by ClustalW scheme: yellow = glycine, green = hydrophobic, red = polar, blue = aromatic, grey = unconserved. Consensus sequence displayed beneath residue numbers. Level of conservation indicated by bar chart; red= < 30 % identity, green-brown= 30 - 99.9 % identity. POR loop region within dotted lines. The alignment highlights the high sequence similarity within the N-terminal region and the low sequence similarity in the C-terminus. The POR loop region exhibits very little conservation within the superfamily, also containing the largest number of gaps within aligned proteins, with the exception of the extreme C-terminus.

3.4.1 ClustalW alignments

Using the Python programming language, scripts were produced (detailed 2.28) in order to perform multiple sequence alignments of POR against the other proteins in the SDR database, utilising the ClustalW program (Larkin et al., 2007). Scripts were also written to select the aligned proteins if they had certain residues conserved in the POR sequence. The conserved residues looked at were Gly154, Gly155, Gly190 and Cys226. The FASTA files of these alignments could be viewed in Geneious and coloured according to a ClustalW colouring system, indicating alignment of conserved residues throughout the sequence. Unsurprisingly, the majority of the conservation shown from these alignments is in the N-terminal region, particularly around the residues in the Rossmann fold NADPH binding pocket (Kavanagh et al., 2008). The C-termini of the proteins in the superfamily tend to be more

variable depending on the substrate, and the lack of conservation observed in this region reflects this, with a clear reduction in aligned residues. The alignments for all SDR proteins, those with the conserved Cys226, and those with Gly154, Gly155 and Gly190, are shown in Fig. 3.3, 3.4 and 3.5, respectively.

Figure 3.4 Alignment of proteins from the SDR superfamily containing an aligned cysteine residue with the POR sequence. Figure displayed as Fig. 3.3. As in Fig. 3.3, the N-terminal region has significantly higher sequence similarity than the C-terminus. However, the cysteine containing proteins appear to exhibit a slightly higher identity within the loop region, with fewer gaps and more conserved residues in the consensus sequence, though this conservation is still low. The alignment with the cysteine within the C-terminal region does not appear to improve the identity within the C-terminus.

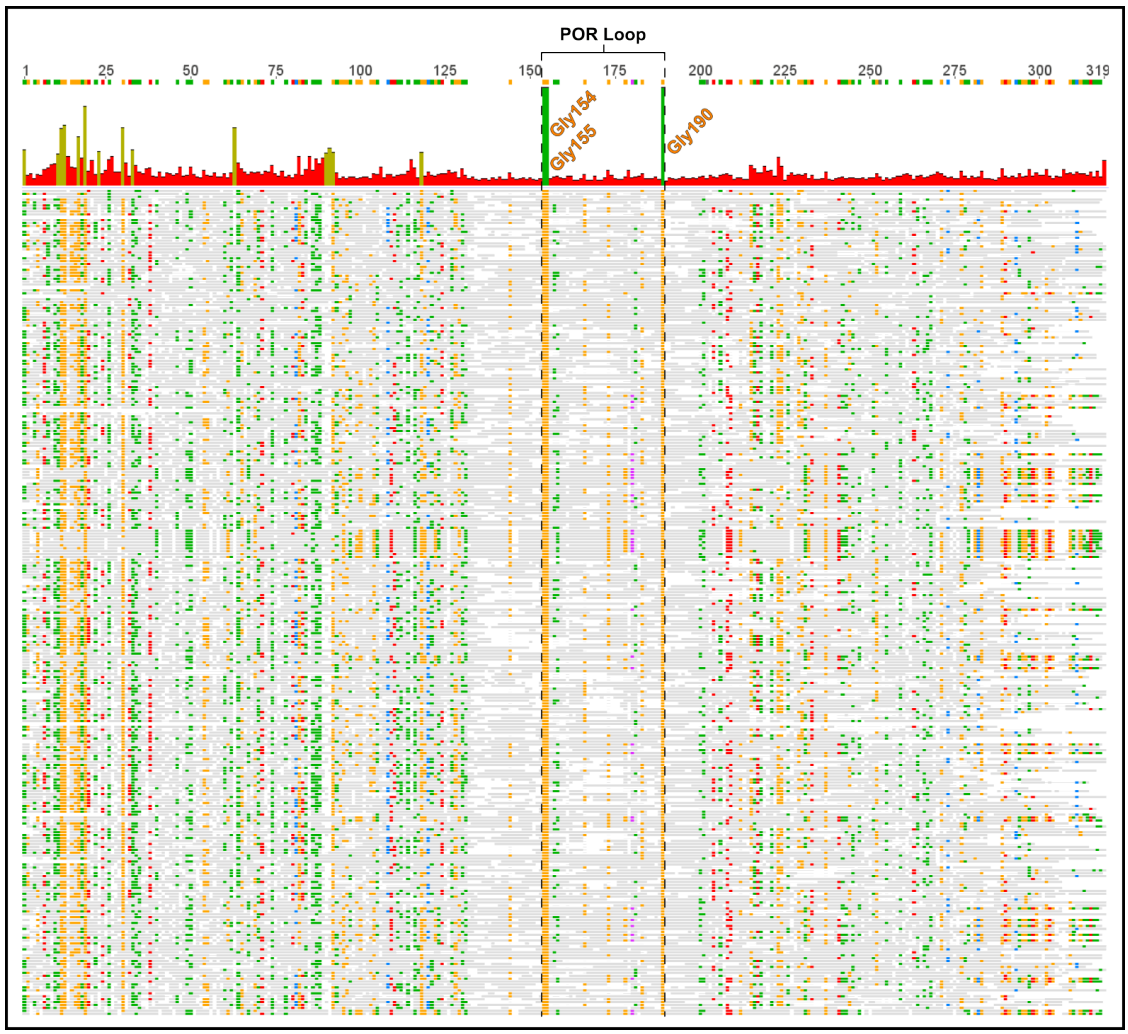

Figure 3.5 Alignment of proteins from the SDR superfamily containing aligned Gly154, Gly155 and Gly190 residues with the POR sequence. Figure displayed as Fig. 3.3. Alignment with the three glycine residues flanking the loop improves the sequence identity at both the N-terminus and the C-terminus. The loop region still exhibits a low sequence identity; though the level of identity of the regions flanking the loop is also greatly reduced.

3.5 Structural Alignment

From the alignments, it was observed that the aligned sequences with the conserved Cys226 appeared to have a higher conservation within the loop region of the POR sequence. This looked like a good selection of proteins to investigate further to determine whether these proteins did indeed contain a loop region and, for any that did, determine their substrates and structure and analyse any with a similarity to POR. The next step was to locate any of these that had an entry in the PDB, so that its structure could be examined with respect to the presence of any loop region. To do this, a FASTA file for all structures within the PDB database was downloaded, in order to detect any from the aligned sequence database using a BLAST search of the aligned sequences against the structure database.

Due to the high throughput nature of this BLAST searching, obtaining all hits in a short space of time, it was decided to do this with the whole of the original alignment, ignoring conserved

residues. This would prevent any of the possible loop-containing structures that did not have the conserved Cysteine residue from being overlooked. Further from this, an HMMsearch was carried out to refine the search to only proteins from the SDR superfamily.

The next step required the deletion of any structures that did not contain a loop region in order to focus attention on those that did; a script was used which overlays pdb files by their structure. Attempts to merge the files into a single pdb file, with all the structures overlaid, were unsuccessful due to the vast number of structures required to be aligned to each other. It was decided to pursue the structural overlay in a less automated fashion, by mass import of 120 randomly selected pdb files (approximately 20 % of the total) into PyMOL and their alignment, using the inbuilt 'alignto' function in PyMOL to align them structurally against the basic Rossmann fold template. The structures without any significant loop region at an equivalent position to POR could then be manually removed to leave just those with a POR-type insert remaining.

3.6 Loop Structure Analysis

After the alignment and removal of structures without loop regions, there were two clear types of loop structure identified. The first of these, of which all were carbonyl reductase (CBR) enzymes, had a loop of ~45 residues with a helix-turn-helix structure (Fig. 3.6A); this loop region lays across the two helices in the structure that are known to be dimerisation helices in many proteins within the SDR superfamily (Kavanagh et al., 2008). The second type of loop structure was more widespread; this loop wraps around the side of the protein away from the dimerisation helices, leaving these exposed (Fig. 3.6B). With the exception of one of these structures, these were smaller loop regions, about 10-15 residues shorter than that of POR, and also appeared to be unstructured along their full length.

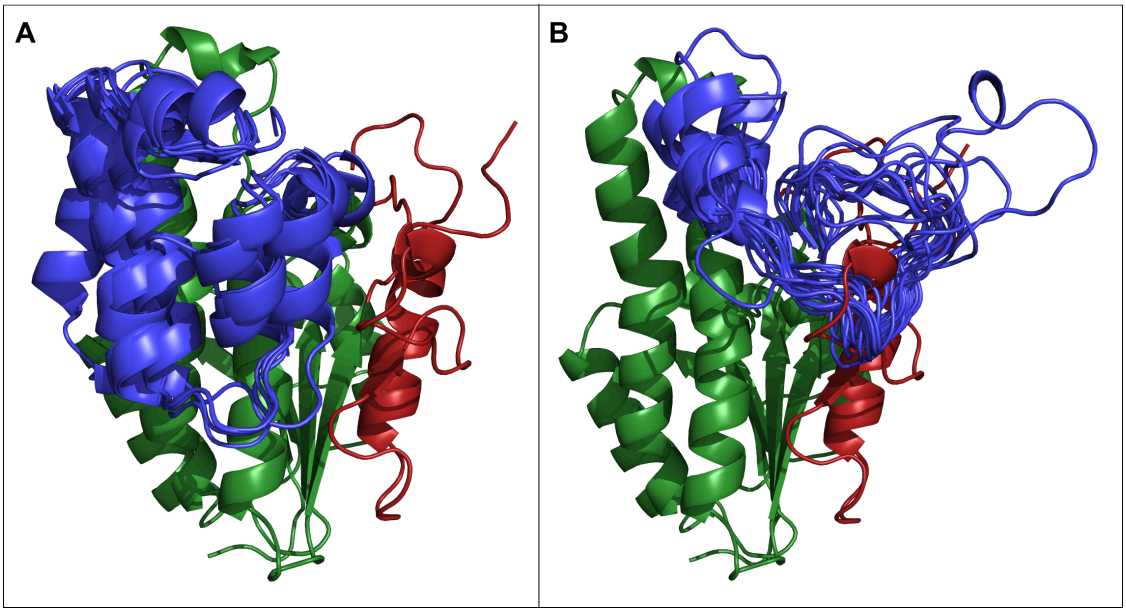

Figure 3.6 POR ITASSER model with the loop region replaced with those from (A) CBR-type loop proteins and (B) non-CBR-type loop proteins. The N-terminus is coloured green, the C-terminus red, and the loop region blue. The CBR-type loops are longer (~45 residues), mainly helical in structure, and lie across the long, hydrophobic helices. The non-CBR-type loops are shorter (~25 residues), mainly unstructured, and wrap around the side of the protein leaving the hydrophobic helices exposed.

The structures acquired from this analysis are listed in Table 3.1, along with information such as length of protein, subunits in the functional protein and substrate. From this it is clear that the main difference between the two types of loop is in the size of protein and whether it exists as a monomer, rather than in dimeric or tetrameric form, with a couple of exceptions. The CBR type loops tend to be from smaller proteins (approximately 275 residues), whilst the non-CBR type loops are from larger proteins (340 residues on average). Also, due to the coverage of the dimerisation helices by the CBR type loop (Fig. 3.7), these proteins exist as monomers, whereas the other type of loop does not block the dimerisation face, thus these exist as dimers or tetramers like the majority of proteins from this superfamily. There are two notable exceptions to this trend: the *C. testosteroni* CBR, which contains the CBR type loop yet still forms a dimer due to the presence of an alternative dimerisation face; and the vestitone reductase (VR) which, despite lacking a CBR type loop, still exists as a monomer. Due to its monomeric nature, the fact that the length of the protein is comparable to that of POR, and the similar reaction carried out, it was decided to look more closely at VR.

Non-carbonyl reductase-type loop:

PDB	Name	Res	SU's	Cofactor	Substrate
1udc	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
1g1a	DTDP-D-glucose 4,6-dehydratase	361	2	NAD+	DTDP-sugar
1uda	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
1udb	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
2cnb	UDP-galactose 4-epimerase	397	4	NAD+	UDP-sugar
1r6d	DTDP-D-glucose 4,6-dehydratase	337	2	NAD+	DTDP-sugar
2iod	dihydroflavonol-4-reductase	337	4	NADPH	flavonoids
1zze	carbonyl reductase	342	2	NADPH	aldehydes
3ehe	UDP-galactose 4-epimerase	313	2	NAD+	UDP-sugar
2x4g	NDP-sugar epimerase	342	2	NAD+	NDP-sugar
1nah	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
2p4h	vestitone reductase	322	1	NADPH	isoflavonoids
1nai	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
3gpi	NAD-dependent epimerase	286	2	NAD+	NDP-sugar
1hzj	UDP-galactose 4-epimerase	348	2	NAD+	UDP-sugar
1lrj	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
1bxk	DTDP-D-glucose 4,6-dehydratase	355	2	NAD+	DTDP-sugar
1i3k	UDP-galactose 4-epimerase	348	2	NAD+	UDP-sugar
1lrk	UDP-galactose 4-epimerase	338	2	NAD+	UDP-sugar
2pzk	WbmG	330	2	NAD+	UDP-sugar
3enk	UDP-galactose 4-epimerase	341	2	NAD+	UDP-sugar

Carbonyl reductase-type loop:

PDB	Name	Res	SU's	Cofactor	Substrate
1wma	human CBR1	276	1	NADPH	aldehydes/ketones
2hrb	human CBR3	274	1	NADPH	aldehydes/ketones
1n5d	porcine CBR	288	1	NADPH	aldehydes/ketones
2pfg	human CBR1	276	1	NADPH	aldehydes/ketones
1fjh	*Comamonas testosteroni* CBR	257	2	NADPH	aldehydes/ketones
3bhi	human CBR1	276	1	NADPH	aldehydes/ketones
3bhj	human CBR1	276	1	NADPH	aldehydes/ketones

Table 3.1 Resulting structures from the structural alignments, listed by the loop-type. Information listed are PDB codes, Protein name, number of residues, number of sub-units, whether NADH or NADPH was the active cofactor, type of substrate. This information shows that the majority of non-CBR-type loop proteins are dimers or tetramers, containing greater than 300 residues, whereas CBR-type loop proteins tend to be monomeric, of length approximately 275 residues. Vestitone reductase (blue) is different from the other proteins with this loop type in that it functions as a monomer; it also contains a comparable number of residues as POR from *T. elongatus.* The CBR from *Comamonas testosteroni* (orange) is different to the other CBR-type loop proteins as it is a functional dimer.

On closer observation of VR, it was discovered that, although the loop generally wraps around the protein away from the dimerisation helices, the latter section of the loop has a α-helical structure that covers the top section of the right-hand dimerisation helix (Fig. 3.7). This section of the helix is significantly more hydrophobic than the outward facing lower section of the helix. The reaction catalysed by VR is the reduction of a carbonyl of (3R)-vestitone to give 7, 2-dihydroxy-4-methoxy-isoflavonol, requiring NADPH. The substrate is

similar in structure to the region of Pchlide at which the POR reaction is carried out, thus the substrate binding pocket may be similar for both.

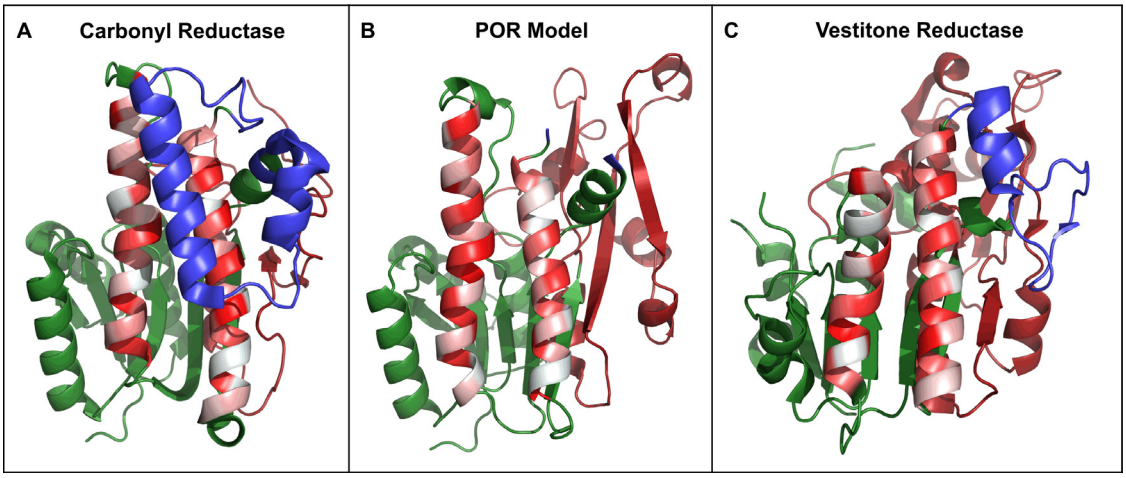

Figure 3.7 Hydrophobicity of SDR dimerisation helices for (A) CBR, (B) POR I-TASSER model (with loop not displayed), (C) VR. Dimerisation helices coloured by hydrophobicity with red representing high hydrophobicity and white low hydrophobicity. The N-terminus of the proteins are coloured green, the C-terminus red, and the loop region blue. The hydrophobicity of the helices on the CBR structure is high in the regions covered by the loop region, yet significantly lower for the exposed regions, as expected for a monomeric protein. The hydrophobicity of the VR helices is highest at the region covered by the loop and at the centre where the helices meet, while the rest is generally quite hydrophilic. The right-hand helix on the POR model appears to be quite hydrophilic, though the left-hand helix has quite high hydrophobicity suggesting that, in order for it to exist as a monomer, this region should not be exposed.

3.7 Refinement of the POR Structural Model

As the current model of POR was shorter than the CBR structure, the C-terminal region of the model was poorly defined. The POR structure was again modelled using I-TASSER, though this time using the VR structure as a template. This structure gave a more realistic representation of the C-terminal region of POR; however the modelling of the loop exposed a large number of hydrophobic residues on the typical SDR dimerisation face and was unlikely to be the loop position in apo-POR existing as a monomer. Therefore a hybrid model was produced, using sections of both models, after structural alignment, in order to produce a new model which best fitted the expectations for a POR structure (Fig. 3.8). The N-terminal region was taken from the Carbonyl-reductase model, up to the conserved Cys226 residue, as were the NADPH, modelled at this region in the Rossmann fold, and the Pchlide, whose position had been refined according to the catalytic residues in this model. From the Cys226 residue onwards, the protein is modelled on the VR-based POR model, therefore giving it a more well-defined C-terminus, which also has a more well-defined binding pocket for Pchlide. Though this model is unlikely to be especially close to the real structure, this gives a good

idea of the overall shape of the protein and the likely positions of some of the key features. This should give a good basis for the design of mutations and other experiments in the future.

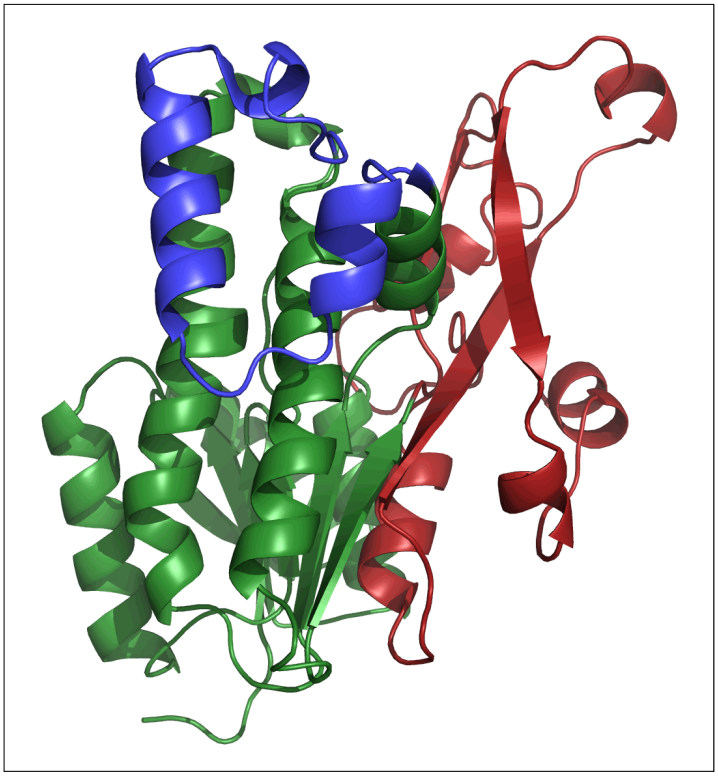

Figure 3.8 Hybrid model of POR using I-TASSER server. The N-terminus (green) and loop region (blue) are from a model based on the human CBR structure, as in the original I-TASSER model. The C-terminus (red) is from a model based on the VR structure. The loop region covers the proposed hydrophobic regions of the left-hand dimerisation helix and the C-terminus wraps around the top of the structure creating a large binding pocket for the hydrophobic Pchlide substrate.

The substrates, Pchlide and NADPH, were docked manually into the model structure, indicating their location and orientation in a POR complex (Fig. 3.9). The NADPH molecule can be quite confidently docked into the Rossmann fold, NADPH binding region, which is well-conserved throughout the SDR superfamily (Kavanagh et al., 2008); the NADPH structure is clearly positioned along the top of the loop containing the GxxxGxG motif, conserved within the SDR superfamily; this loop is highlighted in magenta in Figure 3.9B. The Pchlide molecule is docked with the C17-C18 bond of the porphyrin ring, which is reduced in the POR reaction, in close proximity to the pro-S hydrogen from the nicotinamide ring of the NADPH structure, from which a hydride is transferred in the initial photochemical step. The catalytic tyrosine and lysine (highlighted as magenta sticks in Figure 3.9B) are modelled as expected, with the hydroxyl of the tyrosine in close proximity to the C17-C18 bond of the Pchlide molecule, to enable transfer of a proton in the catalytic mechanism, and the lysine residue nearby to the catalytic tyrosine to lower its effective pKa. There are three glycine residues which flank the

loop region of POR and have been suggested as possible hinge points for loop mobility; these are residues 154, 155 and 190, highlighted with magenta spheres in Figure 3.9A. In the model, these residues are positioned close together at the top of the loop, nearest to the substrate binding site; it is certainly feasible that binding of Pchlide by POR could induce a conformational change at this position, allowing the loop to lift up, away from the α-helices that it covers. These glycine residues, and the loop region as a whole, are studied in more detail in Chapter 6.

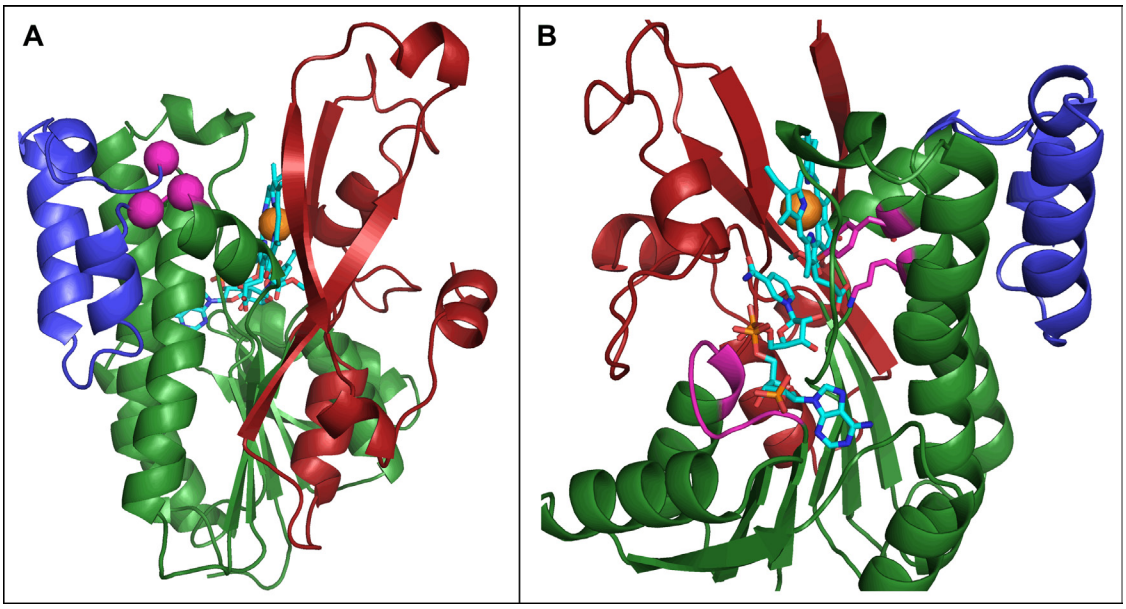

Figure 3.9 Structural model of POR represented by cartoon structure. The protein structure is coloured as in Fig. 3.8. The substrates, NADPH and Pchlide, are displayed as sticks and coloured by atom: carbon = cyan; oxygen = red; nitrogen = blue; phosphate = orange; magnesium = orange sphere. (A) The glycine residues 154, 155 and 190, suggested as possible loop hinge regions, are represented by magenta spheres. These residues are positioned high up the loop, and near to the Pchlide binding site. (B) Image focussing on the substrate binding region of the POR model. The loop region of the Rossmann fold, implicated in NADPH binding, is coloured magenta and positioned close to the NADPH molecule, suggesting it is correctly modelled. The catalytic tyrosine and neighbouring lysine, thought to lower the tyrosine pKa, are displayed as sticks and coloured magenta. The tyrosine residue is positioned close to the C17-C18 bond of the porphyrin ring, to which it is required to transfer a proton. The lysine residue is in close proximity to the tyrosine to affect its pKa.

3.8 Discussion

The unique loop region of POR is approximately 37 residues in length and thought to be involved in substrate binding or protein-protein interactions. Very little has been reported about the structure and function of this region due to the lack of structural data for the protein and the uniqueness of this region within the SDR superfamily. Bioinformatics analysis of this region has shed some light on its likely structural composition and also helped to identify proteins with more significant homology to POR.

Structure and disorder predictions of POR indicate that the loop region has a certain propensity for helical formation, though also a significant likelihood of disorder. This could suggest that the loop is unstructured in the apo protein, with a conformational change occurring upon substrate binding, therefore stabilising a more well-structured form of the loop. The difficulty in crystallising the apo POR protein has been largely put down to the presence of unstructured and/or mobile regions of protein affecting crystal packing, with the bioinformatics data agreeing with this. The large change in macromolecular structure observed upon substrate binding, with aggregates formed upon binding Pchlide, are likely to be driven by protein interactions; thus a large local structural change is to be expected in order to drive this protein-protein association.

The use of locally written homology programs enabled an increased control and specificity in the search for homologous proteins to POR. Due to the limited sequence homology within the SDR superfamily, proteins within the superfamily with solved structures were analysed, looking to isolate any proteins with loop regions similar to POR. Using a combination of ClustalW sequence alignments and structural alignments using the Pymol software, two distinct loop structures have been identified within the SDR superfamily. One type of loop structure is approximately 45 residues in length and found in most carbonyl reductase structures, whereas the other structure is generally shorter at approximately 25 residues and is found in a number of proteins, including some UDP-galactose 4-epimerases.

Analysis of the CBR structures indicated that all but one existed as monomers, despite most proteins in the superfamily existing as dimers or tetramers. This is due to the position of the loop over the two hydrophobic helices, preventing dimerisation. The *C. testosteroni* CBR, which exists as a dimer, has an alternative dimerisation site, whilst one CBR with the other loop structure was also identified, and exhibited the exposed hydrophobic helical faces, enabling dimerisation. This suggests that the POR loop could have a similar role in determining protein-protein interaction by the coverage or exposure of this hydrophobic face.

The second type of loop structure included a large number of epimerases which tend to be slightly longer than POR, especially considering the shorter loop region. They also included the aforementioned CBR structure, dihydroflavonol-4-reductase (DHFR) and vestitone reductase (VR). All of the proteins with this loop structure exist as dimers, with the exception of vestitone reductase which, despite the exposure of the helices usually involved in dimerisation, has been characterised as a monomer. The difference in surface hydrophobicity between these helices in CBR and VR is not dramatic, suggesting that it may

require a particularly specific interaction site in order for dimerisation to occur; thus it is difficult to predict whether this region is involved in POR interactions.

The final model of POR was based on the CBR N-terminus and loop, and the VR C-terminus. The N-terminal Rossmann fold region of the model is remarkably similar to those of both the Townley and Buhr models, as would be expected for such a well conserved motif. The loop region in the model presented here is very different to that of the Townley model, which places it over the proposed Pchlide binding site in order to exclude water from the binding site; our model instead comprises a well structured helical insert away from the substrate binding site and over the SDR dimerisation helices, more in line with the evidence presented by Reinbothe et al. implicating the loop in oligomerisation rather than substrate binding; the loop is not modelled into the Buhr model, however is suggested to exist as a helix-turn-strand as it exists in the Townley model. The other contrasting aspect of the three models is the C-terminal substrate binding region; in our model there is a large amount of β-sheet structure in this region, in addition to unstructured loops, whereas in the other models the C-terminus is predominantly loops, with a small amount of helical structure. It is difficult to comment on the accuracy of modelling in the substrate binding region due to the limited homology to other members of the SDR superfamily in this region.

Chapter 4

4. Preparation and structure determination of chlorophyll precursors

4.1 Introduction

This chapter discusses the purification and characterisation of two different, but related, chlorophyll precursors: protochlorophyllide (Pchlide), discussed in sections 4.2 to 4.4, and a pigment produced following a knockout of Ycf54, discussed in section 4.5. The enzyme protochlorophyllide oxidoreductase (POR) catalyses the light-dependent reduction of either monovinyl (MV) or divinyl (DV) protochlorophyllide (Pchlide) [Fig. 4.1] to form chlorophyllide (Chlide) (detailed in 1.9-1.10). In order to study complexes of POR with its substrates, a large amount of pure Pchlide is required; the substrate is not produced commercially therefore must be produced and purified *in vivo*. The high purity of the pigment is necessary in order to produce high quality protein complexes and enable efficient turnover of the enzyme.

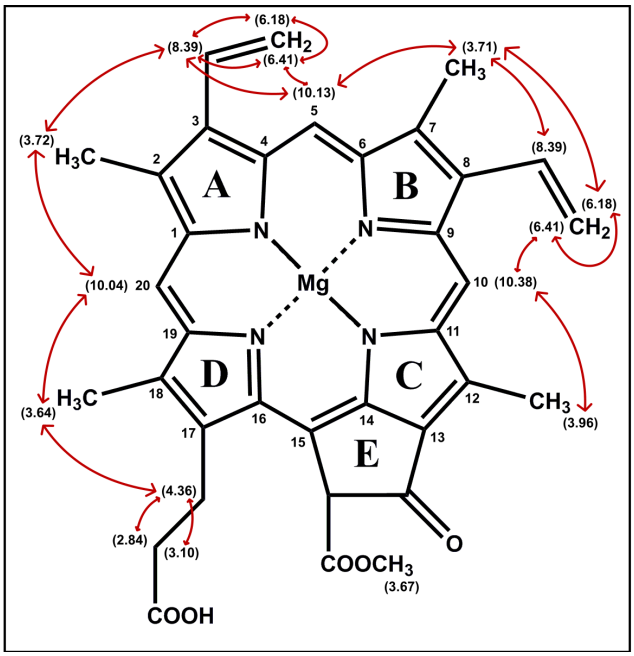

Figure 4.1 Structure of divinyl protochlorophyllide (DV-Pchlide), as confirmed by NMR data from 4.4. The carbon atoms of the porphyrin ring are numbered by the IUPAC system. The chemical shift of each unique proton is displayed in brackets. Red arrows indicate observed NOEs. DV-Pchlide has vinyl groups at both the 3' and 8' positions, whereas monovinyl Pchlide has an ethyl group at the 8' position.

4.2 Initial Protochlorophyllide Purification

A protocol had previously been established for the production of large quantities of Pchlide for use in protein assays and POR complex formation (Proudfoot, 2011); this was a modification of a previous protocol, originally described by (Heyes et al., 2000). This protocol was used to produce large quantities of Pchlide from a relatively simple initial extraction; this resulted in a mixture of MV and DV-Pchlide forms, both of which are functionally active substrate for the POR protein. It was subsequently found, however, that the purity of the pigment was low; an alternative protocol was therefore devised, in order to improve the purity of Pchlide, detailed in 4.3. The original protocol, and subsequent analysis of Pchlide purity, is detailed below.

4.2.1 Initial Method for Pigment Expression and Bung Extraction

Pchlide was produced from the *Rhodobacter capsulatus* ZY5 mutant, which contains a single point mutation in one of the subunits for the light-independent POR (DPOR) protein; thus leading to the accumulation of Pchlide. *R. capsulatus* ZY5 from a glycerol stock was streaked out on a VN agar plate with 25 mg/ml rifampicin, and subsequently incubated for 3 - 4 days at 34 °C. Colonies were picked and used to incubate VN starter cultures, incubated at 34 °C with minor agitation for 2 days, until green in appearance. These were then used to inoculate 2 L VN growths which were then incubated at 34 °C overnight, with agitation.

Autoclaved polyurethane foam bungs were added to the medium and the flasks incubated for a further 24 hours, during which time the pigment was adsorbed into the bungs. These bungs were removed and transferred to 100 % methanol, into which the pigment was resolubilised. The bungs were replaced with fresh, autoclaved bungs and the process repeated until the pigment stopped being excreted. The extracted pigment in methanol was placed in a rotary evaporator and the methanol boiled off under vacuum. This was repeated until all of the pigment remained in only minimal methanol.

4.2.2 Purification of Pigment on CM-Sepharose Column

The methanol sample was taken up to 1 L with methanol and acetone, to a final ratio of 5 % methanol and 95 % acetone. The Pchlide was subsequently purified on an acetone resistant CM sepharose column, using a stepwise increase in methanol concentration for specific elution of molecules by polarity, as described by (Heyes et al., 2003); carotenoids eluted at 15 % methanol in acetone, with Pchlide eluted at 25 % methanol. The eluted Pchlide was

then transferred to a rotary evaporator and the solvent evaporated off until the pigment was highly concentrated in minimal solvent. This solution was then aliquotted and dried overnight in a vacuum centrifuge, while kept in the dark. Dried Pchlide samples were stored in eppendorfs at -20 °C, and wrapped in foil to prevent exposure to light.

4.2.3 NMR of Pchlide

In order to ascertain the purity of the pigment produced by this method, it was decided to analyse a sample of Pchlide, prepared from this protocol, using NMR. The dried pigment from HPLC was resuspended in 500 µl methanol-d4 (Sigma Aldrich), centrifuged to remove any insoluble pigment, transferred to a 5 mm NMR tube and sealed. All NMR experiments were carried out on a Bruker Avance DRX 600 instrument equipped with a cryoprobe, at an acquisition temperature of 298 K.

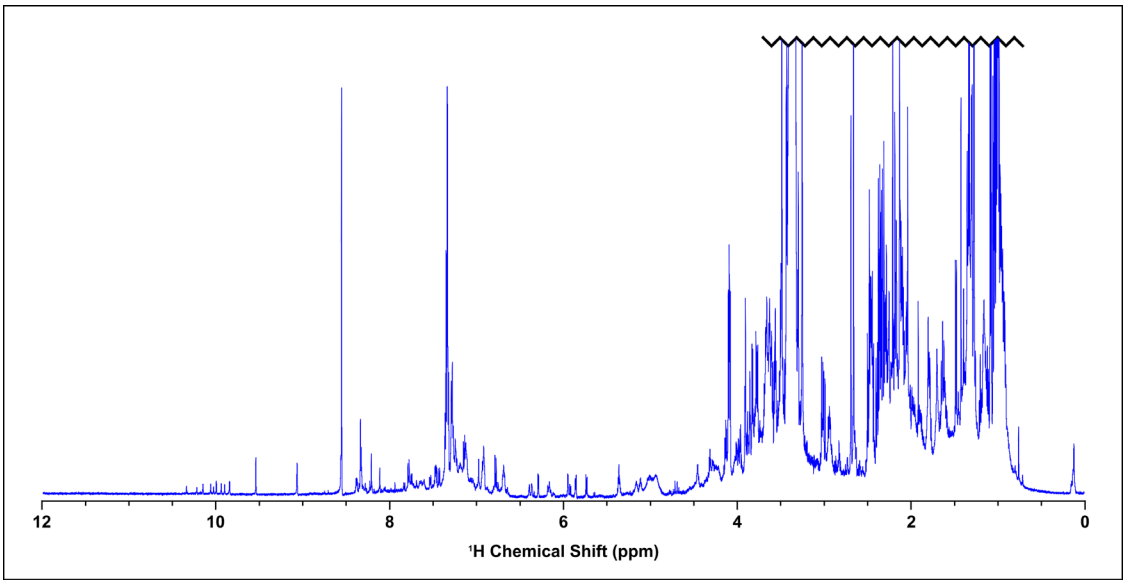

Figure 4.2 One-dimensional proton NMR spectrum of Pchlide preparation from *R. capsulatus* ZY5 strain. A number of characteristic porphyrin meso proton peaks can be seen at around 10 ppm, suggesting that multiple porphyrin molecules are present in the sample. There are massive amounts of contaminating materials upfield of 4 ppm indicating the considerably low purity of the sample.

A one-dimensional ^{1}H experiment was carried out on the sample in order to analyse the signals present and determine the relative purity of pigment (Fig. 4.2). Due to the limited number of protons within the Pchlide structure, the number of peaks expected to be in the spectrum is quite low. This is clearly not the case, as massive amounts of signals are observed, particularly between 1 and 4 ppm. A number of peaks at around 10 ppm represent characteristic porphyrin meso proton peaks, suggesting that there maybe be further

porphyrin molecules present in addition to Pchlide; there are three meso protons in Pchlide, therefore only three signals should be expected here, or six if the signals are different in MV and DV Pchlide forms. These peaks are also low intensity compared to the contaminating materials, indicating that the purity of the sample is extremely low.

4.3 Optimisation of Protochlorophyllide Preparation

One of the significant problems with studying the structure of POR is the high levels of aggregation of the ternary complex. It is thought that this aggregation could be driven by, or at least exacerbated by, the poor level of purification of the Pchlide substrate and therefore the presence of hydrophobic contaminants. In order to improve the quality of these ternary complex preparations, a new protocol was devised in order to produce significant quantities of highly pure Pchlide.

4.3.1 Protochlorophyllide Expression and Solvent Extraction

It was noted that, in other pigment purification procedures, solvents such as diethyl ether were used to extract pigments directly from the media. Solvent extraction was attempted with the *R. capsulatus* ZY5 mutant, however the solvent layer of the extraction contained very high levels of an unknown lipid; this made the solvent extraction procedure unfeasible, as huge volumes of solvent were required and only low levels of Pchlide were extracted, therefore an alternative Pchlide producing strain was identified.

A *R. sphaeroides* Δ*bchJ* strain, which accumulates high levels of Pchlide, was provided by Dr S. Hollingshead for the preparation. This strain was grown (detailed 2.15) with the remaining media, coloured green with Pchlide, kept in the dark to prevent photodamage of the pigment. Solvents were used in order to extract the pigment from the aqueous media. Diethyl ether (Et_2O) was added to the media, agitated and the solution was allowed to settle, with the pigment expected to migrate to the organic solvent layer of the suspension; however this did not occur, with a significant majority of the pigment remaining in the aqueous layer.

When Acetonitrile (ACN) was added to the suspension of media and Et_2O, agitated and allowed to settle, the resultant organic solvent layer was intensely green in colour, compared to the water layer, indicating that Pchlide had migrated into the organic solvent. After testing various combinations of solvent volumes, it was found that optimum extraction required 2 volumes of media to 1 volume of solvent; the solvent composed of 2 volumes ACN and 1

volume of Et_2O. The solution was transferred to a separating funnel and the lower, water-based phase discarded and the green-coloured, solvent-containing upper phase retained. In order for the Pchlide to be taken down to useable volumes, the solvent required evaporating off using a rotary evaporator (detailed 2.15.3). The Pchlide, in minimal volume, could then be aliquotted into eppendorfs and dried overnight in a vacuum centrifuge.

4.3.2 HPLC Purification Using Reverse-phase Column

The solvent extraction of the Pchlide pigment was a relatively crude separation of pigments from the other media components; therefore further purification was required in order to give pure Pchlide (detailed 2.15.4). High Performance Liquid Chromatography (HPLC) enables high quality purification and the use of a preparative C18 column enabled significantly higher quantities of pigment to be run through the column. Dried pigment was resuspended in minimal methanol, and an equal volume of water added to give a 50 % methanol solution. Samples were loaded via a 500 µl loop and a method run in order to gradually increase the concentration of methanol in order to specifically elute the pigments from the C18 resin depending on their hydrophobicity. The HPLC traces for absorbance, at 440 nm and 630 nm, and fluorescence, at 440 nm excitation and 632 nm emission, are shown in Figure 4.3, along with the methanol gradient. The signals are normalised at the 13.8 min peak for clarity.

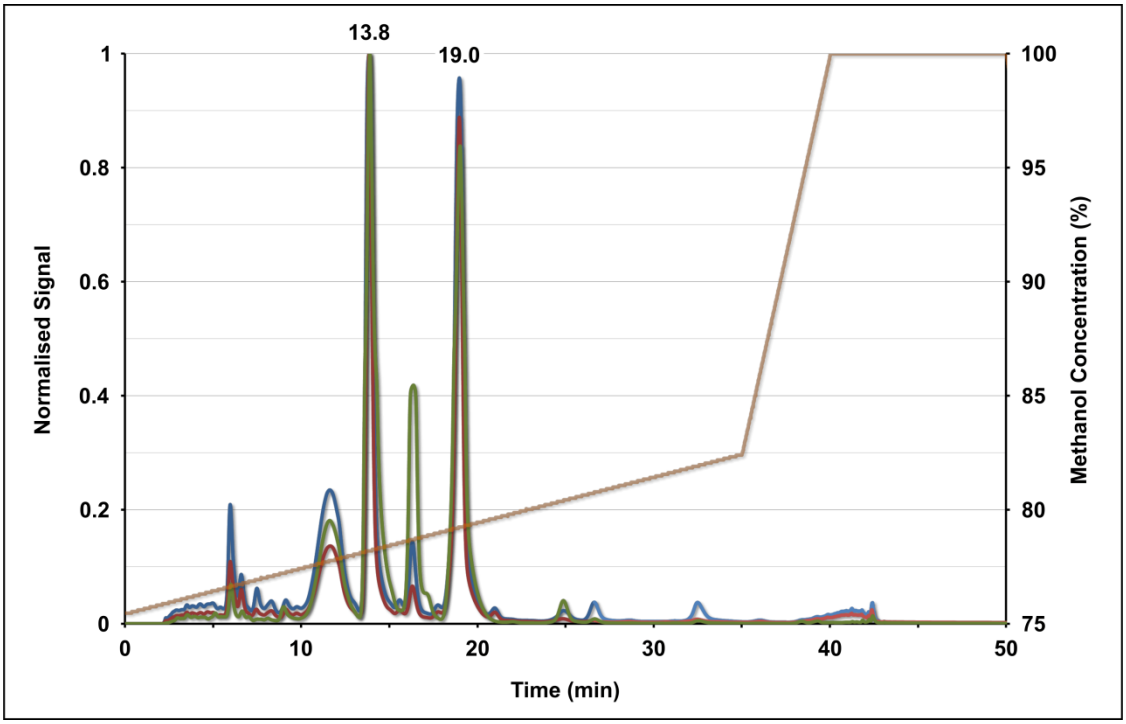

Figure 4.3 HPLC traces of Pchlide sample run on a reverse-phase preparative C18 column. Signals for absorbance at 440 nm (blue), 632 nm (red) and fluorescence (green), with excitation at 400 nm and emission at 630 nm, are shown and normalised to the peak at 13.8 min. The methanol gradient is shown in orange. There are a number of minor peaks present, though the two peaks at 13.8 and 19.0 min are most significant and occur in all traces.

From the HPLC trace it is clear that there are many species present with similar absorption profiles to Pchlide, though most peaks are relatively minor. There are two major peaks at 13.8 and 19.0 min, representing two distinct molecules according to the elution profile. The absorbance spectra of these two peaks, however, are identical (Fig. 4.4), indicating that both contain Pchlide. The discrepancy between the two peaks is not clear from the HPLC analysis so it was decided to carry out NMR on samples from each peak in order to identify the difference between them and to ensure that pure, homogeneous Pchlide was being produced in the purification.

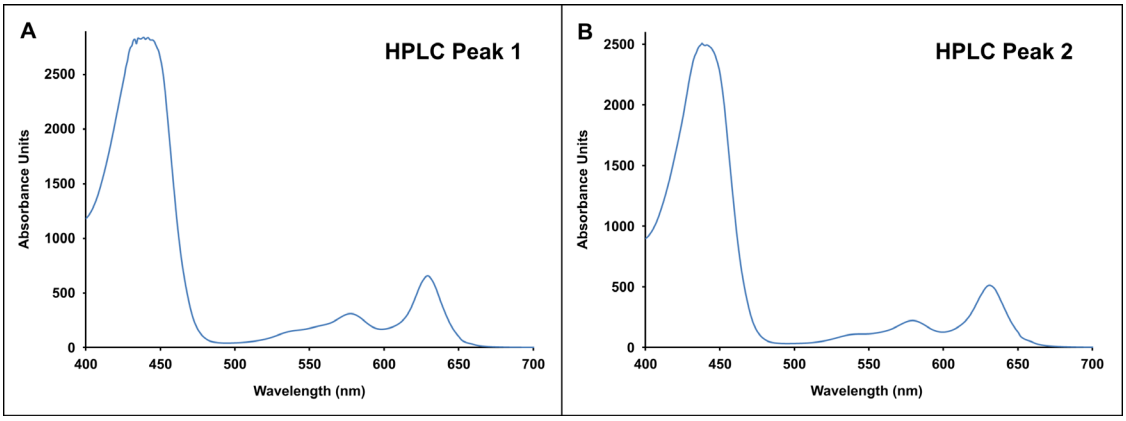

Figure 4.4 Absorbance profiles of the HPLC peaks at 13.8 min (A) and 19.0 min (B). Both spectra appear identical, with the Soret band at 440 nm and the characteristic Pchlide absorbance peak at 630 nm. The Soret band in the 13.8 min peak is slightly distorted due to reaching the maximum absorbance for the detector, however is still consistent with the absorbance spectrum of the 19.0 min peak.

4.4 NMR of Pure Protochlorophyllide

In order to ascertain the purity and quality of the Pchlide samples, HPLC purified pigment was analysed by NMR. The HPLC peaks at 13.8 and 19.0 min were taken to be analysed by NMR in an attempt to determine any differences between them. Further experiments would also confirm the presence of either MV or DV Pchlide and ensure all other structural features were consistent with the structure.

4.4.1 One-Dimensional Protochlorophyllide Spectra

One-dimensional ^{1}H experiments were run on both HPLC peaks, with peaks representing all of the proton signals in the sample. The spectrum for the 13.8 min peak (Fig. 4.5) was remarkably clean, with very few impurities and a spectrum that was characteristic of a single Pchlide species. The spectrum for the 19.0 min peak (Fig. 4.6), however, contained significantly more signals; most of these appearing to represent various porphyrin molecules.

The absorbance spectrum of this peak indicated that the only pigment present was Pchlide, with only a single peak present demonstrating the presence of one distinct molecule. This suggested the presence of a homo-oligomer of Pchlide, which is unlikely to be a functional substrate, thus this peak was discarded from the Pchlide purification and further analysis was carried out only on the 13.8 min HPLC peak.

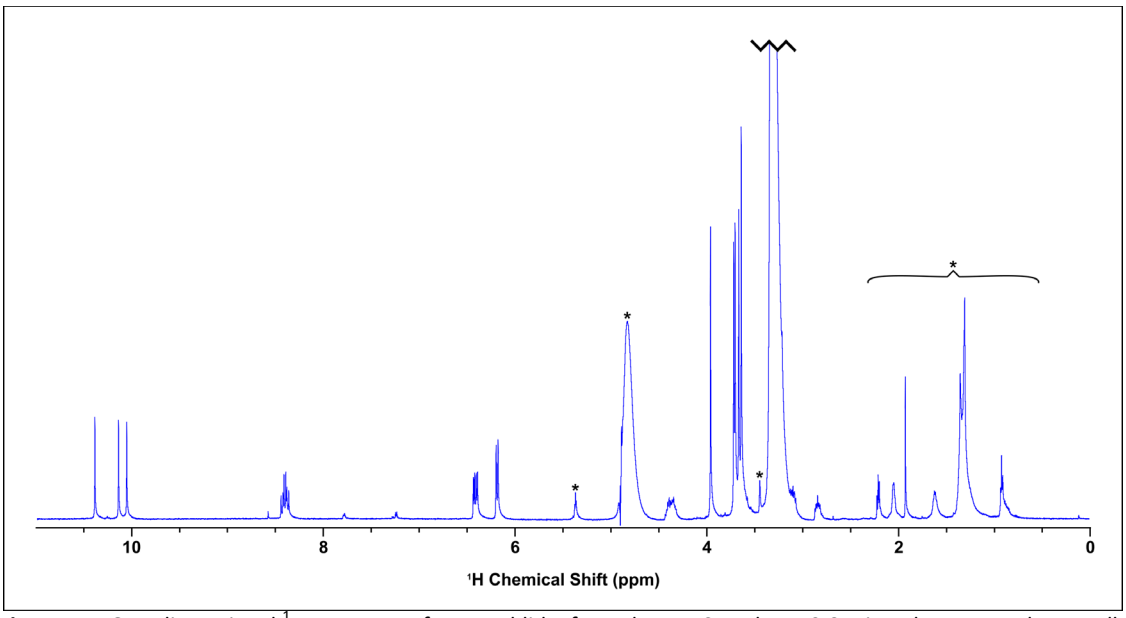

Figure 4.5 One-dimensional ^{1}H spectrum of pure Pchlide, from the HPLC peak at 13.8 min. There are only a small number of contaminants, marked with an asterisk; many of these represent solvent peaks, with water at 4.7 ppm, methanol at 3.34 ppm, and peaks from acetone, acetonitrile and diethyl ether groups expected to come at 1 - 2 ppm. The 3 meso protons can be clearly identified downfield at around 10 ppm. There are a number of peaks between 3.5 and 4 ppm, representing methyl groups. Non-singlet peaks between 2.5 and 8.5 ppm represent protons with coupling to other nearby protons.

From the one dimensional spectrum of this sample, a number of signals can be identified due to their characteristic chemical shifts. The peaks at around 10 ppm correspond to the 3 meso protons, the protons directly attached to the porphyrin ring at carbons 5, 10 and 20 of the Pchlide molecule. The cluster of peaks at 3.5 - 4.0 ppm correspond to methyl signals, yet it was not possible to deduce which is which without further experiments. A number of other coupled peaks between 4 and 8.5 ppm represent CH_2 or CH groups from the vinyl or propionate groups of the Pchlide. A small number of signals are present in the spectrum representing contaminants from the buffer, solvents or HPLC column resin, marked with an asterisk in the figure.

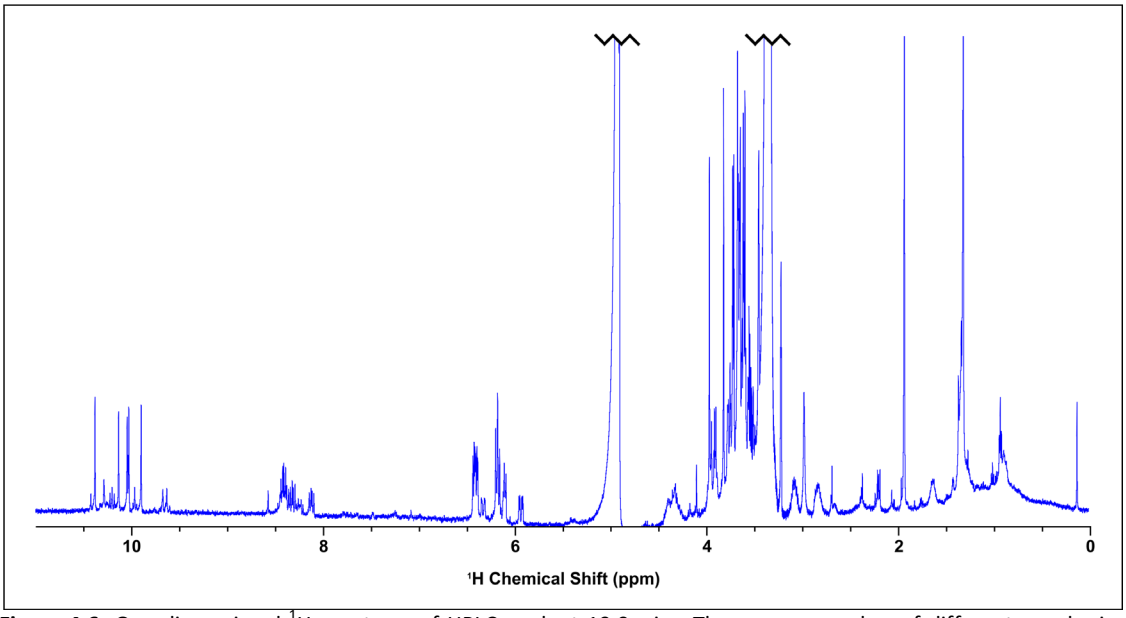

Figure 4.6 One-dimensional ^{1}H spectrum of HPLC peak at 19.0 min. There are a number of different porphyrin peaks visible in the spectrum, notably the meso proton peaks at around 10 ppm. The number of non-porphyrin contaminants upfield of 3 ppm is otherwise quite low, suggesting quite a pure porphyrin sample.

4.4.2 ^{13}C-HSQC Protochlorophyllide Spectrum

In order to discern more specific details about the environment of the proton signals, a two-dimensional, natural abundance ^{13}C-HSQC experiment was run. This experiment transfers magnetisation from the proton to the coupled carbon atom and back again, with signals for every unique CH bond in the sample. The characteristic chemical shift of each carbon atom, depending on its environment, enables assignments to be confirmed for the proton signals, and enables any other signals with inconsistent carbon shifts to be discarded as not belonging to the Pchlide structure. The ^{13}C-HSQC spectrum is shown in Figure 4.7.

The methyl region of the HSQC is expanded inset of Figure 4.7, with the one-dimensional spectrum at this region overlaid. The methyl ester on the E ring of Pchlide can be easily identified due to its carbon shift at 52.0 ppm, downfield of the other methyl shifts which all have characteristic carbon-methyl shifts of around 10 - 12 ppm. These shifts all correspond to methyl groups attached directly to the protoprophyrin ring, whereas the methyl ester signal is shifted significantly downfield due to its bond with the electron withdrawing oxygen atom, reducing its shielding from the magnetic field.

Most of the other features of the Pchlide structure can be discerned from the ^{13}C-HSQC, with the exception of the 13^{2} proton on the E-ring which cannot be identified. The 17^{1} protons are both found at 4.36 ppm in the proton dimension, with a slight separation, and with a carbon shift of 22.6 ppm. The 17^{2} protons have a more significant separation with proton shifts of 3.10 and 2.84 ppm, though both have the identical carbon shift of their shared carbon atom

at 37.5 ppm. Though both 17^1 and 17^2 are CH_2 groups, the 17^1 protons are shifted downfield due to their proximity to the aromatic porphyrin ring; with the 17^2 carbon shifted downfield due to the bonding to a carboxyl group.

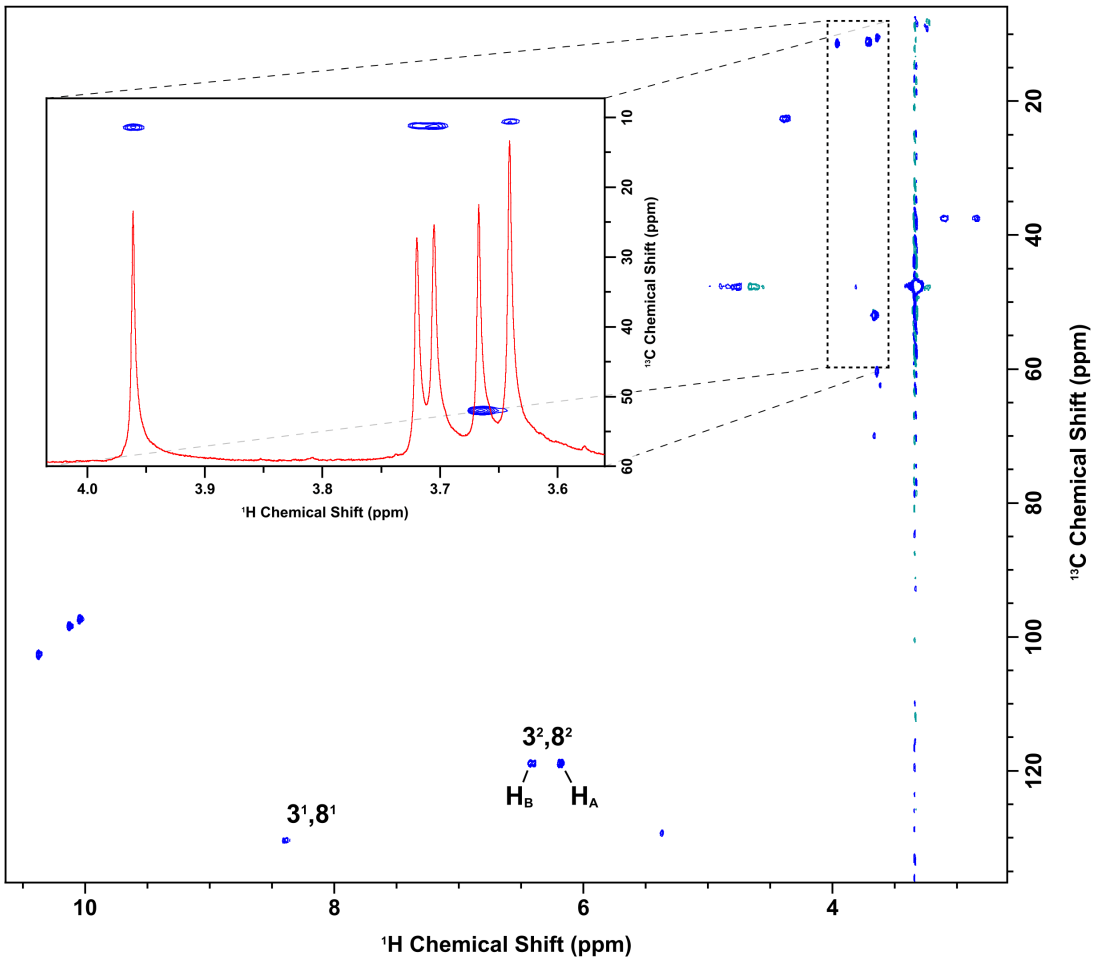

Figure 4.7 ^{13}C-HSQC spectrum of pure Pchlide. Blue and green indicate positive and negative signals, respectively. The methyl region is expanded inset, with the one-dimensional spectrum overlaid. The downfield shifted methyl peak at 52.0 ppm represents the methyl ester at carbon 13^4. Three downfield, two at 119.0 ppm and one at 130.4 ppm, represent signals from the vinyl groups.

The integral of each of the peaks between 6 and 8.5 ppm was calculated to be twice that of the meso proton peaks, indicating the presence of two proton signals. Due to their characteristic signals, it was determined that they represented the 3 and 8-vinyl groups, which are overlapped in both dimensions due to their identical composition and highly similar environment, thus confirming the presence of DV-Pchlide. The 3^1 and 8^1 groups have a proton shift at 8.39 ppm and a carbon shift at 130.4 ppm, characteristic of a carbon shift for a CH alkene group. The CH_2 groups of the alkene have two distinct proton shifts, at 6.41 and 6.18 ppm, but a single carbon shift of 119.0 ppm; this indicates that the distinct proton shifts are representative of the *cis* and *trans* protons of the vinyl group, rather than the protons

from the 3^2 and 8^2 groups having distinct shifts between them. Finally, the meso protons at 10.38, 10.13 and 10.04 ppm have associated carbon chemical shifts of 102.7, 98.4 and 97.4 ppm respectively, though confirming which signal corresponds to which atom on the structure is still not possible.

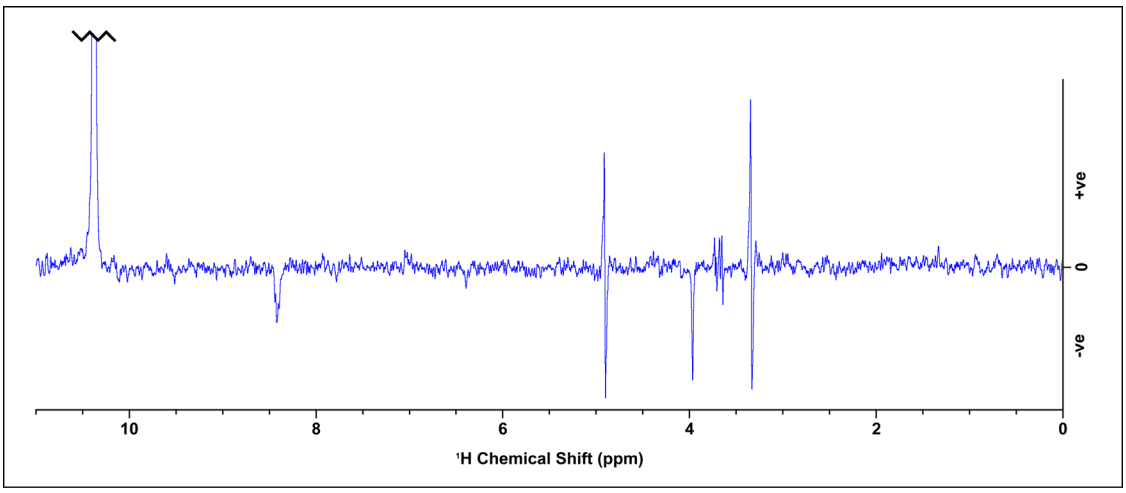

Figure 4.8 Selective NOE experiment on Pchlide sample. The meso proton signal at 10.38 ppm is selectively excited and magnetisation transferred through space to neighbouring protons. Negative peaks indicate protons with NOEs. Solvent peaks give artefacts with both negative and positive signals, due to subtraction errors.

4.4.3 Selective NOE Experiments

The Nuclear Overhauser Effect (NOE) involves the transfer of magnetisation through-space between atoms, due to the proximity of the atoms and their cross-relaxation. An experiment has been developed which enables selective excitation of proton signals, followed by the evolution of NOEs, giving signals for other protons which are close in space (Stott et al., 1995). This experiment was carried out on the Pchlide sample, selectively exciting each signal to identify other signals which are in close proximity. The experiment produces a spectrum with a large positive signal for the excited atom and negative signals from atoms which have an NOE to the excited signal and are therefore in close proximity. All signals could then be assigned to their respective atoms by analysis of their NOE signals in conjunction with their proton and carbon shifts. All assigned atoms, along with their associated proton shifts, carbon shifts and NOE cross peaks, are displayed in Table 4.1, with proton shifts and NOE connectivities also displayed on Fig. 4.1.

Proton	δ [^{1}H] (ppm)	δ [^{13}C] (ppm)	Protons with NOE correlations
2^1	3.72	11.2	3^2 (H_B), 20
3^1 (H_X)	8.39	130.4	3^2 (H_A + H_B), 5
3^2 (H_B, *trans*)	6.41	119.0	2^1, 3^2 (H_A), 5
3^2 (H_A, *cis*)	6.18	119.0	3^1, 3^2 (H_B)
5	10.13	98.4	3^1, 7^1
7^1	3.71	11.3	5, 8^1, 8^2 (H_B)
8^1 (H_X)	8.39	130.4	7^1, 8^2 (H_A + H_B), 10
8^2 (H_B, *trans*)	6.41	119.0	7^1, 8^2 (H_A), 10
8^2 (H_A, *cis*)	6.18	119.0	8^1, 8^2 (H_B)
10	10.38	102.7	8^1, 8^2 (H_B), 12^1
12^1	3.96	11.3	10
13^4	3.67	52.0	-
17^1 (H_A + H_B)	4.36[a]	22.6	17^2 (H_A), 18^1
17^2 (H_A)	3.10	37.5	17^1, 17^2 (H_B)
17^2 (H_B)	2.84	37.5	17^1, 17^2 (H_A)
18^1	3.64	10.6	17^1, 20
20	10.04	97.4	2^1, 18^1

Table 4.1 ^{1}H and ^{13}C-NMR assignments for Pchlide, with NOE correlations. Atoms numbered as in Fig. 4.1. NOE connectivities to either 3' or 8' atoms were deduced from neighbouring signals.
[a] = 17^1 protons overlapped.

4.4.4 Final Pchlide Preparation

The sample prepared in 4.4 can be produced in relatively large quantities by a straightforward protocol; the final Pchlide preparation has been shown to be highly pure by NMR. This preparation has therefore been used for the production of complexes of POR, as described later in the chapter. The increased purity of the pigment seems to have a dramatic effect on the complex formation, suggesting that the previous purification was inadequate for the production of Pchlide suitable for structural studies. This new preparation should, therefore, be considered as the preferred method for production of Pchlide for use in future experiments.

4.5 NMR and Structure Determination of A433 Pigment

The enzyme immediately prior to POR in the chlorophyll biosynthesis pathway is Mg-protoporphyrin Monomethyl Ester cyclase (Mg cyclase); catalysing the conversion of Mg-protoporphyrin monomethyl ester (MgPME) to Pchlide (detailed 1.8). This reaction involves cyclisation of the carbon chain at position 13 of the porphyrin ring, forming the E-ring of the final chlorophyll molecule. The exact mechanism of this reaction is unknown; however it

must occur via multiple steps, due to the complexity of the reaction. Isolation of intermediates from this reaction would greatly enhance our understanding of the enzymatic mechanism of Mg cyclase.

Work on the Mg cyclase in cyanobacteria by Dr S. Hollingshead has led to the identification of the Ycf54 protein. This protein was found to be associated with Sll1214 and Sll1874, the known catalytic components of the Mg cyclase in *Synechocystis*, via pull down experiments using these proteins initially as bait, followed by the reciprocal pull down using the Ycf54 protein as bait. A *Δycf54* strain of *Synechocystis* was made and the pigments produced by this strain analysed with HPLC, by Dr Hollingshead. The product of the Mg cyclase reaction, Pchlide, was only identified at trace levels, while MgPME had accumulated to high levels. This suggested that the Ycf54 protein was not essential for the function of Mg cyclase, however its deletion does significantly affect its function.

In addition to these pigments, an unknown pigment was identified with a Soret band absorbance of 433 nm, in contrast with MgPME and Pchlide which have absorbance at 418 and 440 nm, respectively. Due to the spectral properties of this pigment, with absorbance between the substrate and product of the Mg cyclase reaction, it was hypothesised that this may be an intermediate of the reaction which had accumulated due to the impaired function of the Mg cyclase. In order to identify the nature of this pigment, this pigment was purified and prepared for analysis by NMR to enable the exact structure of the pigment to be deduced.

4.5.1 NMR of Mg-protoporphyrin monomethylester IX

In order to aid the structure determination of the unknown pigment, the substrate of the Mg cyclase reaction, MgPME, was initially analysed by NMR. By assigning NMR spectra for the Mg cyclase substrate MgPME, along with the Pchlide product that was analysed previously, the identity of the unknown pigment can be very confidently resolved by comparison of signals to those in the MgPME and Pchlide spectra. Purified pigment was prepared as in 4.4.1, with all experiments conducted on a Bruker Avance DRX 600 instrument equipped with a cryoprobe at an acquisition temperature of 298 K. The structure of the MgPME pigment is detailed in Fig. 1.13.

4.5.1.1 One-Dimensional Mg-protoporphyrin monomethylester IX Spectrum

The one-dimensional proton spectrum indicates the high purity of the pigment, with very few contaminants apparent, particularly downfield of the methanol peak at 3ppm (Fig. 4.9). There are four peaks present at around 10 ppm, consistent with the four meso protons of the MgPME structure at positions 5, 10, 15 and 20 of the porphyrin ring. The methyl protons are grouped together at approximately 3.7 ppm, while various coupled peaks represent the CH_2 groups on the 13 and 17-chains as well as the 3 and 8-vinyl groups.

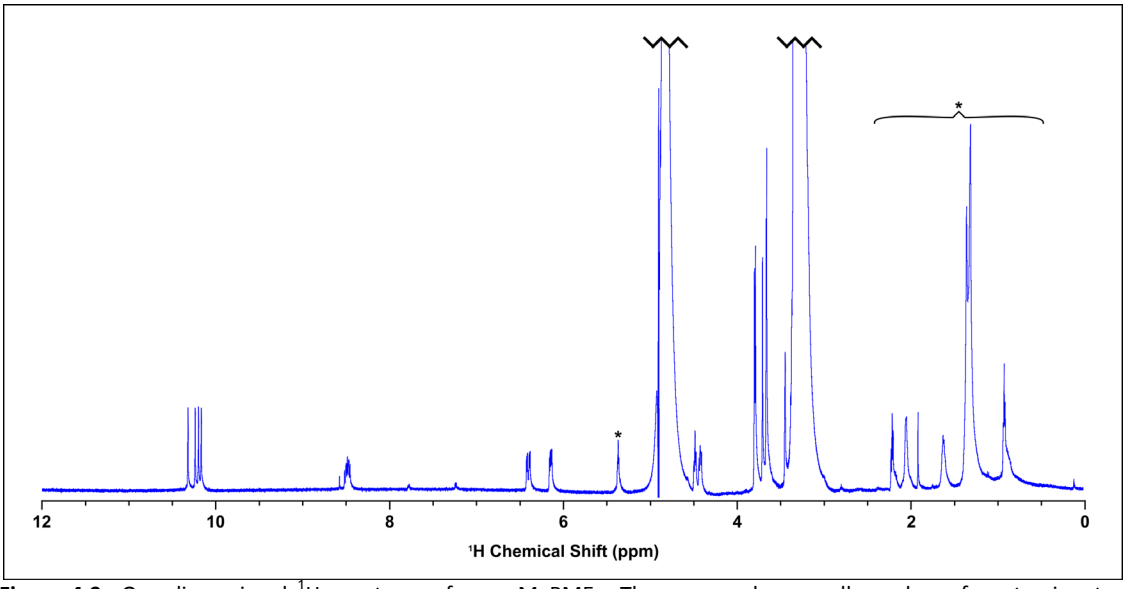

Figure 4.9 One-dimensional ^{1}H spectrum of pure MgPME. There are only a small number of contaminants, marked with an asterisk. The 4 meso protons can be clearly identified downfield at around 10 ppm. There are a number of peaks between 3.5 and 4 ppm, representing methyl groups. Non-singlet peaks between 2.5 and 8.5 ppm represent protons with coupling to other nearby protons.

4.5.1.2 Selective NOE Experiments

To enable assignment of all peaks in the one-dimensional spectrum, a full set of selective NOE experiments were carried out on the sample, as in 4.4. These spectra produced signals for protons which were close in space therefore the proton signals could be mapped to the structure on the basis of signal from neighbouring atoms. All proton peaks in the spectrum could be confidently assigned to their respective atoms in the MgPME structure and these are detailed in Table 4.2, along with the atoms which gave NOEs to each peak.

Proton	δ (ppm)	Protons with NOE correlations
2^1	3.80	20
3^1 (H_X)	8.48	5
3^2 (H_B, *trans*)	6.40	3^2 (H_A)
3^2 (H_A, *cis*)	6.14	3^2 (H_B)
5	10.32	3^1, 7^1
7^1	3.79	5
8^1 (H_X)	8.48	10
8^2 (H_B, *trans*)	6.40	8^2 (H_A)
8^2 (H_A, *cis*)	6.14	8^2 (H_B)
10	10.24	8^1, 12^1
12^1	3.71	10, 13^1
13^1	4.48	12^1, 15
13^2	3.38[a]	-
13^4	3.66	-
15	10.17	13^1, 17^1
17^1	4.42	15, 17^2, 18^1
17^2	3.45	17^1
18^1	3.71	17^1, 20
20	10.20	2^1, 18^1

Table 4.2 ^{1}H-NMR assignments for MgPME, with NOE correlations. Atoms numbered by the IUPAC system.
[a] = proton partially masked by methanol signal in 1D spectrum.

4.5.2 One-Dimensional A433 Spectrum

Purified pigment was prepared as in 4.4.1, with the purity of the pigment again seen to be reasonably good. There are only minor contaminating signals downfield of 5 ppm in the ^{1}H spectrum (Fig 4.10B), with most of the extra upfield signals coming from solvents. The most striking characteristic of this spectrum is the presence of additional signals, downfield from the cluster of meso proton peaks that were previously observed in the MgPME and Pchlide samples. The three proton peaks just above 10 ppm are as observed for previous meso proton peaks; however there are 2 additional peaks at 10.86 and 11.60 ppm.

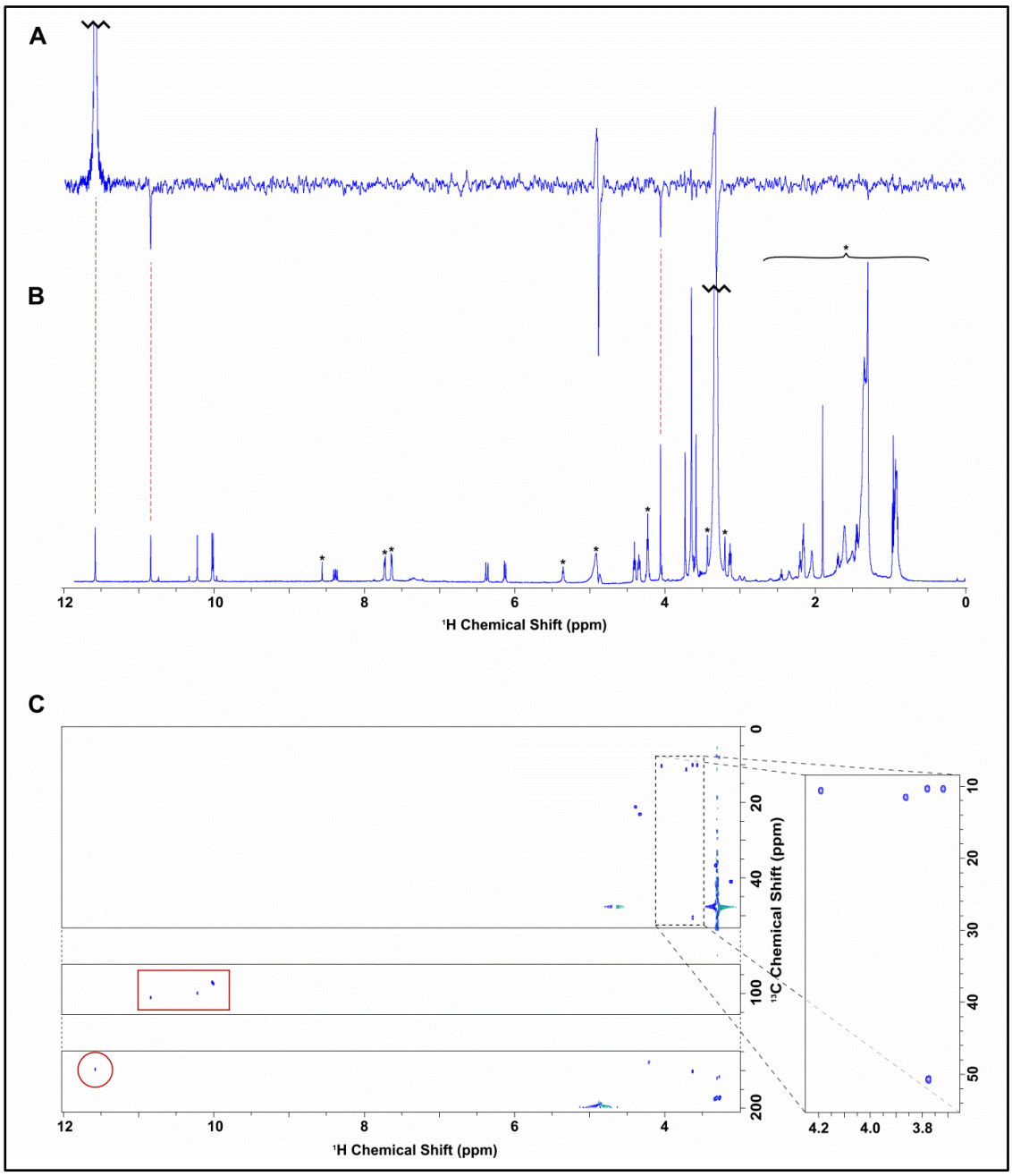

Figure 4.10 NMR assignment of A433. (A) One dimensional selective NOE spectrum of a433, saturating at the 11.60 ppm proton peak. Negative signals indicate NOE correlations with protons 5 and 2^1 at 10.86 and 4.07 ppm, respectively. (B) ^{1}H-NMR spectrum of a433. Signals marked with an asterisk are either solvent signals (methanol, water) or impurities (e.g. column matrix). (C) ^{13}C-HSQC spectrum for a433. Carbon axis is split for clarity. Dashed box indicates methyl signals, expanded on the right. Red box indicates signals from meso protons. Red circle indicates 3^1 aldehyde signal.

4.5.3 13C-HSQC, TOCSY and Selective NOE experiments on A433 pigment

Signals were assigned using a combination of ^{1}H TOCSY, gradient-selected 1D NOE, and natural abundance ^{13}C HSQC spectra. NOE experiments were carried out with selective saturation of all proton peaks downfield from 3 ppm in order to identify protons with through-space correlation (Figure 4.10A). Two ^{13}C HSQC spectra were run, with ^{13}C offsets of 60 and 140 ppm, in order to cover the full range of ^{13}C shifts. Many of the signals have ^{1}H and

^{13}C chemical shifts similar in frequency to those from MgPME, with the expected TOCSY and NOE connectivities, and can therefore be assigned straightforwardly. The ^{13}C HSQC spectra (Fig.4.10C) confirmed the assignments of the four meso protons, including the upfield-shifted peak at 10.86 ppm, and the five methyl groups; with the four porphyrin methyls having ^{13}C shifts of around 10 ppm and the propionate methyl having a shift of 50 ppm.

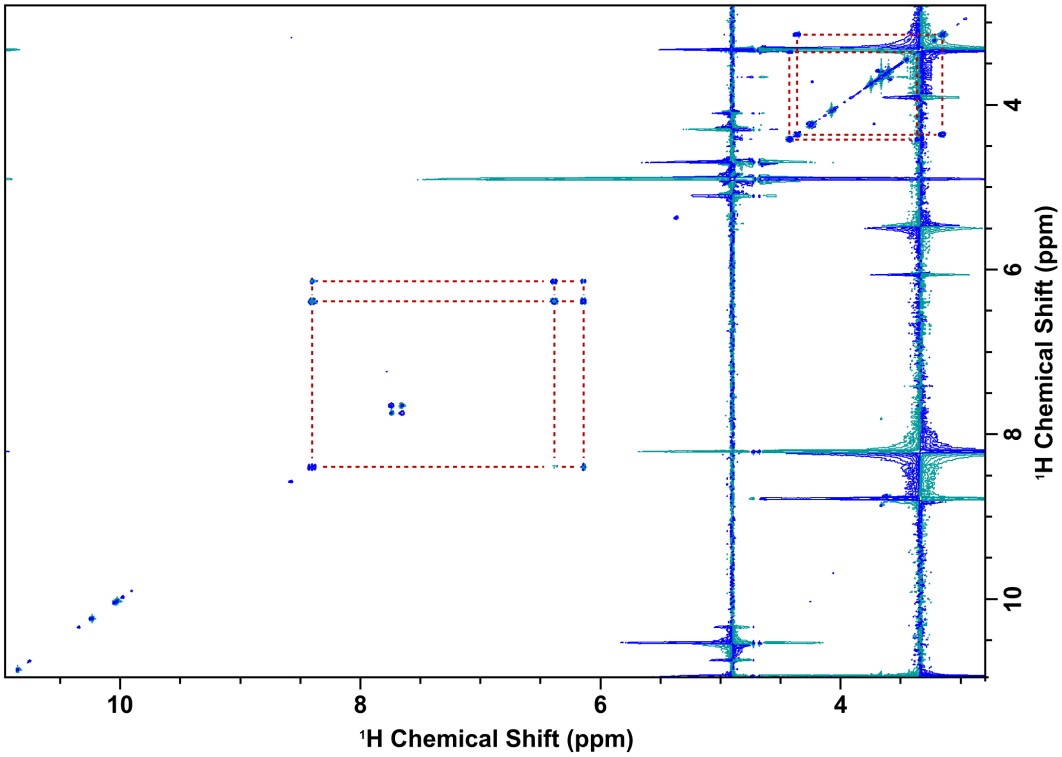

Figure 4.11 TOCSY spectrum of a433 pigment. The large signals at 4.8 and 3.3 ppm correspond to water and methanol, respectively. Coupled peaks are highlighted by adjoining red dotted lines. There are only three coupling systems within the molecules: the overlapped vinyl proton signals, the 17^1 and 17^2 protons and the 13^1 and 13^2 protons.

The signals from the 3-vinyl protons were absent, though the upfield signal with a ^{1}H shift of 11.6 ppm was found to have an unusual ^{13}C shift of 190 ppm, which strongly suggests that it is an aldehyde. This signal has NOEs to both the 5-meso and 2-methyl protons, both of which were shifted downfield, with no through-bond connectivity in the TOCSY, confirming that this was a 3-formyl group which had replaced the 3-vinyl group of the MgPME. The NMR spectrum is thus fully consistent with that of 3-formyl MgPME, the structure of which is indicated in Figure 4.12. Table 4.3 contains the structure assignments, along with their respective NOE and TOCSY connectivities.

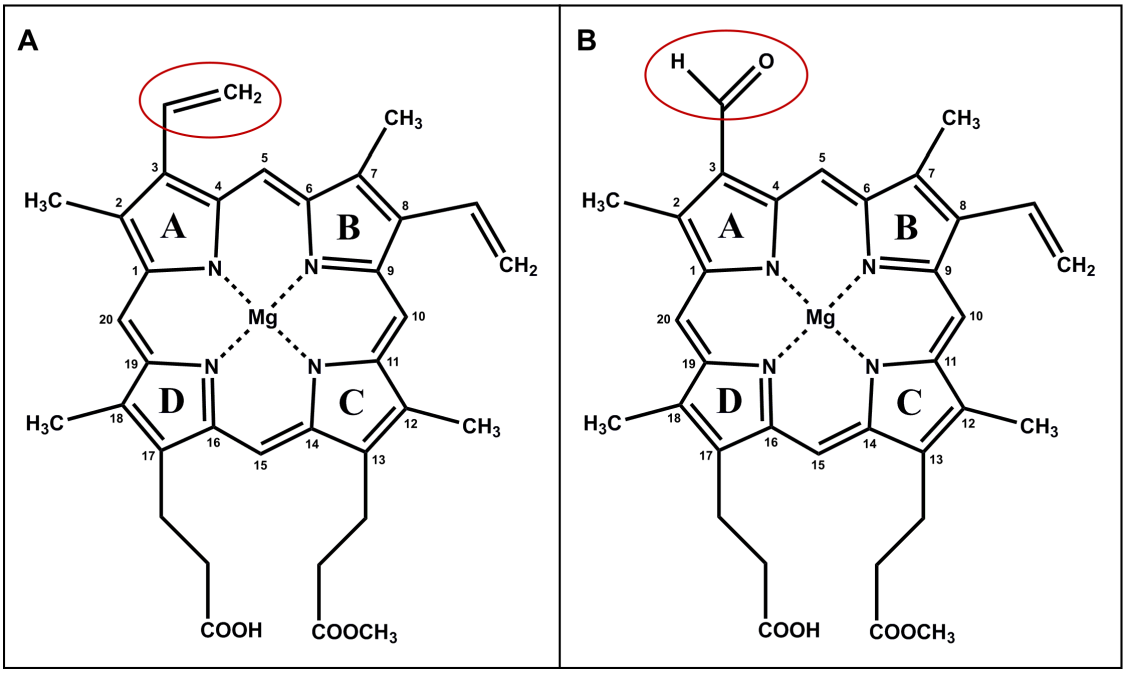

Figure 4.12 Molecular structures of (A) MgPME and (B) a433. Porphyrin carbons are numbered by the IUPAC system. Red circle indicates the vinyl group, in MgPME, converted to a formyl group in a433.

Proton	δ (ppm)	Protons with NOE correlations	Protons with TOCSY correlations
2^1	4.07	20, 3^1	-
3^1	11.60	2^1, 5	-
5	10.86	3^1, 7^1	-
7^1	3.74	5	-
8^1 (H_X)	8.40	8^2 (A,B), 10	8^2 (A,B)
8^2 (H_B, *trans*)	6.14	8^1, 8^2 (A)	8^1, 8^2 (A)
8^2 (H_A, *cis*)	6.38	7^1, 8^2 (B)	8^1, 8^2 (B)
10	10.04	8^1, 12^1	-
12^1	3.60	10	-
13^1	4.42	12^1, 15	13^2
13^2	3.32[a]	-	13^1
13^4	3.66[b]	-	-
15	10.03	13^1, 17^1	-
17^1	4.36	15, 17^2	17^2
17^2	3.15	17^1	17^1
18^1	3.66[b]	17^1, 20	-
20	10.24	2^1, 18^1	-

Table 4.3 ^{1}H-NMR assignments with NOE and TOCSY correlations for a433. [a] = signal masked by methanol peak in 1D spectrum but visible in TOCSY spectrum. [b] = 13^4 and 18^1 protons overlapped in 1D spectrum.

4.5 Discussion

The production of pure, high quality pigment is essential for the formation of POR complexes that can be considered a reasonable representation of their *in vivo* state. Any contaminants within the pigment sample would exacerbate the likelihood of artefacts in any experiments conducted on the POR complex. In order to analyse the purity of the current pigment preparation, samples were prepared for NMR and analysed by one-dimensional ^{1}H spectra. These spectra indicated that the level of purity of pigment was very low, with a massive number of peaks present in the sample, including a number of different peaks representing various porphyrin structures.

The presence of these contaminants would be likely to have a negative effect on the quality of complexes formed, therefore the purification protocol was altered to improve this, while still retaining high levels of pigment; approximately 2 mg of pigment could be produced per litre of growth medium, compared to 3 - 4 mg in the previous protocol. One notable problem in the use of *R. capsulatus* was the excretion of large amounts of lipid into the media, therefore the bacterial strain used for production of Pchlide was changed to *R. sphaeroides*. This strain excreted very little in the way of lipids and therefore pigment could be easily extracted directly from the media by the use of solvents. This enabled the exclusion of the polyurethane foam bungs step thus preventing the possible leaching of bung polymers into the preparation.

With direct solvent extraction from the media chosen as the preferred method of pigment extraction, the optimum solvents, and ratios thereof, were subsequently optimised. The resultant pigment sample was run on a C18 column on HPLC and found to contain mainly Pchlide; however there were a large number of minor contaminants, in addition to multiple Pchlide peaks. The two major Pchlide HPLC peaks were isolated and analysed by NMR, each giving surprisingly distinctive spectra, considering the homogeneous nature of their absorbance profiles.

Peak 1 from the HPLC was considered to be monodispersed Pchlide and, from the analysis of the NMR spectra, this was confirmed to be DV Pchlide. The second peak from the HPLC trace produced a polydispersed NMR spectrum, appearing to contain a number of different porphyrin molecules. This is in contrast with the analysis of the peak via HPLC, as the peak was very narrow and distinct from any other peaks, suggesting it represented a single molecule. The absorbance spectrum for this peak was as expected for Pchlide indicating that any alternative pigments, which would have a distinctly different absorbance profile, were not present in any significant quantities. It is thus most likely that this sample represents a

homo-oligomer of Pchlide, quite possibly a dimer with the porphyrin rings stacked against each other. For this reason, only peak 1 from the HPLC purification was taken for use in subsequent experiments, to ensure a homogeneous sample of monomeric Pchlide; this is essential for subsequent structural studies, to ensure efficient binding of the intended substrate, reducing the likelihood of artefacts.

Knockout of the Ycf54 protein, associated with the MgPME cyclase in *Synechocystis*, produced an unknown pigment with a λmax between that of the substrate, MgPME, and product, Pchlide, of the Mg cyclase reaction. The use of a variety of NMR experiments, carried out on each of MgPME, Pchlide and the mystery pigment, enabled a comprehensive structural assignment of the unknown pigment. The unusually large downfield shift of the aldehyde group of the pigment, helped to indicate the oxidation of the 3-vinyl group of MgPME to form the magnesium 3-formyl protoporphyrin IX monomethyl ester (3-formyl MgPME) molecule.

Due to the nature of the alteration of the MgPME molecule, with the oxidation reaction occurring far from the region of the molecule upon which Mg cyclase acts, this pigment cannot be considered as an intermediate of the cyclase reaction. Due to the large increase in accumulated MgPME in this mutant, it is likely that some of this substrate was poorly contained and was therefore exposed to oxidation away from the usual pathway. Whether this has any link to the production of chlorophyll *b* in the cell is unknown. It is clear, however, that the Ycf54 protein clearly has an important function with respect to the efficient turnover of MgPME to Pchlide and therefore must be considered as a subunit of MgPME cyclase; though the lack of apparent redox or electron transfer sites suggest it is unlikely to be a catalytic subunit.

Chapter 5

5. Formation and Macromolecular Structure of POR Ternary Complex

5.1 Introduction

The structure of POR cannot be truly understood without consideration of its organisation upon binding its substrates, NADPH and Pchlide. Two isoforms of POR in plants, PORA and PORB, have been shown to be major components of prolamellar bodies (PLBs), the large aggregated structures observed in dark-grown etioplasts (Griffiths, 1978; Lebedev et al., 1995; Oliver and Griffiths, 1982). It has also been shown that the expression of cyanobacterial POR proteins, from *Gloeobacter violaceus* and *Synechocystis sp.*, is able to restore the structure of PORA deficient prolamellar bodies (Masuda et al., 2009). This implies that cyanobacterial POR isoforms also have an innate ability to form similar structures.

Due to the large aggregated complexes formed upon POR substrate binding, it is difficult to study the POR ternary complex at an atomic structural level; techniques such as NMR and X-ray crystallography are not applicable to these aggregates. However, there is still much scope to study the arrangement of these complexes at a macromolecular level, using a number of different techniques. Firstly, the preparation of ternary complex required optimising to produce high quality ternary complex samples for experimental analysis. This, initially, enabled the examination of aggregates of POR ternary complexes; further refinement of the preparation subsequently enabled the identification of more detailed structural arrangements.

In this chapter, 'ternary complex' normally refers to the complex formed between POR, $NADP^+$, and Pchlide; though the true Michaelis ternary complex would be a complex between POR, NADPH and Pchlide. The NADPH complex is stable in the dark, however, with many experimental techniques, it is not possible to ensure complete darkness at all times; therefore the presence of $NADP^+$ rather than NADPH, prevents turnover of the complex in these instances. $NADP^+$ binds tightly to POR and is easily water soluble, therefore addition of significant excess will ensure it is bound; NADPH binds even more tightly, and so can easily displace $NADP^+$, in order to form the reactive complex.

5.2 Optimising protein purification of POR

Although a purification protocol had been established for POR previously (Proudfoot, 2011), this protocol was altered slightly in order to improve protein purification and quality. Due to problems with cleavage of the protein during the purification protocol, especially during the nickel affinity column where EDTA cannot be present, it was decided to develop an ammonium sulphate purification step to attempt to bypass the nickel affinity chromatography column. The previous protocol has an ammonium sulphate addition step subsequent to cell breakage, adding 80 % ammonium sulphate in order to precipitate the maximum protein possible. It was thought that if a lower concentration of ammonium sulphate was added at this point then the purity of POR could be increased while still keeping a high protein yield.

5.2.1 Ammonium sulphate precipitation

WT POR cells were grown up (detailed 2.7), the cells broken (detailed 2.9) and the supernatant retained and kept on ice, with an aliquot of this taken for later analysis. Ammonium sulphate was added to supernatant gradually, to a concentration of 20 %, continuously mixed for 15 min at 4 °C and centrifuged at 18,500 rpm for 15 min at 4 °C. The protein pellet was retained and a 10 µl sample of the supernatant taken, with the remainder of the supernatant retained. The volume of the remaining supernatant was checked and ammonium sulphate added gradually on ice to give a final concentration of 30 %. The steps above were then repeated and the whole process repeated for 40 %, 50 % and 60 % samples. The protein pellets were resuspended in 20 ml Low Salt buffer and 10 µl samples taken of each to run on an SDS-PAGE (detailed 2.17) alongside the supernatant samples (Fig. 5.1).

The SDS PAGE gel indicates the relatively high expression of POR in the supernatant prior to ammonium sulphate addition. At 20 % ammonium sulphate concentration only a small amount of POR precipitates, though it appears relatively pure, however in the supernatant there is clearly still a significant amount of POR remaining in solution. At 30 %, more POR is precipitated than at 20 % with purity higher than seen in the original expression; however there is still a relatively significant amount of POR remaining in the supernatant. The precipitant in the 40 % sample is similar to 30 %; however the supernatant in this sample appears to have minimal amounts of POR remaining. The precipitant from the 50 % sample only has a relatively small amount of POR in addition to a larger number of contaminating proteins, with the 60 % sample giving precipitate with only trace amounts of POR.

These results suggest that addition of ammonium sulphate could be used as a further purification step, without reducing yield of POR significantly. It is clear from the SDS PAGE that up to 40 % ammonium sulphate, most of the POR is precipitated with less contaminants than in the original sample, while above 40 % ammonium sulphate there is little POR precipitated and a number of contaminants. This indicates that a protocol with the addition of 40 % ammonium sulphate could improve purity of POR at this step, though not anywhere near as effectively as the nickel affinity chromatography. It was therefore decided that the nickel affinity purification step would definitely be required in order to give the level of purity required for NMR and other techniques. Thus, it was decided not to alter the ammonium sulphate precipitation protocol, retaining the 80 % ammonium sulphate addition step, in order to give the maximum protein possible.

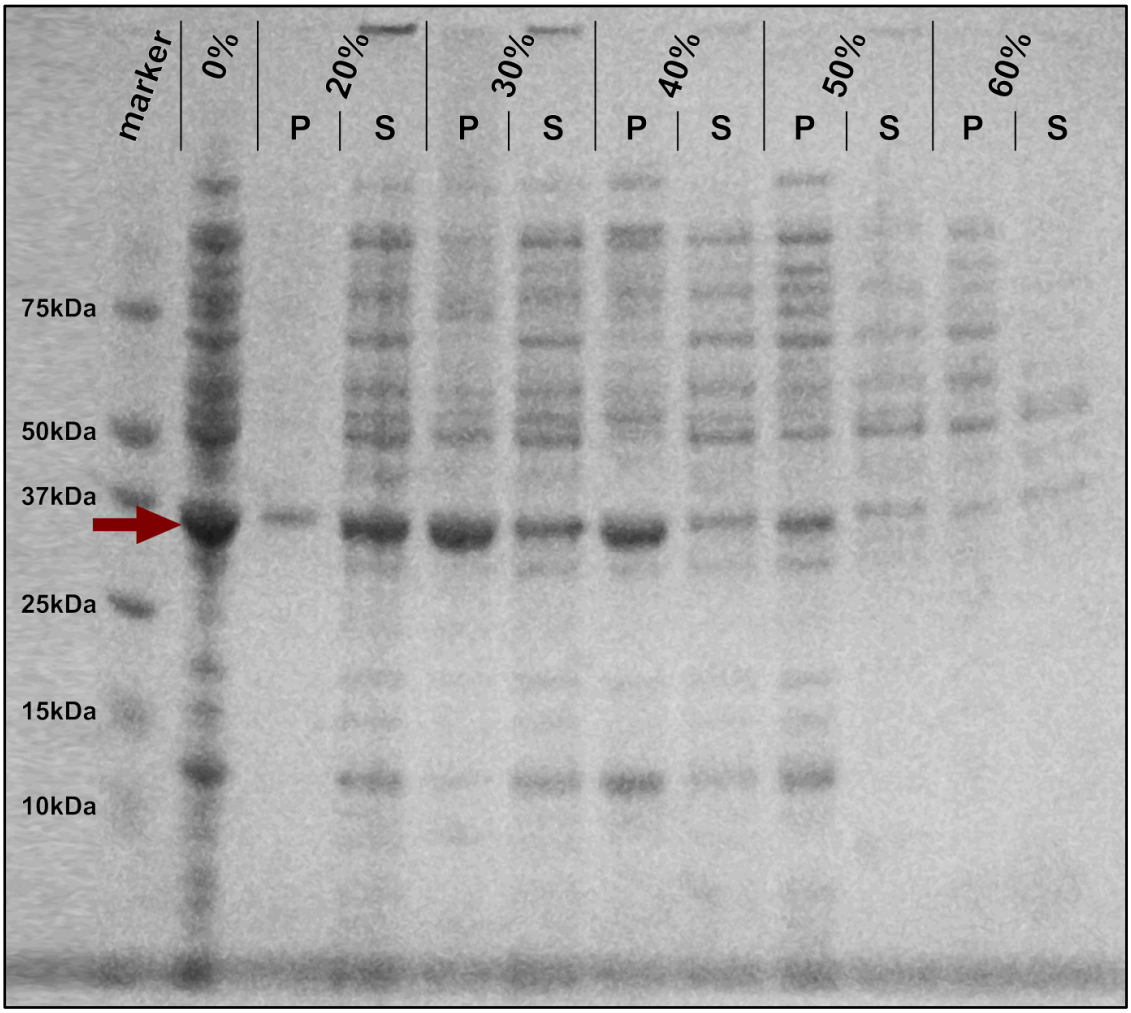

Figure 5.1 SDS PAGE gel showing protein levels in various ammonium sulphate additions. Marker lane is at the far left with marker protein masses labelled on the left. The second lane is the protein expression prior to ammonium sulphate addition. Each set of 2 lanes thereafter includes the pellet and supernatant after centrifugation of each ammonium sulphate addition, for 20 %, 30 %, 40 %, 50 % and 60 % ammonium sulphate. The position of POR on the gel is indicated by a red arrow.

5.2.2 Nickel affinity chromatography

The optimisation of the POR purification was, ultimately, to produce the best possible sample of deuterated protein for NMR triple resonance experiments. Expression of protein grown in deuterated medium was expected to be low therefore the purification protocol was required to be optimised for low concentrations of POR. The nickel affinity resin chosen was Protino Ni-TED resin, selected due to its high purification levels of polyhistidine tagged proteins and high specificity. The protein was purified under gravity (detailed 2.10.3), using LEW Binding buffer, LEW Wash buffer and LEW Elution buffer (detailed A3.1 – 3.3). Protein was eluted directly into sample tubes containing DTT (to final concentration of 1 mM) and Roche complete protease inhibitor.

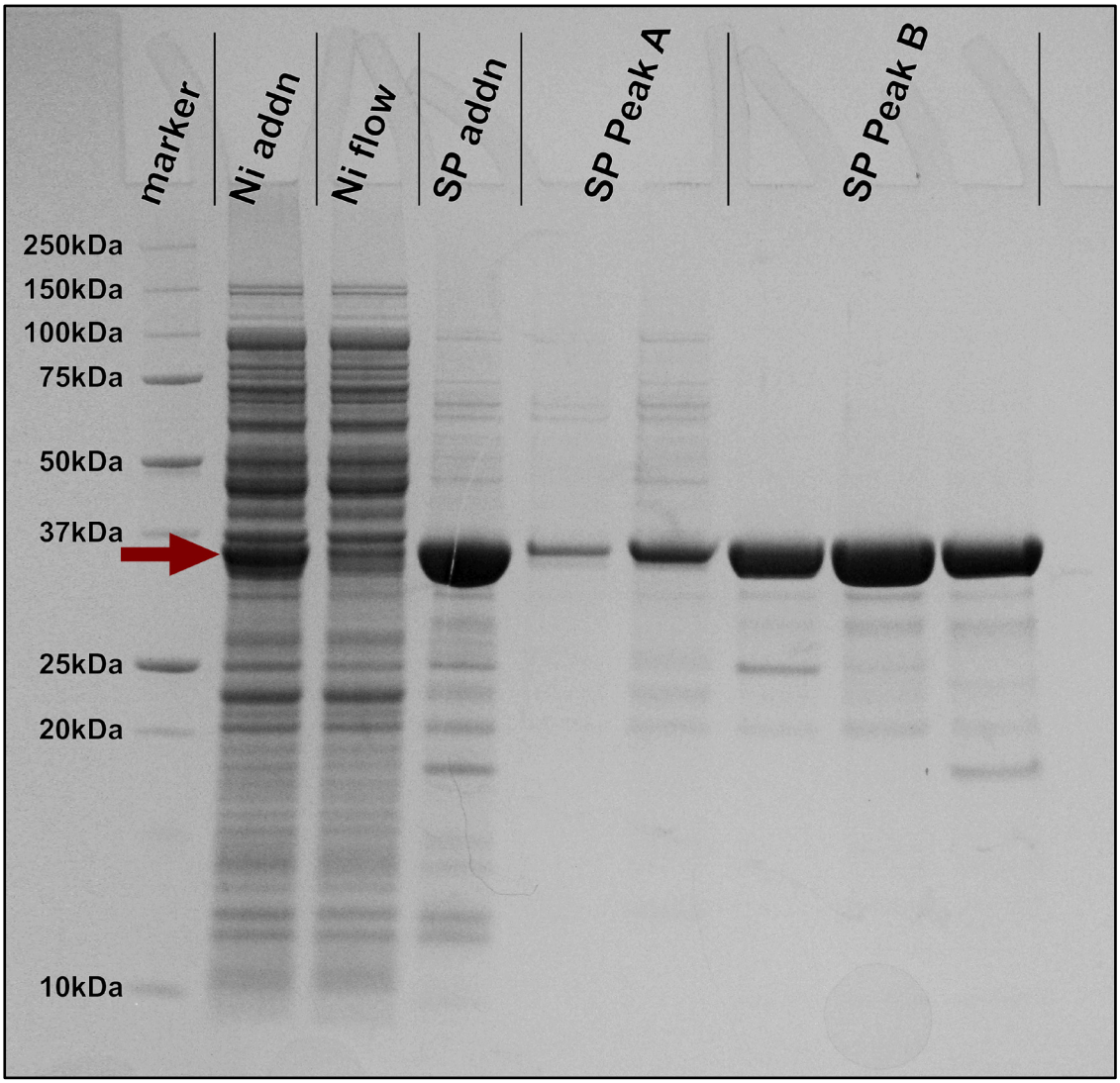

Figure 5.2 SDS PAGE gel monitoring levels of POR during the purification. Marker lane is at the far left with marker protein masses labelled on the left. POR expression prior to nickel affinity purification is in lane 2. Protein that did not bind the nickel column is in lane 3. Nickel column elution, subsequently loaded to the SP sepharose column, is in lane 4. Lanes 5 - 9 contain protein samples taken from the SP sepharose column, with lanes 5 and 6 from peak A and lanes 7 - 9 from peak B. The sample eluted from the nickel column clearly contains highly pure POR, with very little protein lost in the flow-through. The SP sepharose column removes a number of further contaminants, in lanes 5 and 6, giving a final sample with even higher purity.

Samples from the nickel column addition, flow through and elution were run on SDS-PAGE (Fig. 5.1) in lanes 2, 3 and 4 respectively. The addition in lane 2 shows that the expression of POR was good, with it clearly being the most expressed protein with the largest band in the gel. In the flow through sample in lane 3, very little, if any, POR can be seen indicating that it had bound very efficiently to the resin. In the elution sample, which is also the sample that was loaded to the SP sepharose column, there is a very large band for POR with only very weak bands of contaminants therefore the purity after the nickel column was good.

5.2.3 Concentration of POR

After elution, the samples containing protein were pooled, transferred to a spin concentrator and centrifuged at 3,000 rpm at 4 °C until reaching the desired volume. Previously, the use of a spin concentrator had been replaced with an Amicon stirred ultrafiltration cell due to the issue of protein precipitation at the bottom. However it was found that if the concentration step was performed immediately after elution, with the relatively high salt concentration from the buffer, then the protein would remain in solution throughout the concentration process. This was the preferred method of concentration due to the increased control possible, with the volume of the sample significantly more accurate to determine.

5.2.4 Anion Exchange Chromatography

In the original protocol, the anion exchange chromatography step was carried out using a SP sepharose column, run using Tris pH 7.4 buffers. Due to the requirement for the protein to be in phosphate buffer for NMR, this meant that the protein would have to be thoroughly buffer exchanged before the sample is ready for NMR. It was decided to change the buffer for the SP sepharose column to a phosphate pH 6.5 buffer for two reasons. Firstly, this pH is further from the theoretical pI of 9.4 for POR therefore the ionic association with the column is likely to be stronger; secondly, the pH is close to the pH required in NMR experiments, therefore very minimal adjustments must be made to the sample following elution from the SP sepharose column.

The SP sepharose column was run as detailed in 2.10.3, with the elution profile exhibited in Fig. 5.3. The UV absorbance at 280 nm (blue line) was monitored from the point of protein injection onto the column to follow the protein elution from the column. The pink line displays the change in salt concentration throughout the run, with the values displayed on the right hand axis. Protein that did not bind to the column was eluted at around 20 min

during the wash step, labelled Peak A. During the gradient step, one large, broad peak (Peak B) with a left hand shoulder was eluted. No excess protein was eluted during the high salt wash step, indicating that all protein that was bound the column had been eluted during the gradient.

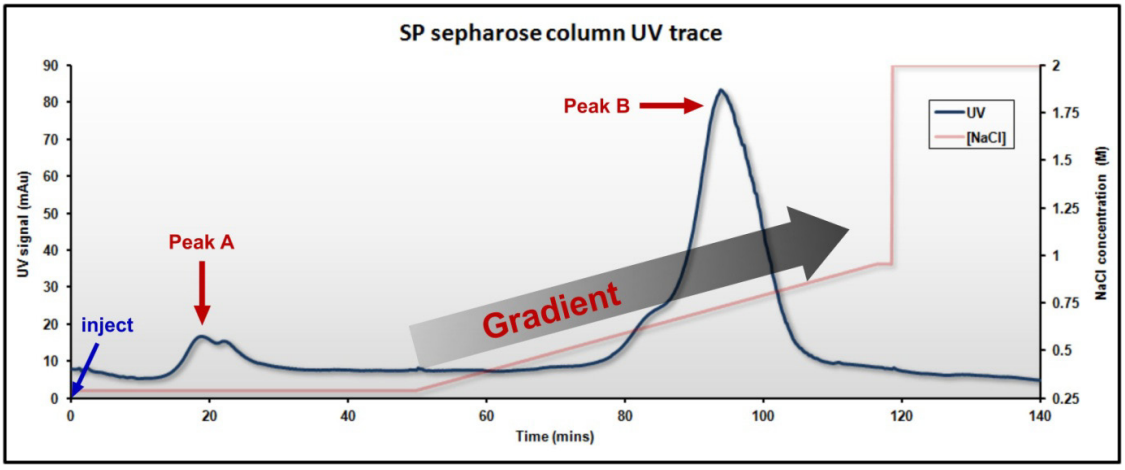

Figure 5.3 UV trace for SP sepharose purification. After protein injection, the column was washed for 50 min, followed by a salt gradient over 67 min and a final high salt wash for 20 min. UV trace is shown in blue, with the UV signal displayed on the left hand y-axis. Salt concentration is shown in pink, with the values displayed on the right hand y-axis. Peaks A and B are highlighted by red arrows.

Samples from both peaks were run on SDS-PAGE (Fig. 5.2) with samples from the front and end of peak A in lane 5 and 6 respectively, a sample from the peak B shoulder in lane 7 and samples from the front and back of peak B in lanes 8 and 9 respectively. Both samples from peak A contain POR, however at quite low concentration, while there are also a number of protein contaminants. All samples from Peak B, including the shoulder of the peak, contain a high concentration of POR and, though a few contaminants can be seen due to the overloading of the gel lanes, these are minimal compared to the levels of POR; therefore, all the samples from this peak were taken and pooled. The purity of POR after the anion exchange column can therefore be considered to be very good and of sufficient quality for NMR.

5.3 Ternary Complex Preparation using Solvent Solubilised Pchlide

5.3.1 Formation of Ternary Complex

A previously devised protocol for the production of POR complexes (Proudfoot, 2011), was used for formation of POR ternary complex in this study (detailed 2.17). Formation of complex initially involved the preparation of apo POR sample, in a buffer containing 20 mM

βOG detergent (to ensure efficient delivery of Pchlide to the protein); this was followed by the addition of 100-fold excess of $NADP^+$ to POR. Pchlide was solubilised in 100 % methanol and added to the POR, $NADP^+$ and detergent sample in a 2-fold excess, due to limited pigment concentration. The final preparation using this method contained a large amount of insoluble green-coloured material which was subsequently removed by centrifugation at 10,000 rpm for 5 min; this was expected to be insoluble aggregate of either Pchlide or highly aggregated ternary complex. The resulting sample was still substantially green in colour, however, indicating that although much of the Pchlide had been removed, there was still a significant amount remaining in the sample. It was not possible to quantify the ratio of substrate existing unbound and bound to POR; mainly due to the spectral shift of Pchlide peaks, using room temperature spectroscopy, being minimal upon binding to protein.

5.3.2 Purification on SP Sepharose Column

The inability to quantify the proportion of complex to apo protein and unbound substrate, meant that the homogeneity of protein in the complex could not be confirmed. In order to minimise the amount of unbound components in the sample, it required further purification. Previous attempts to separate and analyse the POR ternary complex by gel filtration had been unsuccessful due to the large hydrophobic aggregates of the complex blocking the resin. It was therefore decided to attempt to separate apo and complexed POR by charge using a SP sepharose column. This purification was identical to that of apo POR purification, with protein complex bound to the column and eluted specifically via a gradient of increasing salt concentration (detailed 2.10.3).

The UV trace for the SP sepharose purification is shown in Fig. 5.4, along with the corresponding NaCl concentration. The UV signal corresponds to protein rather than Pchlide, therefore the presence of Pchlide had to be determined by eye from the level of green colour observed in the elution. The majority of the material was eluted at the start of the run, suggesting that most of the protein was unable to bind the column. This peak was clearly green in colour, indicating that it also contained a large amount of Pchlide; though it could not be determined whether this substrate was bound to protein.

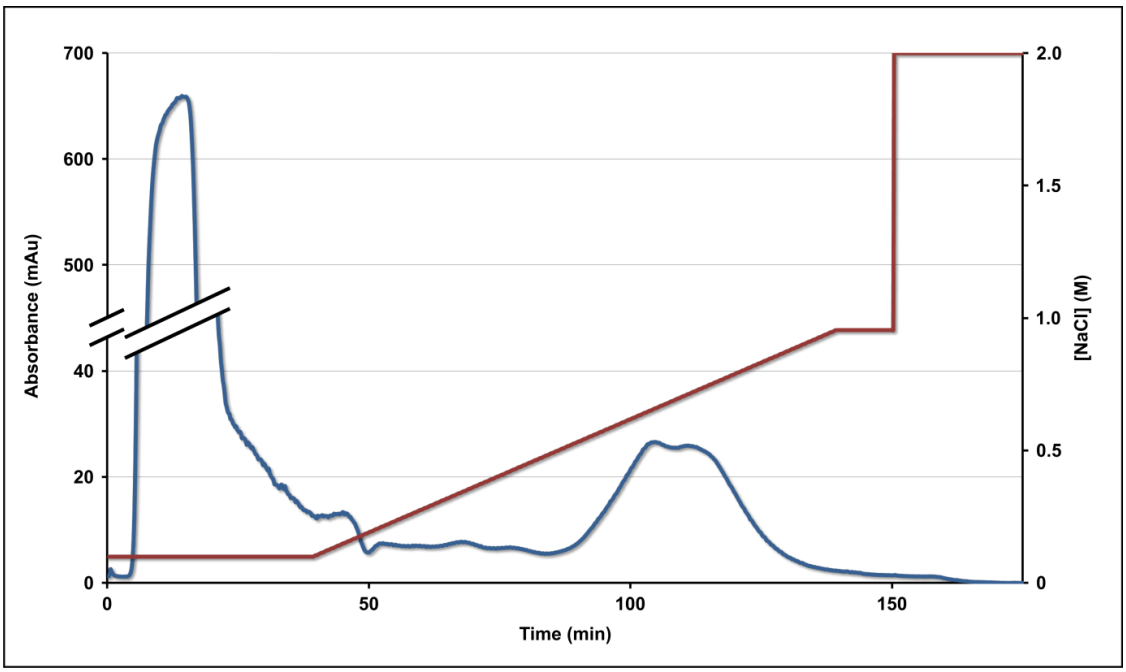

Figure 5.4 UV trace of ternary complex purification on SP sepharose column. UV absorbance, at 280 nm, is monitored by the blue line with values on the left-hand, split axis. The change in salt concentration is indicated with the red with values on the right-hand axis. A very large initial peak is evident prior to application of the salt gradient; this peak was green in colour. Two overlapped peaks can be observed at approximately 105 and 115 min, green and colourless in appearance, respectively.

During the salt gradient, two significantly overlapped peaks are observed, indicating protein elution from the column. Of these peaks, the first one was green whereas the second was clear; indicating that the first peak contains protein bound to pigment and the second peak contains only protein. Though these peaks are overlapped, the first half of the green peak should contain very little of the protein in the clear peak. It is likely that all of the unbound Pchlide is contained within the peak at the beginning of the trace; therefore the other green peak was taken to be the most pure sample of POR complex. The absorbance spectrum for this sample confirmed the presence of Pchlide; NADPH was subsequently added and the sample incubated at 55 °C and irradiated for 10 sec, with the resultant absorbance spectrum indicating turnover of the Pchlide to Chlide (Fig. 5.5). The sample was, therefore, confidently determined to contain active ternary complex, and thus used for all following studies.

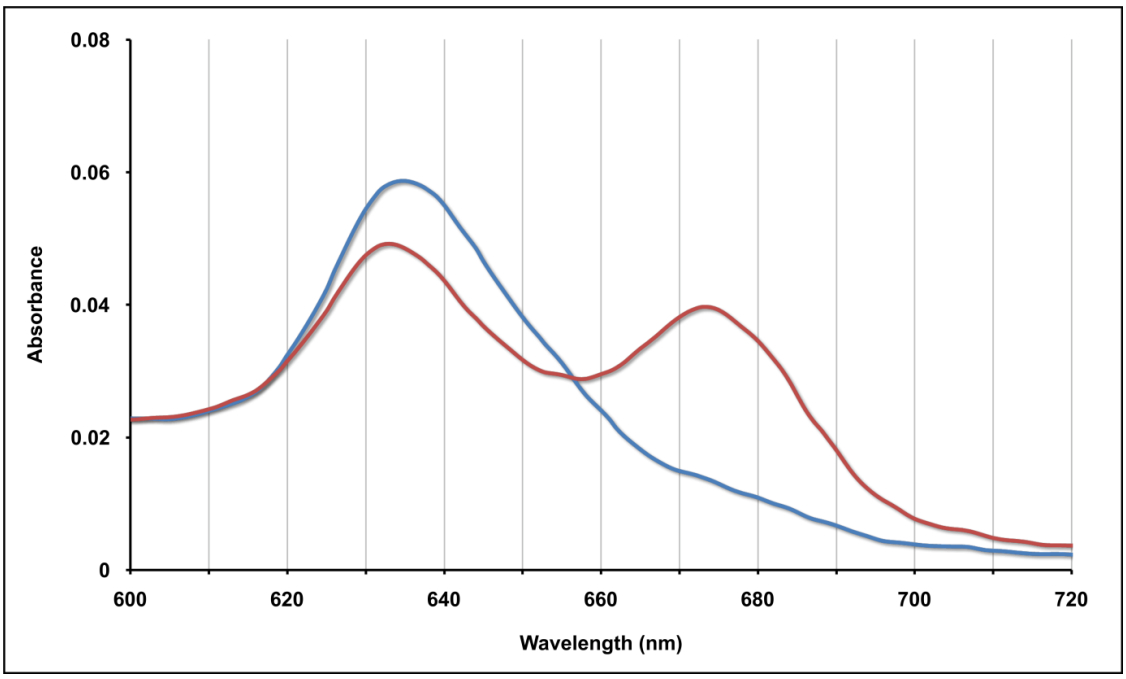

Figure 5.5 Absorbance spectra of purified POR ternary complex, before (blue) and after (red) irradiation. The increase in Chlide absorbance at 670 nm, following irradiation, indicates turnover of the complex.

5.3.3 Native Protein Gels

To analyse the size of the POR ternary complex samples under standard buffer conditions, a number of samples were run on native polyacrylamide gel electrophoresis (detailed 2.19). Molecules migrate through the gel at different rates depending on their size, as in SDS PAGE, however the samples are not subjected to denaturing conditions, and thus protein complexes remain intact. The gels were characterised by the appearance of smearing of protein samples at the top of the gel lanes, with only small amounts of discrete banding visible (Fig. 5.6A).

The migration of the marker bands was plotted against their molecular weight to enable estimation of the molecular weights of the discrete sample protein bands on the native gel (Fig. 5.6B). The estimated molecular weights do not match up well with exact polymers of POR, for example an estimated molecular weight of 99 kDa corresponds to 2.5 molecules of POR. This is likely to be due to a number of factors, including the nature of the sample buffer and the shape of the protein complexes. There is clearly no monomeric POR in the sample with the complex seeming to range from dimer/trimer size upwards. There also appears to be no regularity in complex sizes; aside from the discrete banding observed, there is no obvious trend in the way that larger complexes could be built up from the smaller complexes.

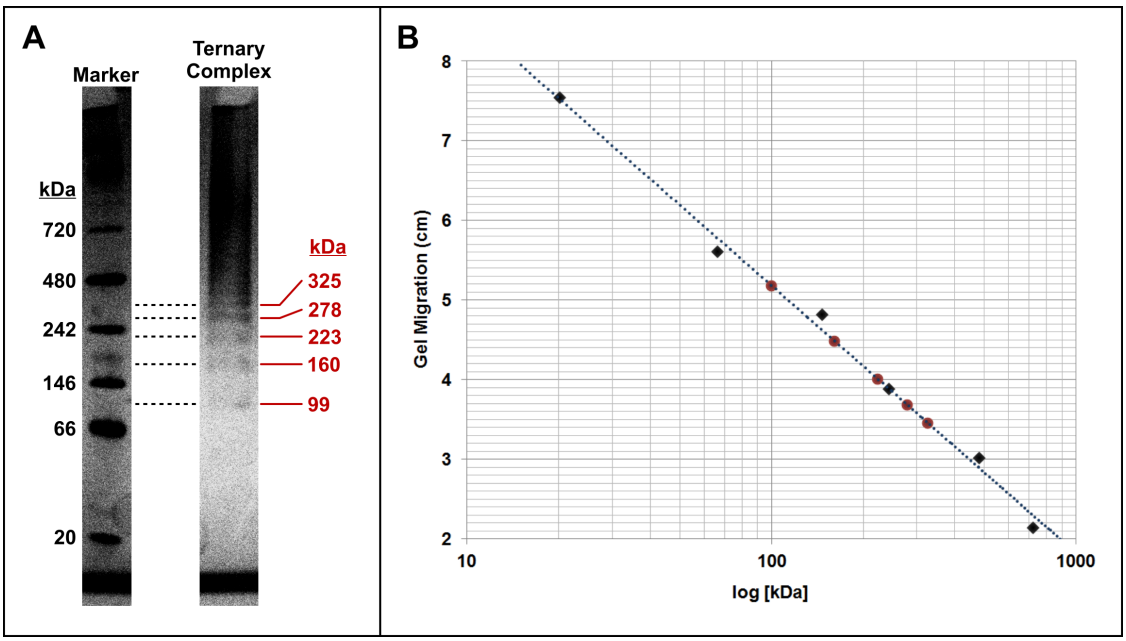

Figure 5.6 (A) Slices from native gel of POR showing marker, left, and ternary complex sample, right. Molecular weights of marker bands are shown on the left. Red lines indicate discrete bands in the sample, with predicted molecular weights shown, also in red. (B) Graph plotting marker molecular weights alongside their gel migration distance (black diamonds), in order to calculate predicted weights for the sample lane (red dots).

5.3.4 Dynamic Light Scattering (DLS)

Dynamic Light Scattering (DLS) was identified as another technique to characterise the size distribution of POR complexes. This technique detects light scattered by a sample in order to determine the distribution of particles in the solution. POR samples were run, by staff scientists from Malvern Instruments Ltd., on a Zetasizer Nano with a green laser (detailed 2.27). Samples were run for both apo protein, ternary complex and product complex; with any changes in the sizes of the complexes monitored (Fig. 5.7). Apo POR gave a single peak at approximately 15 nm in the fitting of the light scattering data, which is significantly larger than the diameter of around 5.5 - 6 nm calculated for monomeric POR. Although it is expected that apo POR is predominantly monomeric, it is highly hydrophobic and is therefore expected to form a number of transient complexes. DLS is significantly more sensitive to higher molecular weight complexes, therefore light scattering from a small number of larger complexes could inundate the scattering from the smaller molecules; and therefore the size estimation for a predominantly monomeric sample may increase.

Analysis of the ternary complex sample indicated that the size of complexes drastically increases from those of the apo protein sample; with a peak representing particles of almost 1 µm in diameter. This does not however indicate the shape and, therefore, mass of the molecule; with calculations for the mass of a globular complex of this size expected to be 25 GDa (875,000 monomers), whereas a linear molecule be calculated to be 420 MDa (10,000

monomers). Considering the shape and structure of *T. elongatus* (Hsueh et al., 2007), with a diameter of approximately 1 µm and length of 5 - 10 µm, a more linear structure would be likely. As DLS is so strongly biased towards detection of larger particles, these are likely to be the upper estimates of the structures. This sample was subsequently irradiated and further light scattering data collected over intervals of approximately 5 min. After each subsequent experiment, the representative size of the peak exhibited had decreased; from 850 nm diameter in the non-irradiated complex, through 430 nm, 240 nm and finally to 200 nm, approximately 15 min after irradiation. These results indicate that the complexes form extremely large aggregates upon binding of substrates, which then break down upon the conversion of Pchlide to Chlide.

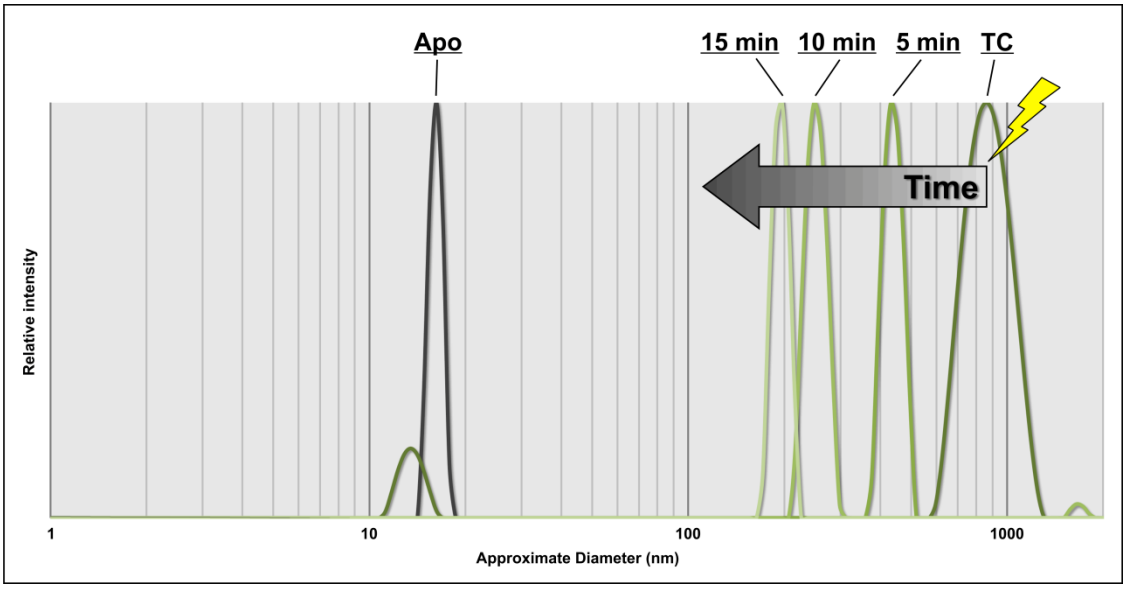

Figure 5.7 Graph of DLS data with peaks representing modeled particle sizes. Relative peak intensities are plotted against a logarithmic scale, normalised to the largest peak. Apo protein is shown in dark grey. Initial ternary complex is shown in dark green, with each of the three lighter green plots representing a 5 minute experiment, subsequent to irradiation.

5.3.5 Analytical Ultracentrifugation (AUC)

Another technique for analysing the sizes of complexes in a sample is Analytical Ultracentrifugation (AUC). This involves the centrifugation of samples at very high speeds, monitoring the sedimentation velocity of the components of the sample, in order to determine the size of all particles within the sample. Sedimentation can be monitored by absorption at different wavelengths as well as with interference (non-specific signal from all particles in the sample), enabling size information for various different components of the sample. By monitoring absorbance at 280 nm, protein sedimentation can be followed; while

sedimentation of components containing pigment can be monitored by absorbance at 440 nm.

The samples analysed were apo POR, ternary complex and irradiated ternary complex; these were prepared and taken to the NCMH at the University of Nottingham, with all AUC experiments and analysis of raw data conducted by Dr David Morou-Besong (detailed 2.23). Samples were run alongside buffer samples in order to obtain data for the background sedimentation of the buffer, to give more accurate sedimentation profiles for the samples. Samples were spun at 50,000 rpm overnight with sedimentation monitored throughout, with the resulting sedimentation profiles shown in Figure 5.5.

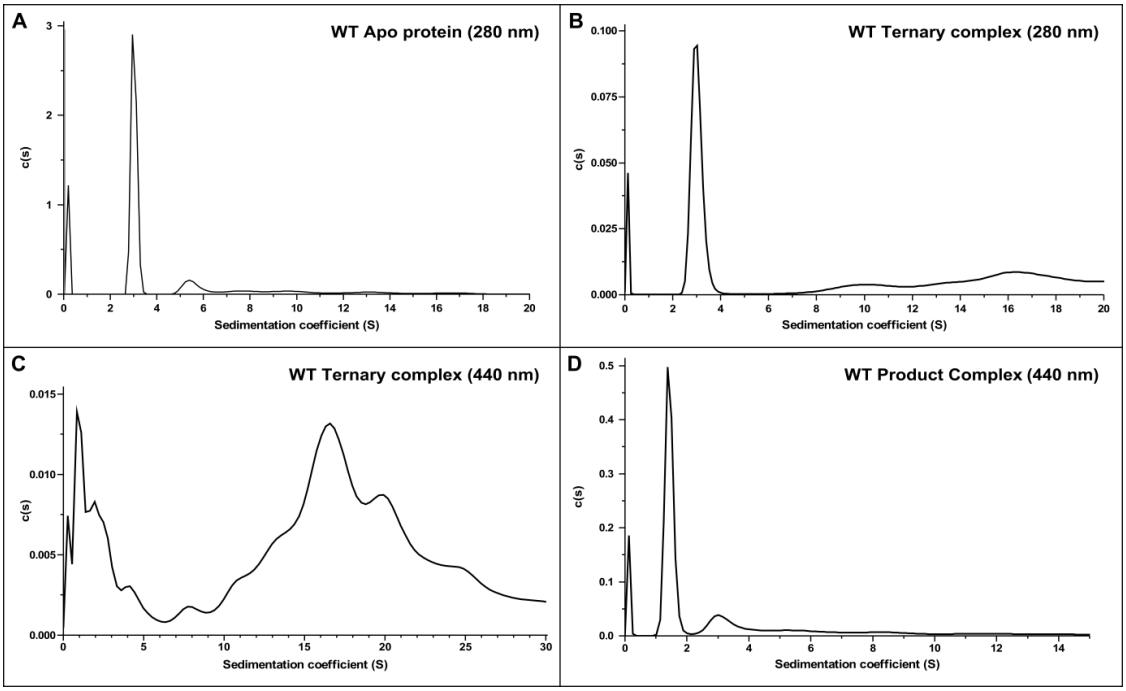

Figure 5.8 AUC data from WT POR samples of: (A) apo protein interference signal; ternary complex at (B) 280 nm and (C) 440 nm; (D) product complex at 440 nm. The sedimentation coefficient is related to molecule size, though also depends on differences in molecular shape. The c(s) value indicates the relative proportion of each component.

The sedimentation profile of the apo protein indicates one predominating peak, with a sedimentation coefficient of approximately 3 S; this is expected to represent monomeric POR as apo POR is known to be predominantly monomeric. There is a second, relatively significant peak at approximately 5 S, suggesting the presence of small amounts of dimeric apo POR; a 1.5 fold increase in sedimentation coefficient is indicative of a doubling in particle size, therefore these peaks are consistent with monomer and dimer. There are also trace amounts of larger complexes in the sample, though the levels of these peaks is insignificant compared to the monomer and dimer peaks.

Ternary complex POR was analysed at two absorbance wavelengths, 280 nm and 440 nm, corresponding to the absorbance of protein and Pchlide, respectively. The 280 nm spectrum indicates that there is a large amount of protein at 3 S, as in the apo protein sample; this almost certainly corresponds to excess unbound POR protein, as it is not present in the 440 nm spectrum, and this implies that there is a significant fraction of unbound POR present in this preparation; approximately 50 % of the sample as measured from the areas of the peaks. In both spectra, there is a large amount of signal at high sedimentation coefficient, thus pertaining to large aggregated material. The 440 nm spectrum indicates that this larger aggregated material is the predominant species in the sample and has no apparent regulation in size that would suggest preferential formation of certain oligomers; the aggregates observed go beyond the measured value of 30 S, suggesting the presence of extremely large, megadalton range species.

The irradiated ternary complex sample was also measured at 440 nm, this time to monitor the absorbance of Chlide in the product complex. In this sample, it is clear that the large aggregates, previously observed in the non-irradiated sample, have dissipated, with mainly smaller particles remaining. The absorbance spectrum is remarkably similar in appearance to that of the apo protein, however, having been monitored at the Chlide absorbance, it cannot correspond to unbound protein. Despite the visual similarity with the apo protein, the sedimentation values for the observed peaks are significantly lower than for the apo protein peaks. As the irradiated complex cannot have feasibly decreased from monomer size, it suggests that there may be a difference in particle shape or possibly buffer components to enable such a reduction in sedimentation value. It is however clear that POR forms large aggregates upon complex formation, which then dissipate upon conversion of Pchlide to Chlide.

5.4 Ternary Complex Preparation using Detergent Solubilised Pchlide

The method of ternary complex formation described above, though successfully producing ternary complex, did not result in a high proportion of substrate binding, as observed in 5.3.5; it also involved the loss of a large amount of the protein, which did not bind the SP sepharose resin in the purification step. It was thought that the addition of pigment to the sample in solvent may have caused the pigment to aggregate instantly, upon making contact with the hydrophilic environment of the water-based buffer. We therefore reasoned that, by solubilising the pigment in a water-based detergent solution, the environment of the porphyrin molecule will not be changed as dramatically upon addition to the lower detergent

concentration. This should enable improved pigment delivery to the protein and thus increase both the concentration and quality of the ternary complex samples.

5.4.1 Formation of Ternary Complex

POR ternary complex was prepared (detailed 2.17) using HPLC purified Pchlide substrate as produced in 4.3. Dried pigment was resolubilised in Solubilisation buffer (detailed A3.11) containing βOG detergent at 50 mM, with βOG diluted to its CMC of 20 mM in the final sample; in order to facilitate pigment availability for binding. The ternary complex sample produced using this method was centrifuged to remove any aggregates and exhibited only minimal amounts of insoluble green-coloured material. The resulting sample was not visibly less green than before centrifugation, indicating essentially 100% of the Pchlide was still present. As mentioned previously, it was not possible to quantify the ratio of substrate existing unbound and bound to POR. The sample was further diluted after binding, in order to reduced the detergent concentration further and prevent any removal of bound pigment from the protein.

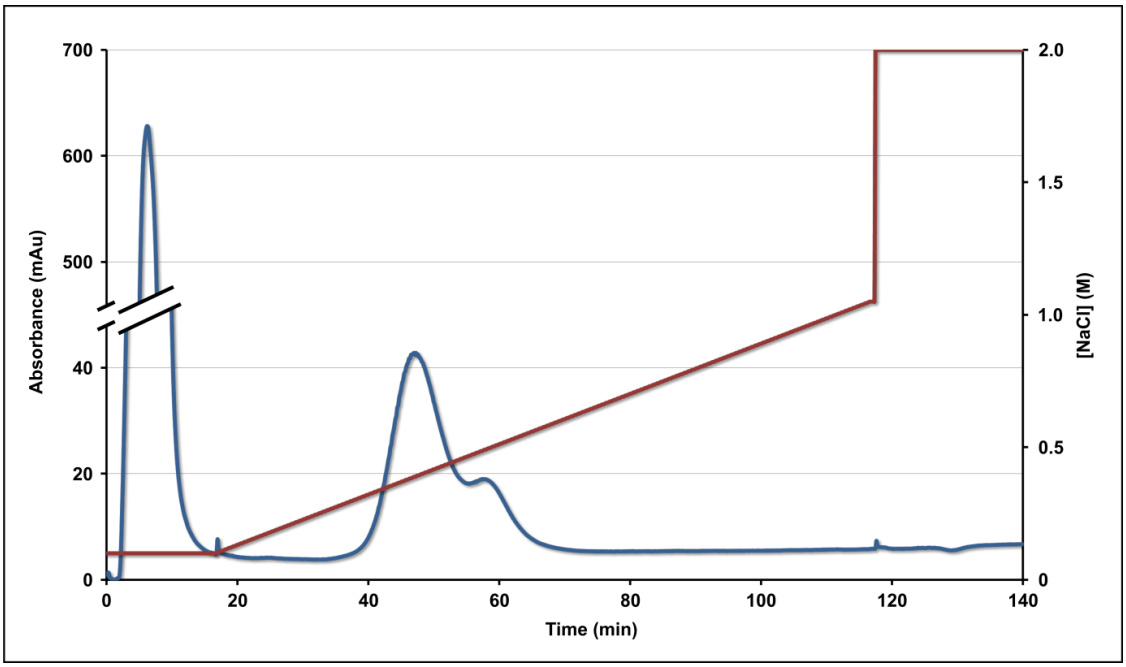

Figure 5.9 UV trace of ternary complex purification on SP sepharose column. UV absorbance during the run is monitored by the blue line with values on the left-hand, split axis. The change in salt concentration is indicated with the red with values on the right-hand axis. A very large initial peak is evident prior to application of the salt gradient; this peak was green in colour. A broad peak with a right-hand shouder can be observed at approximately 40 - 60 min; both peak and shoulder were colourless in appearance.

5.4.2 Purification on SP Sepharose Column

The ternary complex sample was purified on a SP sepharose resin by the same method as described in section 5.2 for the previous protocol. The purification is shown in Figure 5.6, exhibiting a similar trace to the one observed in the original preparation with a very large, green-coloured peak prior to the salt gradient. There were also two overlapped peaks at around 0.4 M NaCl in the salt gradient; though neither of these was green in colour, indicating that they both contain apo POR. This suggests that any protein correctly bound to substrates cannot bind the resin, and therefore that all POR ternary complex is contained within the first peak. This is in contrast to the previous protocol, where visibly green samples were shown to bind to the column, and were subsequently shown to contain active POR complex.

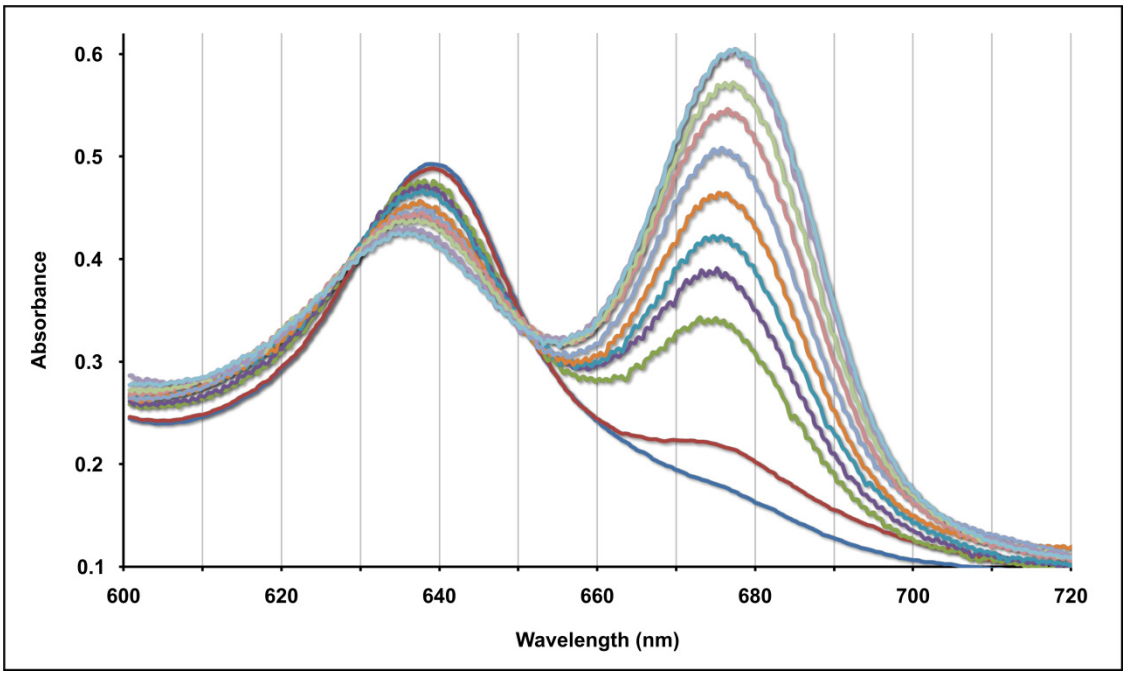

Figure 5.10 Absorbance spectra of purified POR ternary complex, indicating turnover of POR. The increase in Chlide absorbance at 670 nm, following irradiation, indicates turnover of the complex.

Binding of protein to the SP sepharose column is a property of the charge of the protein, therefore the initial peak, which does not bind the column, is likely to have a different surface charge to both apo POR and the ternary complex bound to the column in the previous protocol. Kinetic analysis of the 5 min peak in Figure 5.6 also showed that this contained active complex (Fig. 5.10), therefore active POR ternary complex has been shown to both bind and not bind the SP sepharose resin. It is most likely that, in the presence of βOG, the majority of protein complex is surrounded by detergent; as βOG molecules have a neutral

charge, this would give an overall neutral charge to the surface of the complex, preventing binding to the resin. The ternary complex bound to the resin in the previous protocol is, therefore, likely to represent a small amount ternary complex which is not sequestered by βOG molecules, and thus has a surface charge that facilitates binding to the resin. The green peak, eluted at 5 min on the SP sepharose column in this preparation, was retained as the ternary complex sample; as it is likely that much of the apo protein bound to the resin, and has therefore been removed, indicating that the first peak is likely to be relatively pure ternary complex. This improved preparation was subsequently studied using a range of biophysical methods, as described in the following sections.

5.4.5 Electron Microscopy (EM)

Electron microscopy was carried out by Dr Pu Qian (detailed 2.25), with images gathered for two separate samples; both prepared using the method described above. Despite being prepared identically, the two samples were quite distinct when observed via EM.

5.4.5.1 Aggregated Ternary Complex Sample

The first sample was prepared, as detailed in 5.3.1, and kept in darkness to prevent photoexcitation of the Pchlide. 20 µl of the sample was transferred to the EM grid, fixed and stained whilst remaining in darkness; this would represent unreacted ternary complex in the experiment. NADPH was added in 100-fold excess and the sample incubated at 55 °C for 2 min; the sample was then irradiated for 10 sec. Immediately after irradiation 20 µl of sample was transferred to the EM grid, fixed and stained; this fixing should prevent any subsequent breakdown of aggregates from this point. This was then repeated for samples at timepoints of 2, 4, 8, 16 and 60 min and all grids were imaged in the dark by transmission electron microscopy.

EM micrographs of each sample are shown in Figure 5.8. The POR complex sample kept in the dark can be clearly seen to exhibit significant amounts of aggregated material, though smaller particles are also observed in minor amounts (Fig. 5.11A). The sizes and shapes of the aggregates are not regular; aggregates measure up to 200 nm across and appear to involve random association of smaller components, to form massive aggregate structures. In addition to the larger scale aggregation, there are lesser amounts of small complexes, less than 10 nm in diameter; again there appears to be no order in the structure of these complexes, though the resolution is not high enough to be certain of this.

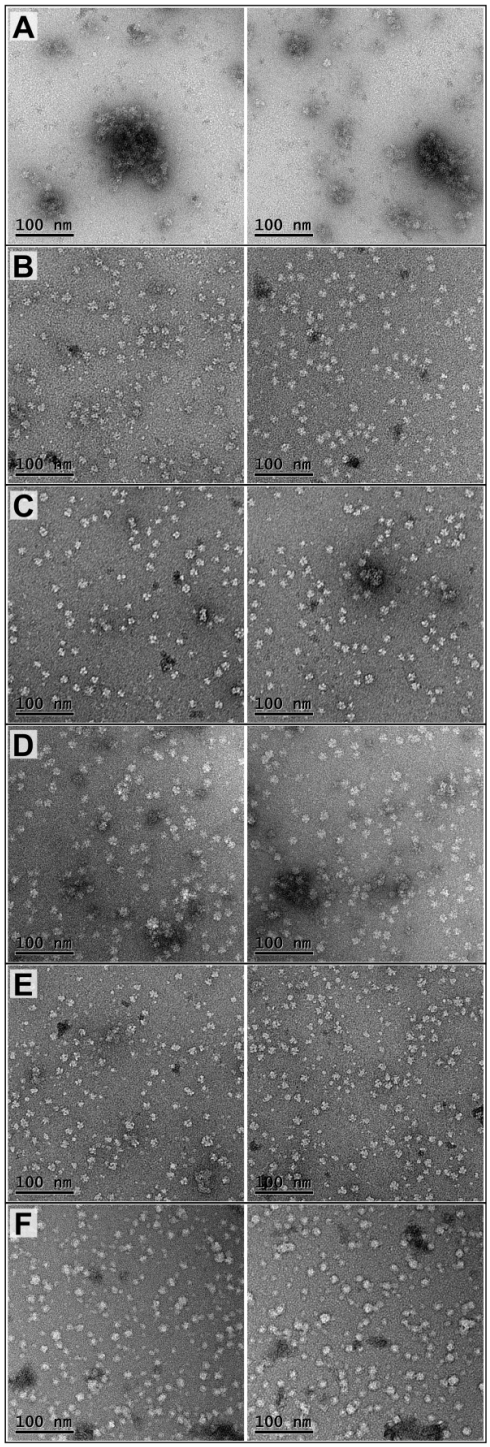

Figure 5.11 Negative stain EM micrographs of POR ternary complex at various time points with respect to irradiation, with two fields of view displayed for each: (A) Dark complex; (B) 0 min; (C) 4 min; (D) 8 min; (E) 16 min; (F) 1 hour. Scale bars placed at bottom left corner of each image. Prior to irradiation (A), very large aggregates (> 100 nm) are visible, along with small amounts of apparent ring structures and monomer. Upon irradiation (B), few of the very large aggregates remain, with various ring structures of 4 - 10 monomers predominating. After incubation for longer time periods (C - F), more apparently monomeric protein is visible, though the ring structures still remain.

Subsequent to irradiation of the complex, the large aggregates observed in the dark sample are immediately broken down (Fig. 5.11B-F). This must have occurred on a timescale less than 2 min; once the sample is fixed on the EM grid it should not undergo any further

physical changes, with the fixation process taking no more than 2 min. This breakdown of the large scale aggregates leads to the observation of predominantly smaller complexes, observed in minor amounts in the dark sample. Again there appears to be little regularity in these complexes; these species range in diameter from 8 to 12 nm with some appearing in ring-type structures while others are roughly globular.

Despite the breakdown of the majority of the larger aggregates, they are not completely eradicated, with some mid-sized aggregates observed in irradiated samples. This does not appear to be a time dependent phenomenon with aggregates noted in some later experiments (Fig. 5.11C right hand image) while not in the initial irradiated sample; these could be representative of ternary complex which had not turned over. It is also apparent that, after longer time periods subsequent to irradiation, a number of apparently monomeric species are observed in the images; this suggests a further breakdown of complexes over time.

5.4.5.2 Non-aggregated Ternary Complex sample

A different sample, though also formed as detailed in 5.3.1-2, was prepared as a dark complex for electron microscopy; this sample however did not exhibit the levels of aggregation observed in the previous sample (Fig. 5.12). This sample clearly exhibits significantly less aggregate, while the smaller complexes, observed after irradiation of the previous sample, are the predominating feature in this sample. The improved resolution of these images enable the structural features of these complexes to be better ascertained, as individual monomers can now be identified within the complexes. Again, it is clear that these complexes exist in various forms, yet these forms can now be more clearly determined.

The main images in Fig. 5.12A, at the reduced magnification, obviously indicate the presence of a number of ring structures; this is highlighted in the far left expanded image with red circles. Despite these images being at the lower magnification, many particles can be identified as clear 4 and 5-mers (centre expanded). There is also a larger aggregate species apparent in the right hand image which has been further magnified (far right expanded); with a diameter of approximately 40 nm, this is not as large as the previously observed aggregates, yet still indicates that the complex still retains a propensity to form larger aggregates.

At the higher magnification (Fig. 5.12B) even more information can be discerned from the images. As observed in the previous images, ring structures appear most common (far and centre-left expanded); again these can be seen to vary in the number of monomers present

per particle. These images also enable the identification of central monomers at the centre of some of the larger ring structures (centre-right expanded). Ring structures at an angle to the plane of the image can also be observed (far right expanded), suggesting that the rings tend to exist as single layers, rather than stacked rings.

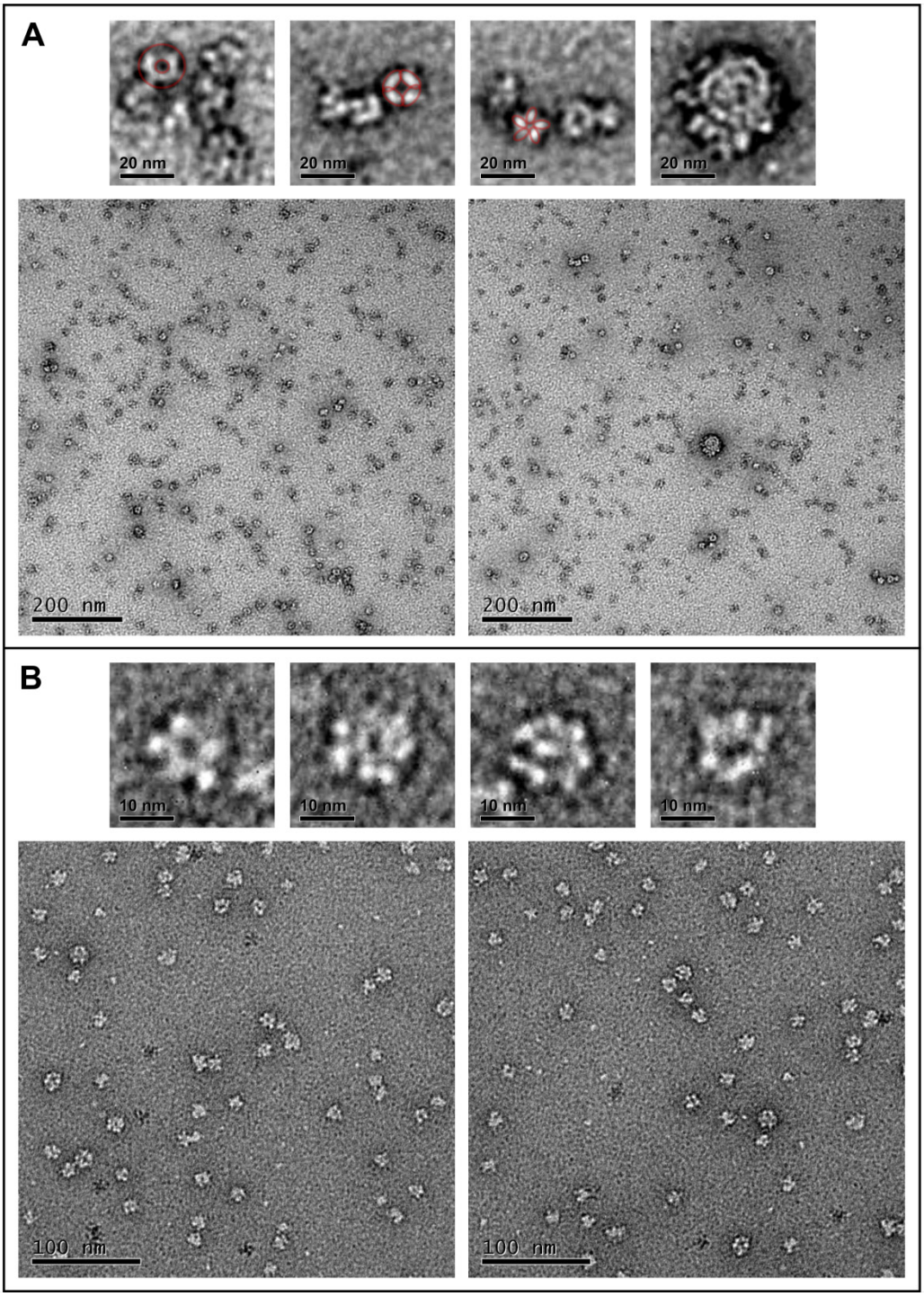

Figure 5.12 Electron micrographs of POR ternary complex sample imaged at two magnifications: (A) 16,000x, (B) 32,000x. Selected regions are expanded above. Scale bars are placed at the bottom left corner of each image. Expanded images from A show, from left to right: ring structure, tetrameric structure and pentameric structure (all highlighted in red), and finally aggregate structure. Expanded images from B show, from left to right, pentamer structure, apparent ring structure (~7mer), ring with central component (~7mer + 1 central molecule) and a ring structure apparently tilted, indicating only a single layer of ring structures.

The reason for the difference in structures observed between the two samples is unclear. The breakdown of the initial large aggregates in the first sample is clearly light-dependent, and therefore must be a functionally related phenomenon; these aggregates subsequently break down into similar structures as observed in the second sample. It is feasible that these smaller complexes are components of the larger structures, which in the case of the second sample, have not formed into these aggregates. This could be due to a slight increase in detergent concentration, which prevents association into the larger aggregates; alternatively, a slightly different concentration of Pchlide may have been present and this may not have permitted the association of these complexes.

5.4.6 Atomic Force Microscopy (AFM)

The technique of atomic force microscopy (AFM) involves the scanning of a surface with a tiny probe, enabling mapping of 3-dimensional images of the surface. This technique has optimal resolutions that approach 0.1 nm vertically and 1 nm laterally; samples can also be imaged in native buffer conditions (Sturgis et al., 2009). AFM experiments were carried out by Dr Cvet Vasilev on the POR ternary complex sample, from the second of the EM preps (detailed 2.26), in order to gain further insight into the structure of these complexes. Samples were placed on an activated mica disk and incubated for 1 hour in order for sample material to attach to the surface, before being washed with buffer. Experiments were carried out using peak force tapping mode on a Bruker Nanoscope 8.0 instrument.

The two-dimensional image from this experiment is shown in figure 5.10A; height of a feature is indicated by colour changes from the dark red surface, through yellow at around 4 nm, to light pink above 5.5 nm. The observed particles are well dispersed and appear to be very similar in size; most have a diameter of between 10 and 20 nm and heights of between 3.5 and 5.5 nm. There is one larger aggregate at the bottom of the image with a diameter of 30 nm and a height of 11 nm; none of the other complexes have a height above 6 nm.

On closer analysis of individual particles, more detailed structural organisation can be identified, with the resolution high enough to distinguish individual protein subunits. Most of the complexes in the sample, in line with the EM images, are ring structures, many with central protein components; some examples of this are 2 times magnified on the left of the main image. The top two images appear to show a central protein surrounded by a ring of 6 - 8 proteins; the bottom two images show slightly smaller structures, which appear to have rings of 5 proteins, the top one with a central pore and the bottom one containing a central component.

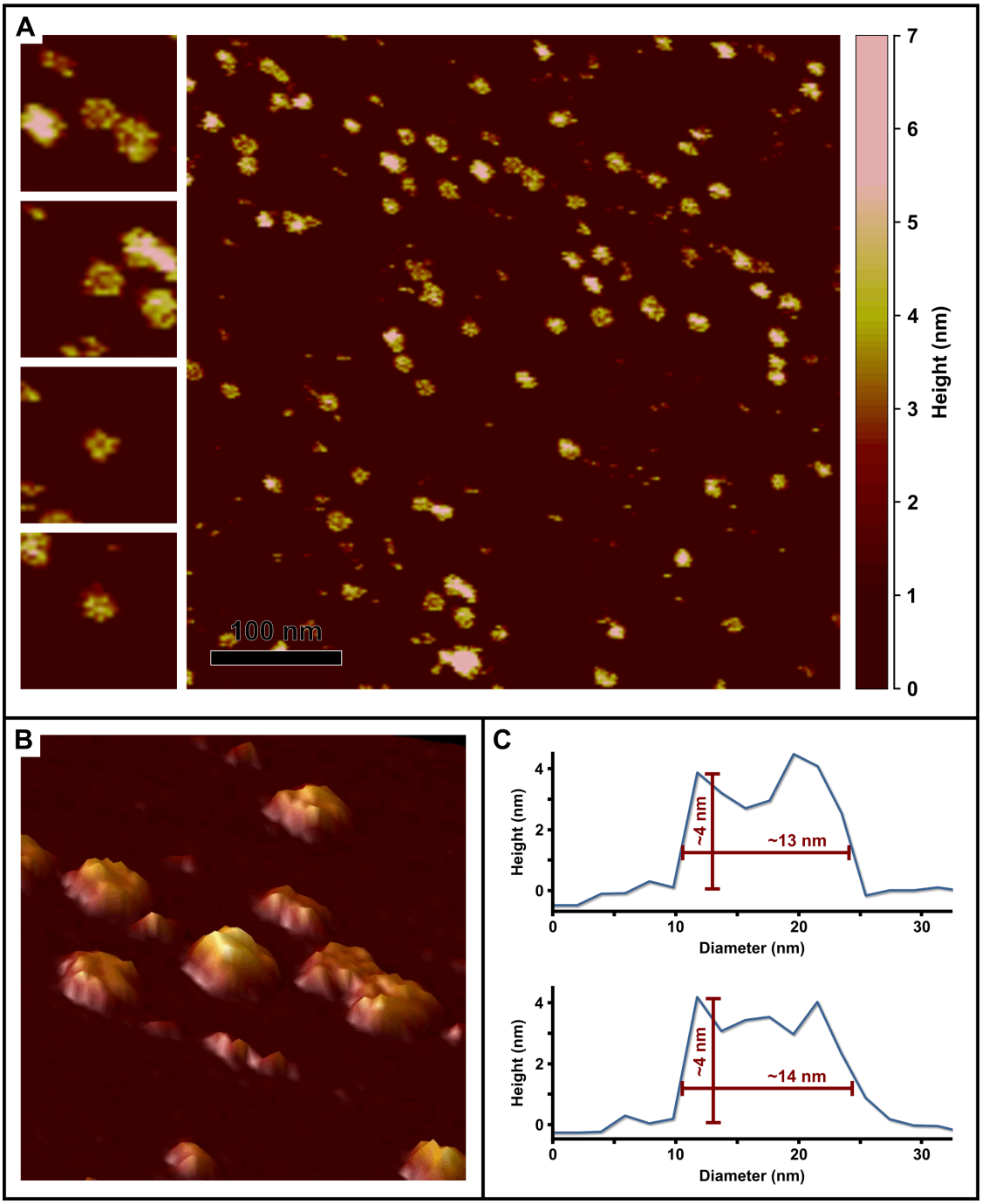

Figure 5.13 AFM data of POR ternary complex. (A) Overhead image with changes in height monitored by colour spectrum displayed on the right hand side. Scale bar shown bottom left corner of main image. Selected regions, on the left-hand side, are magnified 2 x from the main image. A number of complexes of comparable size are present; ring structures are prevalent, some containing a central component. (B) 3D map of protein complexes, using the same colour scale as A. Ring forms, both with and without a central protein, are visible along with more globular structures. (C) Cross-sectional profile of complexes with ring structure, both with (below) and without (above) the central component. Red bars and annotations indicate height and diameter of the complexes. Despite the differences in structure, both examples have comparable dimensions.

A three-dimensional map of the surface (Fig. 5.13B) provides a better idea of the heights of the complexes, and highlights some of these structural features further. The central complex is an example of one of the slightly taller complexes; it appears to be approximately twice the height of the neighbouring peaks, suggesting that it could be two complexes on top of each

other. The complex to the right of this is an ideal example of a complex with a central protein component; the central protein is surrounded by a ring of 6 proteins, with all proteins of an even height. The bottom left sample is a ring structure with a central pore; though it is not as well ordered as the other ring structure, the central pore can be clearly made out in the 3D map.

Figure 5.10C contains analysis of the cross sections of the two common types of ring structure present in the sample, one with a central component, one without. Despite the different in structural morphology, with and without the central protein, both complexes are a similar height and diameter. A height of 4 nm is consistent with the diameter of a single POR protein, indicating that these ring structures are only one protein in height. Both have a diameter of approximately 13 – 14 nm, the approximate diameter of 3 POR proteins; this is unsurprising in the complex with the central protein, however suggests that the ring structure with no central protein contains a pore large enough for another POR protein.

5.4.7 Dynamic Light Scattering (DLS)

The non-aggregated ternary complex sample (i.e. the second ternary complex sample, shown in Figs. 5.8 and 5.9) was also taken to Malvern Instruments Ltd to carry out dynamic light scattering experiments on this sample. Samples were analysed at 55°C on a Zetasizer Nano instrument with a green laser; this would ensure that the light would be scattered and not absorbed by the pigment. A number of output formats are available; intensity data (Fig. 5.14A) indicates the intensity of the light scattered from various particle sizes, and strongly emphasises the largest particles present; volume data (Fig. 5.14B) takes into account the difference in scattering intensity from the variously sized particles, and is more representative of what is truly present in the sample, though it is a less direct representation of the experimental data. Larger particles scatter significantly more light than smaller particles; as the volume data takes this into account, it is a much more reliable indication of the relative amount of particles of each size present in the sample.

The data for a single sample is displayed from 3 sets of scattering data, and all correlate well with each other; any variation observed pertains to the larger aggregate signals. The intensity data indicates that most signal comes from particles with diameters in the hundreds of nanometres. There is also significant signal for particles at approximately 20 nm diameter, as well as for particles with diameters in the μm range. If the intensity data is converted to volume, indicating the amount of protein present at each particle size, then it is clear that most of the protein exists as particles of approximately 15 nm in diameter. There are still

some very large aggregates present in this sample; they represent a relatively small amount of the total sample, however.

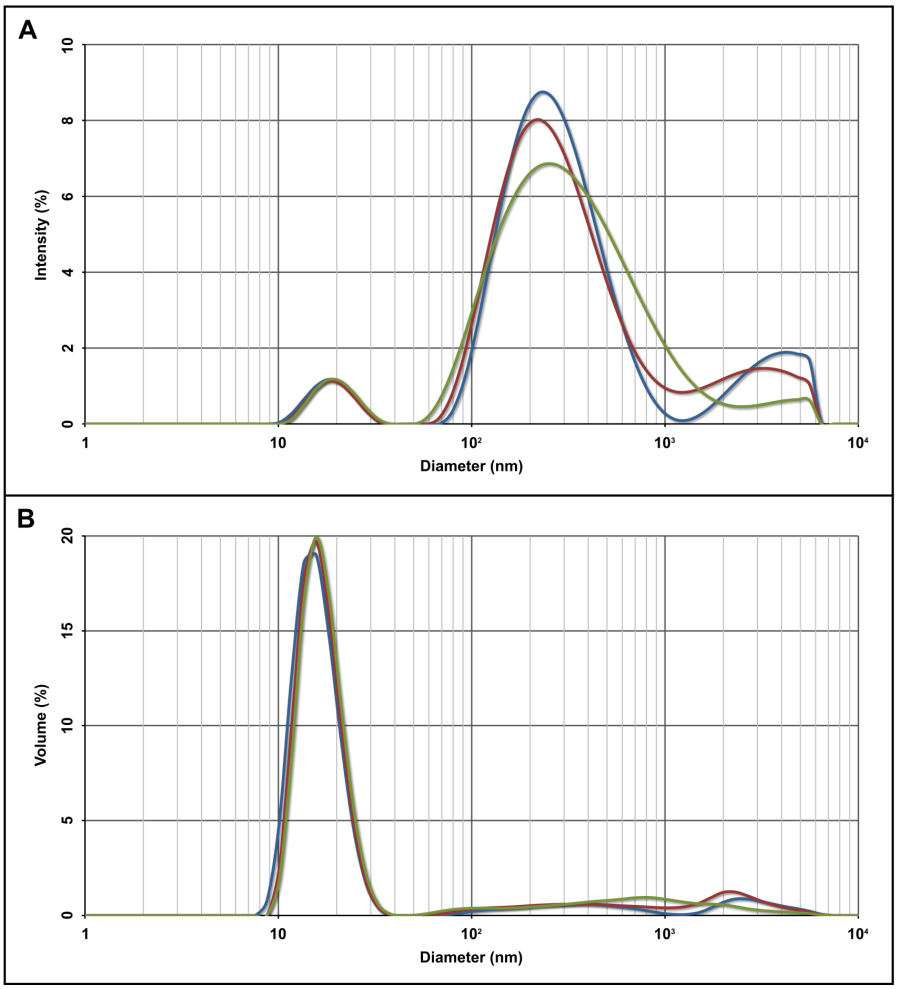

Figure 5.14 DLS data of POR ternary complex. (A) Scattering intensity data for 3 samples, fitted to particle diameter. Particles of approximately 200 nm give the highest intensity of light scattering in a broad fitting of the data. There is also significant intensity at regions around 20 nm and 2 – 5 µm. (B) The intensity data above is recalculated for volume, i.e. mass of particles, as intensity data is skewed towards larger particles. The predominant peak here is at approximately 15 nm and is a relatively narrow peak; suggesting that the vast majority of protein in the sample is in complexes of this size. There are also small amounts of larger particles present, with very broad peaks suggesting a large spread in size of these.

The Zetasizer software includes a calculator tool for rough calculation of the molecular weight of particle based on the hydrodynamic diameter given in the data. A diameter of 15 nm equates to a molecular weight of 380 kDa for a globular protein complex; this is equivalent to 10 monomers of POR, if it was a single globular protein. The EM and AFM data from this sample indicate that it exists in ring structures, rather than globular complexes, therefore the calculation of molecular weight is likely to be a considerable overestimate. This data therefore compares well with the EM and AFM data; it also indicates, as also observed in

EM, that some massively aggregated complexes are still present, though in smaller amounts compared to previous samples.

5.4.8 Analytical Ultracentrifugation (AUC)

In order to obtain a fully comprehensive analysis of the macromolecular structure of ternary complex in this sample, AUC experiments were also carried out; all AUC data was again gathered and analysed by Dr Morou-Besong at the NCMH. Experiments were carried out on apo protein, ternary complex (the sample used for Figs. 5.9, 5.10 and 5.11), product complex and pure Pchlide samples. The product complex sample contained ternary complex with excess NADPH, with the absorbance spectrum monitored during sample irradiation to ensure the reaction had gone to completion. The data is presented in figure 5.12A – D with the top graphs for each sample representing absorbance data and the bottom graphs representing data from the interference signal.

The apo protein experiments are shown in figure 5.12A; this sample was analysed at protein absorbance (280 nm), along with interference, at concentrations of 0.5 (black) and 0.25 (red) mg/ml. In each set of data, there is a predominating peak at a sedimentation coefficient of approximately 3 S, expected to be monomeric POR. In all but one of the datasets, there is also a second peak at approximately 5 S; this is indicative of a dimer as a doubling in particle size should give an approximately 1.5 times larger sedimentation coefficient. There also appears to be small amounts of larger particles in addition to the monomer and dimer peaks. This aggregated material indicates the tendency for POR to form aggregates even as apo protein; this does not seem to be dependent on concentration, however, with both concentrations exhibiting similar levels of aggregate.

The non-irradiated ternary complex sample was analysed at Pchlide absorbance (440 nm) and interference, at concentrations of 0.3 (black) and 0.15 mg/ml (red) (Fig. 5.15B). The 0.3 mg/ml absorbance data indicates that there are no small monomeric complexes present, with a broad peak ranging from 10 to 20 S; this suggests the presence of complexes of approximately 6 - 20 POR monomers. There are no peaks above 22 S suggesting that the prevalence of massive MDa-sized aggregates is very low. At the lower concentration, the size of the complexes is reduced; there are no particles larger than 15 S at this concentration, in addition to the presence of small amounts of apparent monomer and dimer. The interference data also has two large, sharp peaks at 1 and 2.5 S, likely to represent detergent molecules and apo POR, respectively. The frictional ratio of sample components was also calculated (Fig. 5.16), with a higher frictional ratio suggesting a more elongated molecule; this

showed that the smaller complexes were mainly relatively compact in nature, whereas the larger, aggregated particles had more tendency towards elongated structures.

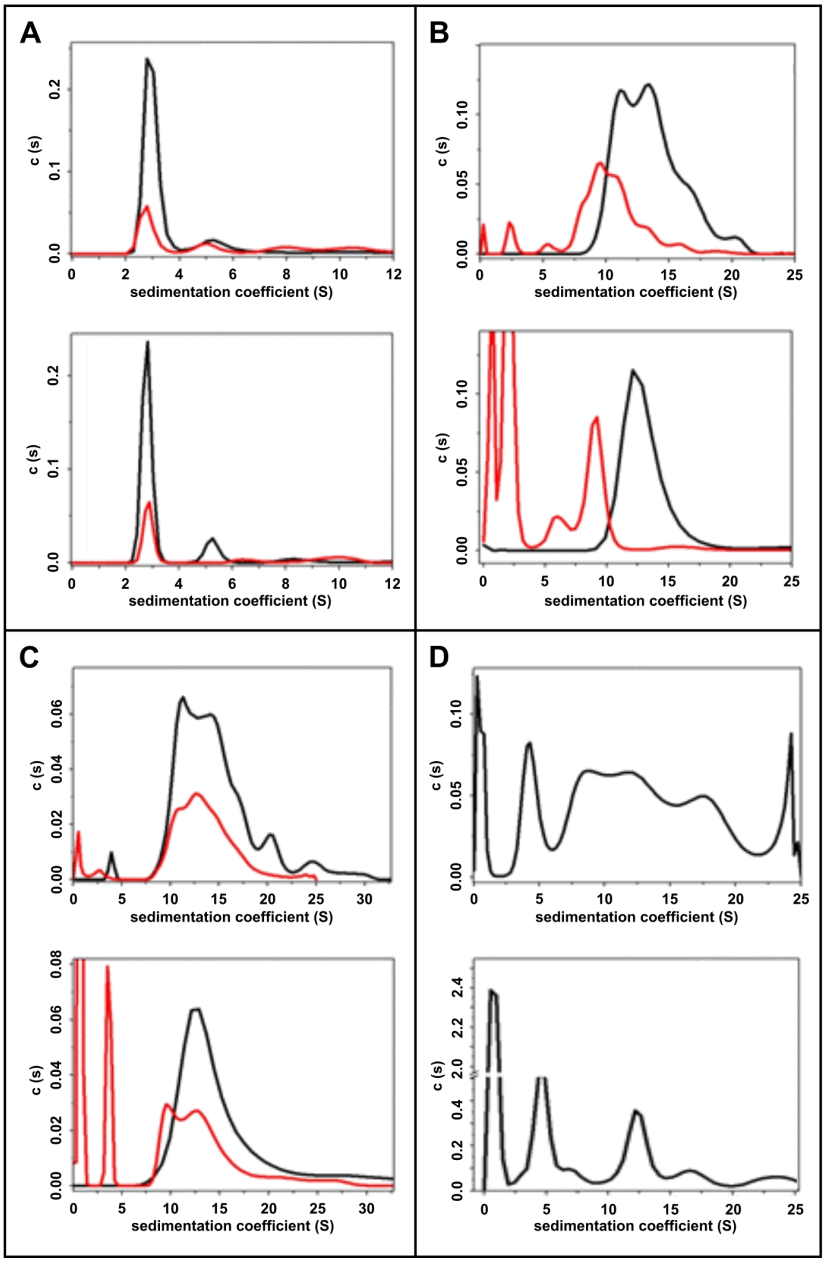

Figure 5.15 Analytical Ultracentrifugation data from: (A) apo POR; (B) ternary complex; (C) irradiated ternary complex; (D) Pchlide. Top and bottom graphs in each panel represent absorbance (for protein signal in A and pigment signal in B, C and D) and interference (signals for all particles, non-specifically), respectively. Graphs in A, B and C have data from two concentrations; the red curve indicates a sample at half the concentration of the black curve. The c(s) value is indicative of the relative proportions of each species in the sample.

The product complex sample was also analysed at 440 nm absorbance, due to Chlide absorption at the same wavelength as Pchlide; the same concentrations as the ternary complex sample were also analysed. The data for this sample was remarkably similar to the ternary complex with complexes from 10 S up to 25 S in the 0.3 mg/ml sample. The

reduction in concentration also seems to have a similar effect in reducing the number of larger complexes present. This suggests that the morphology of the complexes is not altered upon turnover of the protein. The differences in frictional ratios between complexes in this sample were not as profound as observed for the non-irradiated complex.

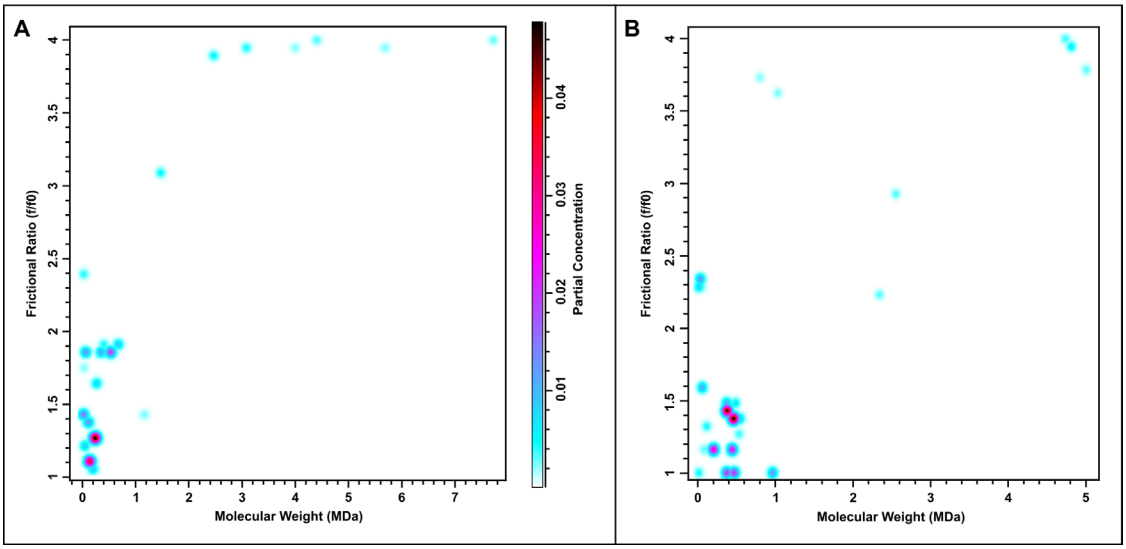

Figure 5.16 3D map of AUC absorbance data for (A) ternary complex (B) irradiated ternary complex. The molecular weight, in MDa, is plotted on the the x-axis and the frictional ratio on the y-axis; the frictional ratio corresponds to the shape of the particle – a spherical particle will have a frictional ratio of 1, with more elongated molecules having a higher ratio. All of the particles below 1 MDa, the majority in both samples, have relatively low frictional ratios, thus are roughly globular. At extremely high molecular weights, above 2 MDa, particles tend to have a higher frictional ratio, thus are likely to be elongated or flat particles; this is particularly emphasised in the non-irradiated ternary complex sample.

The Pchlide sample was at a lower concentration (25 µg/ml) with absorbance readings taken at 630 nm; this is the wavelength of the smaller absorbance peak of Pchlide, selected because the absorbance recorded at 440 nm was too high to enable accurate data. The absorbance data contains a number of peaks ranging from just above 0 S all the way to approximately 25 S, suggesting that there are lots of very large particles that contain Pchlide. These are likely to be formed predominantly of detergent, considering the particle size compared to the size of the pigment. The interference spectrum has a very large peak at 1 S, along with two other large peaks at 5 S and 12.5 S; these are likely to represent detergent particles, with the 1 S peak representing free detergent molecules and the other peaks detergent micelles.

5.5 Discussion

Aggregated complexes of POR in the dark are a feature of the protein in plants, when grown in the dark, with the presence of large structural features called prolamellar bodies (PLBs). PLBs in dark-grown plants are known to be composed significantly of PORA and PORB, along with other components, which dissipate upon irradiation (Reinbothe et al., 1999). The removal of any one of these proteins disrupts PLB structure; though expression of cyanobacterial POR proteins has been shown to restore this structure (Masuda et al., 2009). The structure of isolated cyanobacterial POR ternary complexes has not been studied previously, therefore it was decided to use a number of techniques to analyse these complexes.

Analysis of initial samples, by DLS, AUC and EM, suggested that, upon binding substrates, POR aggregates into massive structures of molecular weight greater than 1 MDa; there also appeared to be no regularity in the complexes formed, with apparent random aggregation. Upon irradiation, and therefore turnover, these aggregates were seen to dissipate; AUC experiments suggested the formation of monomer, while EM suggested it broke down in small complexes of around 5 monomers, possibly further breaking down into monomers over time. These results suggested a similar mechanism to the plant isoforms of POR, forming large structures in the dark which dissipate upon irradiation.

Improvement of the sample preparation protocol, including the use of significantly more pure Pchlide substrate, led to the production of improved ternary complex samples. Unlike the previous samples there was still evidence of some large scale aggregation, however it was much less significant; the predominating complexes were instead approximately 10 - 20 nm in diameter. More detailed EM images of these complexes suggested that these smaller complexes were predominantly ring structures. The ring structures appeared to be quite varied from the EM data, suggesting an apparent randomness of complex morphology; this meant that a 3D reconstruction of the particles in the sample was not possible. Two different sample preparations, both using the improved method described in section 5.3, gave somewhat different results. The data obtained using the second preparation was of particularly interest, due to the clarity of the structures observed, and this sample was used for the majority of the experiments reported here.

The use of AFM gave further insight into the structures of these complexes, while also enabling the complexes to be viewed under native buffer conditions. The ability to gather data on the height of the complexes, as well as the longitudinal structure of the complex meant that the overall shape of the complexes could be better determined. Again, there

were a large number of ring structures present in the sample, some with central pores and others with proteins at the centre; the heights of these ring complexes was consistent with the height of a single POR protein, confirming that these were single level ring structures.

The 2 most common forms of complex structure in both the EM and AFM images, consistent with the AUC and DLS data, were 5-membered ring structures with a central pore, and 6 or 7-membered ring structures with a central protein. The observation of large, aggregated material, in all other preparations of POR ternary complex, suggests that these complexes may associate further; in the presence of a membrane surface, these ring structures could feasibly associate to form large arrays. What function these ring structures of POR could serve is unknown; however it is likely that this behaviour is important for the *in vivo* activity of POR in cyanobacteria.

It is possible that the intermolecular associations within ring structures, with a relatively ordered arrangement, are driven by protein-protein interactions, while the larger scale aggregation observed in initial experiments, which was clearly less regular, may be caused by association of the hydrophobic Pchlide molecules. Previous studies have indicated the ability of porphyrin molecules to form dimers (Mathis and Sauer, 1973, 1972; Vaughan and Sauer, 1974) and AUC of Pchlide in the absence of protein (Fig. 5.15D) indicated apparent random aggregation of the porphyrin. Thus, it is feasible that association of Pchlide molecules helps to drive the association of POR ternary complex, while the conversion to Chlide leads to the disassociation of these aggregates.

Overall, this data indicates an inherent ability of cyanobacterial POR protein to form oligomers upon binding of Pchlide, which then dissociate once Pchlide is converted to Chlide. One of the caveats in the model of chloroplast development via the dark-grown etioplasts is that the organisms must be grown in the dark for a number of days, whereas the plants would be exposed to sunlight more regularly in their native environment. Many cyanobacteria, however, are found deep in oceans where the light levels are low, and thus this process could be more readily applicable to these organisms. This could be an important regulatory process in ensuring that chlorophyll is not produced in too high an excess when not required, a function that could have remained in plants throughout evolution.

Chapter 6

6. Mutational analysis of the POR loop region

6.1 Introduction

One of the most interesting structural features of the POR protein is likely to be a loop region of approximately 30 residues, positioned directly before the catalytic residues. This region is quite unique amongst members of the SDR superfamily, though a similar insert has been discovered in carbonyl reductase enzymes (Baker, 1994); bioinformatics studies in Chapter 3 have also identified another, similar insert in a small number of alternative SDR structures. In plant isoforms of POR, it has been proposed that this region may be implicated in protein-protein interaction events between substrate-bound complexes (Reinbothe et al., 2003b); though other work has suggested a role in membrane anchoring (Birve et al., 1996). In order to understand more about the function of this region of POR, a number of point mutations were made in an attempt to reduce the flexibility of this region; a number of techniques were used to analyse the loop mobility and protein-protein association of these mutants.

6.2 Point Mutations of possible loop 'hinge' residues

The position in which the loop is modelled, across the typical dimerisation helices within the SDR superfamily, is away from the active site and isolated from much of the rest of the protein. This suggests that, in order for the loop to exhibit any functional importance, as investigated later in this chapter, a significant loop movement would be required to position it in a more functionally relevant position. Such a sizeable rearrangement is usually driven by a 'hinge' motion, with most of the movement acting around an axis of one or two flexible residues (Gerstein et al., 1994). Any reduction in specificity of the loop mobility would likely result in less directed movement and impair loop function. It was therefore decided to identify potential hinge points for the loop region in POR and carry out point mutations on these residues in order to reduce the flexibility of the loop.

Due to the nature of hinge flexibility, in that the movement is around very few residues, it is likely that any hinge motion would be driven by glycine residues (Schwarzinger et al., 2002). This is due to the lack of side-chain on the residue which enables greater flexibility around the bond, which would otherwise be restricted by bulkier side chains. A number of glycine residues were targeted as potential hinge points, with glycine residues 144, 154, 155 and 190

identified as mutation targets. It was decided to replace these residues with alanines in order to restrict the movement around the peptide bond. This was thought to be better than a more drastic mutation to a valine as the large bulky side chain would be more likely to impede folding and increase hydrophobicity, rather than only restricting the backbone mobility as desired.

Primers were designed for each mutation and ordered from Sigma-Aldrich. Each point mutation was made (detailed 2.12), the plasmid DNA amplified by mini prep (detailed 2.13) and sequenced (detailed 2.14). Double, triple and quadruple mutants were also produced by performing mutagenesis on plasmid DNA already containing the required background mutation. The constructs made are detailed in Table 6.1.

Number of point mutations	Constructs
Single	G144A; G190A
Double	G154/155A; G144/190A
Triple	G154/155/190A
Quadruple	G144/154/155/190A

Table 6.1 Table outlining combinations of mutations to produce various constructs for POR overexpression.

6.4 Enzymatic Assays of Point Mutants

The conversion of Pchlide to Chlide by POR can be followed spectroscopically, due to the distinct absorbance wavelengths of substrate and product, at 630 nm and 670 nm respectively (Fig. 6.1). In order to determine how the various mutations had affected enzyme function, enzyme assays were carried out on each mutant with respect to WT (detailed 2.16.3). The rate of Chlide production for each mutant was calculated from the change in the absorbance at 670 nm, measured at the linear part of the curve (Fig. 6.2), and compared to WT. This gave the activity of each mutant with respect to the WT protein, indicating how significant an effect the mutations had had on the enzyme function.

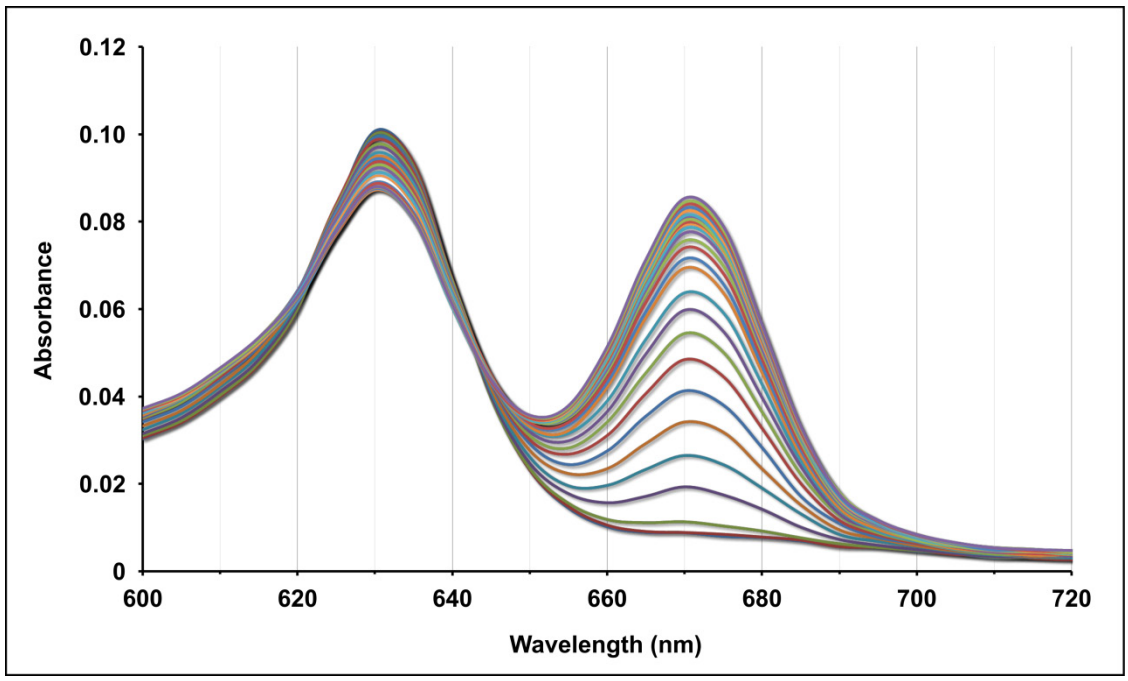

Figure 6.1 Absorption spectra recorded throughout POR assay as Pchlide (632 nm) is converted to Chlide (670 nm). Reaction rate can be monitored by the increase in Chlide absorption over time. The increase at 670 nm is much greater than the decrease at 632 nm due to the higher absorbance by the Chlide molecule, which also retains some absorbance at 632 nm.

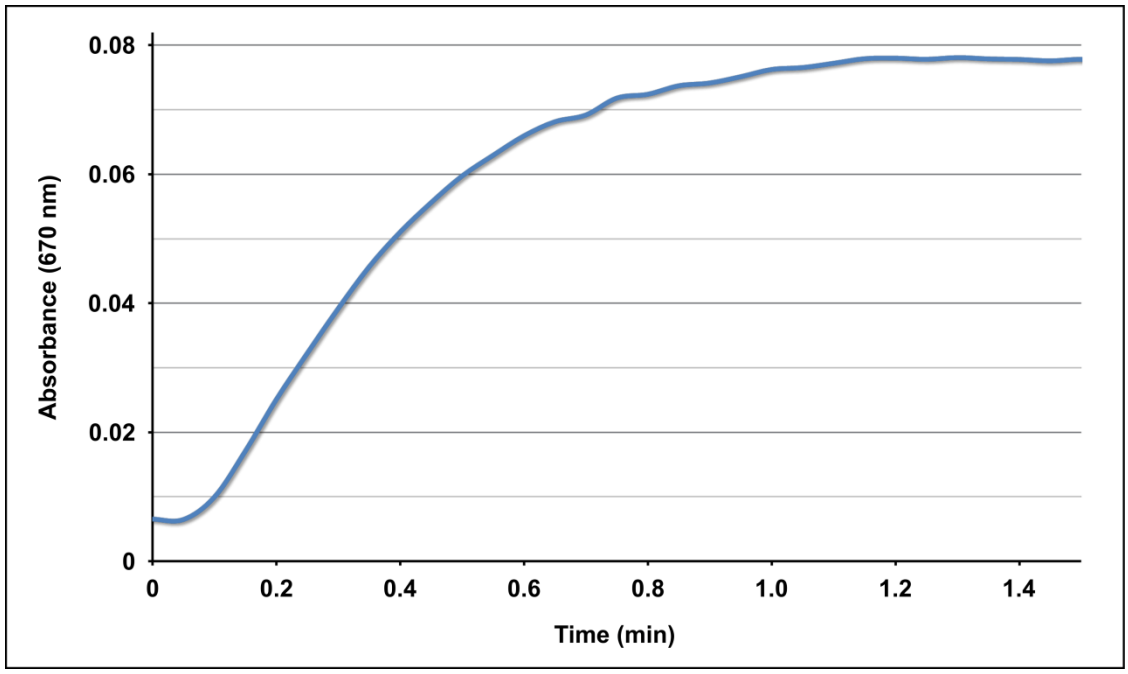

Figure 6.2 Single wavelength spectrum monitoring Chlide absorption (670 nm) over time. The reaction rate is calculated from the linear portion of the curve from approximately 0.1 - 0.3 min in this spectrum.

The activity of each mutant, measured as the initial rate of Pchlide reduction, can be seen in Fig. 6.3. All the mutants were able to convert Pchlide to Chlide, though with reduced rates. The G190A mutant produced the biggest decrease in activity of the single mutants with the activity reduced by more than half, compared with a reduction by about a third of the G144A

mutant. Residues 154 and 155 were mutated as a pair due to their adjacency and this mutant gave a reduction in activity similar to that of G144A.

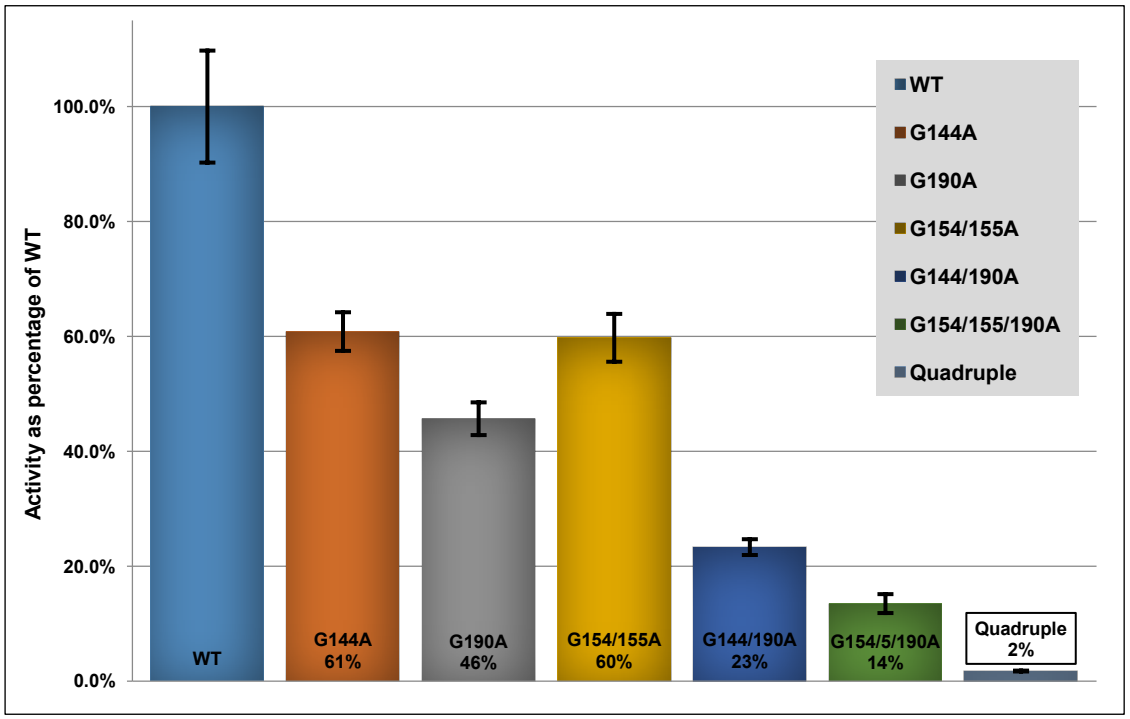

Figure 6.3 Bar chart displaying percentage activity of each mutant with respect to WT. Data are displayed as bars with standard error, based on data gathered in triplicate.

Combinations of these mutations gave further reduction in activity, with the combined G144A/G190A mutant giving an activity roughly similar to the sum of the two single mutations. The combined G154/155/190A mutant exhibited a reduction in activity higher than the sum of the G154/155A and G190A mutants separately. Further to this, the quadruple mutant (G144/154/155/190A) exhibits a 50-fold reduction in activity with respect to WT. It is therefore likely that none of these glycine residues is particularly essential to the mobility of the loop region on their own. They may however all contribute to the intrinsic flexibility of this region, the reduction of which causes a large decrease in the enzyme activity.

Though previous work has suggested that the loop is unlikely to be involved directly in substrate binding (Reinbothe et al., 2003b), the mutations in this proposed region could still affect the binding of Pchlide due to the propagation of structural changes through the protein. Though the mutations are directed to alter the mobility at a sequence in the loop, it is reasonable that any structural changes could affect the overall folding of the protein; these mutants still turnover so must bind substrate, however it is possible that the binding of Pchlide is less optimal than in the WT enzyme.

6.5 EPR of the POR loop mutants

6.5.1 EPR of C37S/C89S/D164C/C198S mutants

In order to directly probe the flexibility of the loop region and the effect on its mobility by the mutations, it was decided to use Electron Paramagnetic Relaxation (EPR). A MTSL (*S* [1 oxyl 2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl]methyl methanesulphonothioate) spin label was used, which gives characteristic line shapes in EPR spectra depending on the freedom and rate of the motion exhibited by the label; the more mobile the spin label, the sharper the signal and the smaller the hyperfine splitting. The label can be attached via a covalent bond to a cysteine residue in order to monitor the mobility of a particular protein region. Mutants were provided by Dr Amanda Brindley, which had all native cysteine residues replaced with serines, with the exception of the essential Cys226 residue; a further cysteine residue had also been introduced into the loop region, in place of aspartate residue 164. The mutations enable a spin label to be attached to the loop region to directly probe the motion of the loop by EPR.

Also provided by Dr Brindley were POR constructs which involved both the mutations above, in addition to each of the loop glycine mutations from 6.4; these were made in order to investigate how each of the mutations affects the loop mobility. In conjunction with Dr Brindley, a total of five protein samples were produced; the D164C mutant (C37S/C89S/D164C/C198S) plus the D164C mutant with G144A, G190A, G144/190A and G144/154/155/190A mutations. $NADP^+$ was added to the samples at 100-fold excess in order to protect the Cys226 residue from labelling before the proteins were labelled with MTSL (detailed 2.20). Each sample was then assayed, alongside an unlabelled D164C sample, to analyse the effect of labelling on activity of each mutant (Fig. 6.4). Activity of each mutant compared to the D164C background was generally similar to the original loop mutant activities compared to WT, with the exception of the D164C and G190A double mutant, which had a significantly lower activity compared to the respective WT background mutant.

EPR samples were run by Dr Steve Rigby, at the University of Manchester, on a Bruker ELEXSYS E580 spectrometer operating at X-band (~9.5 GHz). Each of the samples was prepared (as detailed in 2.10) and EPR spectra obtained for each of the samples at 100 µM POR with $NADP^+$ bound. EPR spectra for the D164C sample were also acquired in various other substrate bound states, each with 0.1 % Triton X-100 added to ensure correct delivery of the hydrophobic Pchlide to the active site.

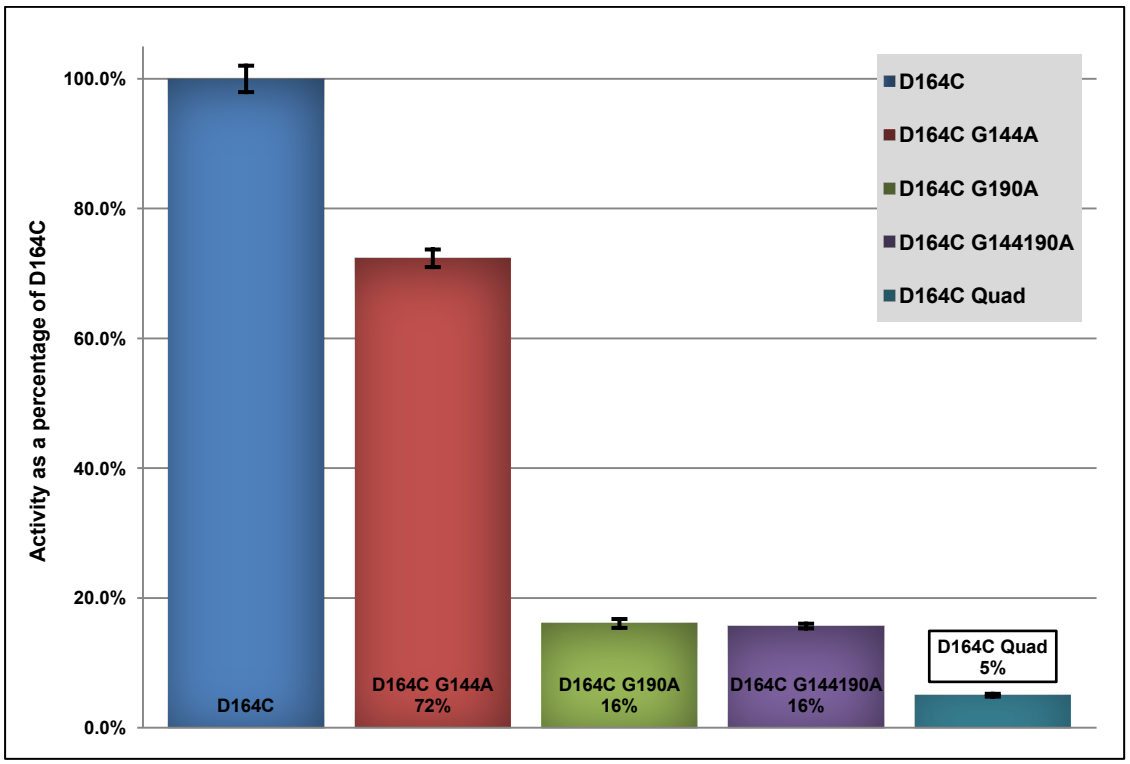

Figure 6.4 Bar chart displaying percentage activity of each EPR mutant with respect to D164C (C37S/C89S/D164C/C198S mutant) as a background rate. Data are displayed as bars with standard error, based on data gathered in triplicate.

The EPR spectra for each apo POR sample are shown in Fig. 6.5A, in addition to the spectrum for free MTSL label in solution. There are two modes of movement apparent in the protein spectra; fast movement corresponding to highly mobile spin label (purple dashed line), and slow movement corresponding to relatively immobile spin label. This is likely to reflect two populations of spin label each in a different conformation; one of which allows a lot of movement independent of the amino acid backbone (side chain motion) and one in which movement is restricted to that of the labelled residue. The relatively immobile label observed in these spectra is similar to that observed for labels on secondary structural elements, suggesting that the labelled loop is not intrinsically mobile on the EPR timescale (ns). The spectra are all quite consistent suggesting that the motion of the label is not altered by the mutations, though the difference in the relative intensities of the two components in the D164C + G144A and D164C + Quad mutants could indicate a change in loop conformation in these mutants.

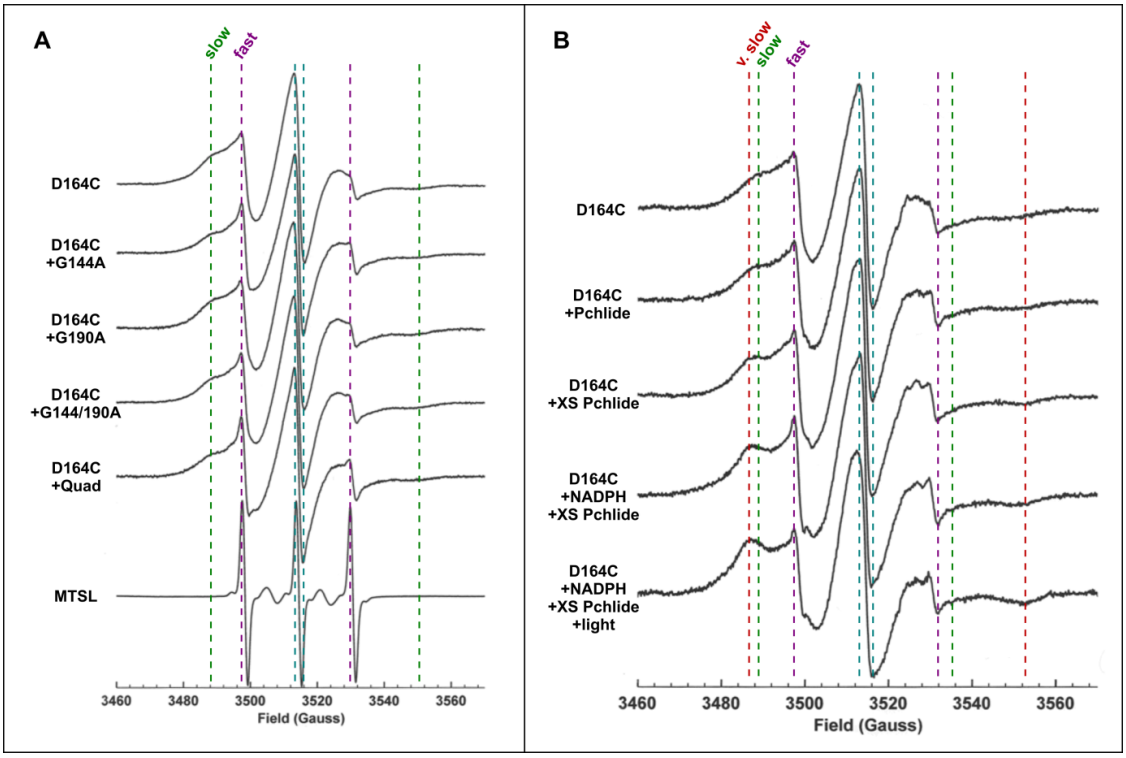

Figure 6.5 EPR spectra. (A) Apo protein samples with free MTSL label for reference. (B) Spectra for D164C mutant at various states: apo; 60 µM Pchlide; 150 µM Pchlide; 150 µM Pchlide and 10 mM NADPH; 150 µM Pchlide, 10 mM NADPH and illuminated. Turquoise dotted lines indicate the broadness of the central EPR peak. Purple, green and red dotted lines indicated the modes of motion designated as fast, slow and very slow, respectively.

EPR spectra were also collected for the D164C protein with various states of substrate binding (Fig. 6.5B). Initially, 60 µM Pchlide was added to give a partially bound ternary complex followed by 150 µM Pchlide which was expected to be closer to being fully saturated with Pchlide. Increasing concentrations of Pchlide introduced a broader signal indicative of a further reduced mobility than seen in the apo samples. The addition of 10 mM NADPH to this sample further increases the proportion of this slower species, suggesting that binding of both substrates has an effect on the label dynamics. The sample was then irradiated with light and further EPR spectra collected on the complex after turnover. The signal for this sample was greatly reduced despite no alteration to the sample other than the addition of light. This is likely due to the reduction of the free radical on the spin label, therefore suggesting that some kind of free radical transfer had occurred during the reaction, something that has previously been suggested in the mechanism (Belyaeva et al., 1988; Lebedev and Timko, 1999). The slower signal seen in the previous samples is seen to increase further after the turnover of substrate indicating that the product complex may remain closed and inflexible.

Due to the apparent free radical transfer that occurred on the addition of light to the sample, it is likely that the label was also attached to the Cys226 residue; for transfer to occur, the

label would have to be in close proximity to the substrates which is unlikely to be the case with the Cys164 label. Although the Cys226 residue was thought to be protected from labelling by the addition of $NADP^+$, this substrate binding is reversible whereas the binding of label is an irreversible reaction. It is therefore understandable that a significant proportion of the protein would have label bound to Cys226 as well as Cys164. The EPR spectra gathered for all the samples is therefore an ensemble of both Cys164 and Cys226 labels so does not represent purely loop mobility.

6.5.2 EPR of C37S/C89S/C198S mutants

The EPR experiments on the C37S/C89S/D164C/C198S mutants clearly indicated that both Cys164 and Cys226 had been labelled, therefore it was decided to repeat the experiments without the D164C mutation so that the proteins were labelled on Cys226 only. This would enable the isolation of this data from the previous experiments and also enable any change in mobility of this residue to be analysed on substrate binding. All proteins were purified (detailed 2.10) and 100-fold excess of NADP+ added to each sample prior to labelling, as in 6.6.1; this would ensure identical conditions to the previous set of experiments.

EPR spectra were recorded for apo protein (Fig. 6.6A), 10 mM NADPH added (Fig. 6.6B), and further addition of 150 µM Pchlide (Fig. 6.6C). Unlike the previous set of experiments, a clear difference can be seen between the spectra of the various loop mutations. There are 3 modes of motion in these spectra, fast, slow and very slow; the ratio of these varying throughout samples. The EPR spectrum of the so-called WT protein (C37S/C89S/C198S) looks similar to that of the Cys164/Cys226 labelled protein, with a combination of fast and slow components. In the G190A, G144/190A, G154/155A and Quadruple mutants, the two modes of motion present are the fast and very slow signals. This suggests some further restriction of the backbone mobility, possibly due to altering the protein conformation, or a restriction in the freedom of movement by the label. There is also a difference in the ratio of fast and slow components for each mutant, with G144A and G154/155A seeming to show more fast motion than slow, while G190A and G144/190A mutations are weighted more towards slow components.

Upon addition of NADPH, all of the proteins switch to a combination of fast or very slow components. Again the ratio between the two types of motions is different, with G144A and G154/155A exhibiting a greater proportion of fast components; suggesting the label is in a conformation that promotes intrinsic label motion. Addition of Pchlide to the samples leads to a further mode of motion in the G190A, G144/190A and Quadruple mutants, even slower

than the very slow component, designated as extra slow. The spectra for the remainder of the proteins do not change on addition of Pchlide suggesting that the label environment is not altered upon addition of Pchlide. It appears that the presence of the G190A mutation somehow alters the environment of the label, causing it to be further restricted in its movement upon addition of Pchlide.

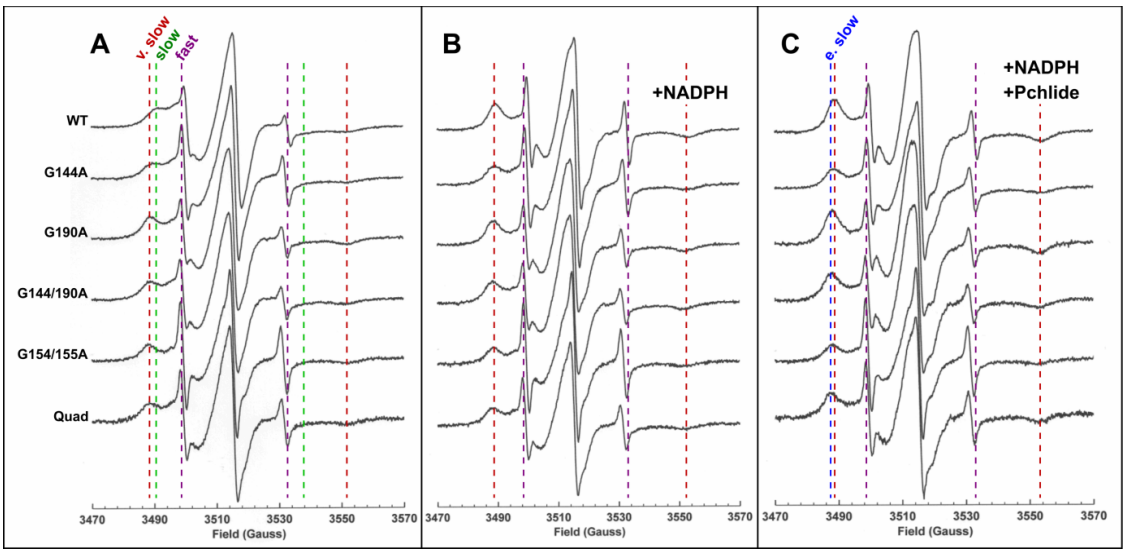

Figure 6.6 EPR spectra of Cys226 labelled samples. (A) Apo protein. (B) 10 mM NADPH added. (C) 10 mM NADPH and 150 µM Pchlide added. Purple, green, red and blue dotted lines indicated the modes of motion designated as fast, slow, very slow and extra slow, respectively.

6.5.3 Activity of C226S mutant

Due to the issues in labelling only Cys164 when Cys226 is also present, the only way to observe solely loop mobility by EPR would be to remove the Cys226. The concern with this mutation is that it is known that Cys226 is essential and thus removing it may cause the protein to be inactive. This would lead to any structural inferences being based on a non-functional enzyme which would not be representative of the native protein. A C226S only mutant was therefore produced and assayed (detailed 2.16) in order to ascertain the activity of this mutant with respect to WT and ensure that it retains function. The activity calculated for C226S with respect to WT was 30.9 ± 1.3 %. This is a higher value than had been determined for *Synechocystis* POR (Menon et al., 2010), though does suggest that this mutant is sufficiently active to give a realistic representation of the structural behaviour of WT POR.

6.5.4 EPR of C37S/C89S/D164C/C198S/C226S mutants

The D164C mutation, in addition to the removal of all the native cysteines, was provided by Dr Brindley, giving a C37S/C89S/D164C/C198S/C226S mutant, which could be specifically labelled on only the Cys164 residue. Proteins were purified (detailed 2.10), labelled (detailed 2.20), and assayed along with WT protein (detailed 2.15), with activities calculated with respect to WT protein (Fig. 6.7) and the D164C background (Fig. 6.8). Interestingly, these show that the reintroduction of a cysteine, though on the loop region of POR, seems to give an increase in the activity; with D164C with Cys226 deleted having an activity of around 50 % compared to the activity of around 30 % for the C226S mutant. With respect to the D164C background protein, all the other mutations seem to have comparable activities to those seen for the D164C protein in the presence of Cys226. For these mutants the activity of the G190A protein is even further reduced compared to the other mutations, reinforcing the possibility that the D164C and G190A mutations are somehow intrinsically linked.

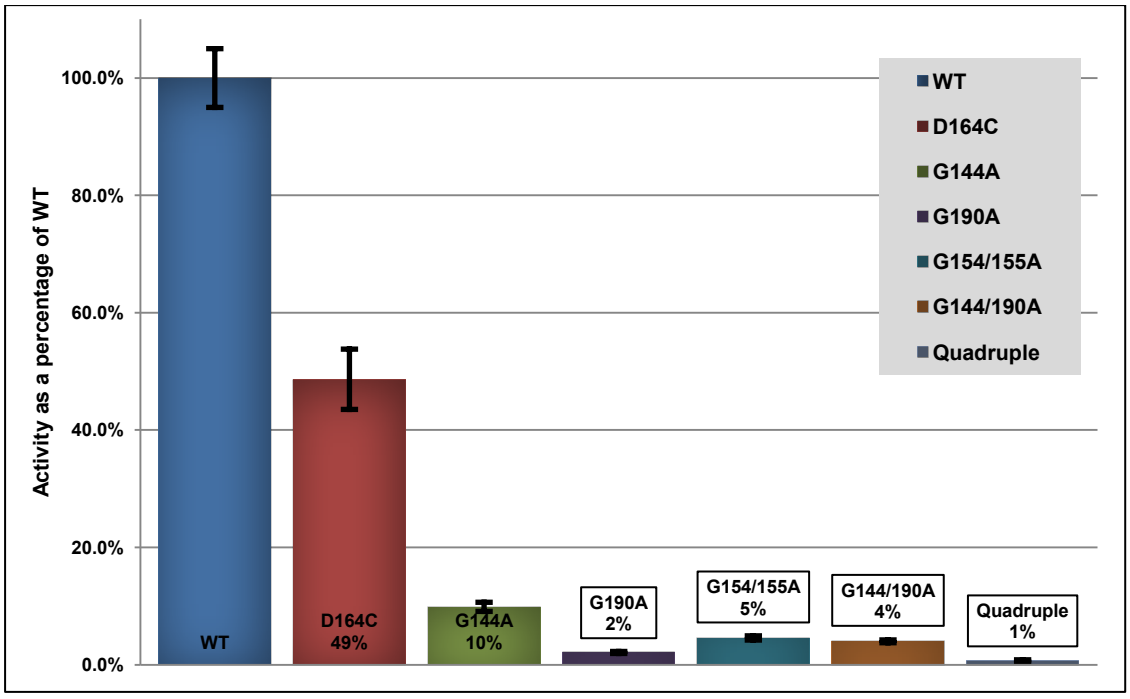

Figure 6.7 Bar chart displaying percentage activity of each EPR mutant with respect to WT. Data are displayed as bars with standard error, based on data gathered in triplicate.

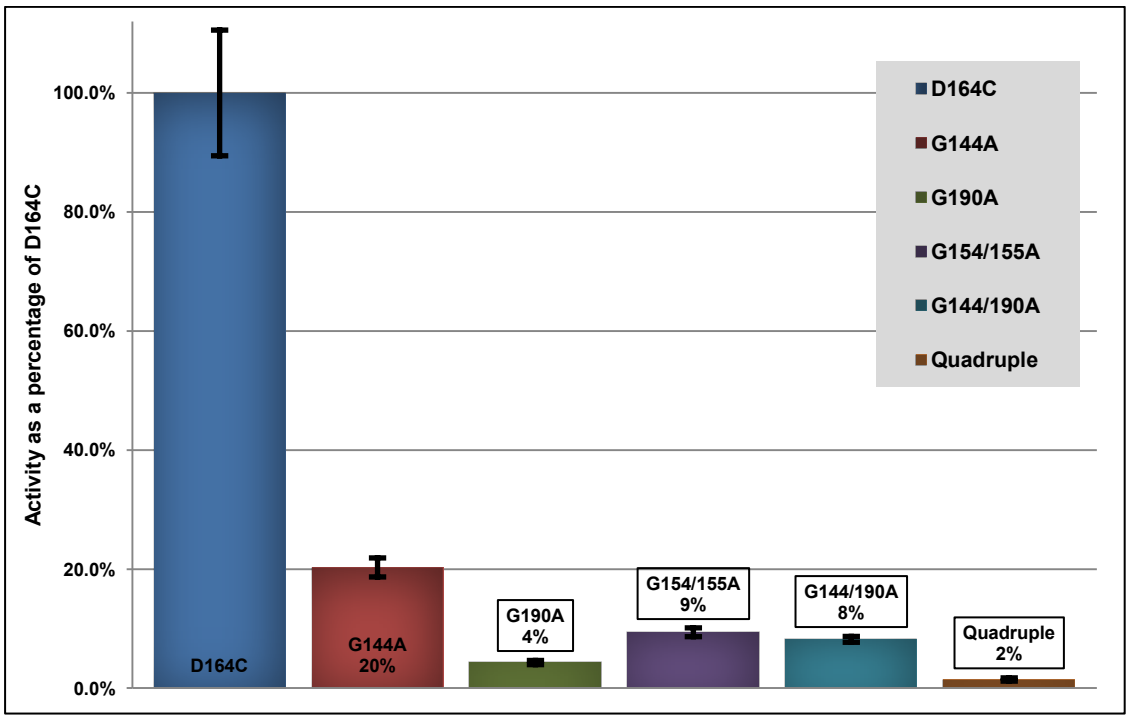

Figure 6.8 Bar chart displaying percentage activity of each EPR mutant with respect to D164C labelled protein (C37S/C89S/D164C/C198S/C226S mutant) as a background rate. Data are displayed as bars with standard error, based on data gathered in triplicate.

EPR spectra were again collected for apo, NADPH bound and ternary complex of each protein. All apo protein samples exhibited 2 modes of motion, fast and slow, with fast motion dominating as expected for the loop region (Fig. 6.9A). G144A and G154/155A again exhibit a higher proportion of fast motions while WT and G190A show a higher proportion of slow motions. Generally, it appears that none of the mutations have significantly restricted the global mobility of the loop region; in fact the G144A and G154/155A mutations appear to have increased the mobility. This may not be the case and instead could be due to a conformational restriction of the loop, as a restriction in the possible loop positions could appear faster via EPR due to the increased rate of movement between two arrangements.

Unlike the Cys226 labelled protein, there was no apparent change in the types of motion in each spectrum upon addition of NADPH (Fig. 6.9B). There is, however, a general shift towards a higher proportion of slow component in all samples, with WT and G190A exhibiting a more prominent slow feature. Upon addition of Pchlide to form ternary complex (Fig. 6.9C), each spectrum reveals a very slow component, indicating that the loop mobility has been constrained by substrate binding. It is possible that this is consequence of the oligomer formation, exhibited in Chapter 5, with the loop being in a fixed conformation in the ternary complex structures. This is clear evidence for loop function in substrate binding events, either directly or indirectly due to conformational rearrangements.

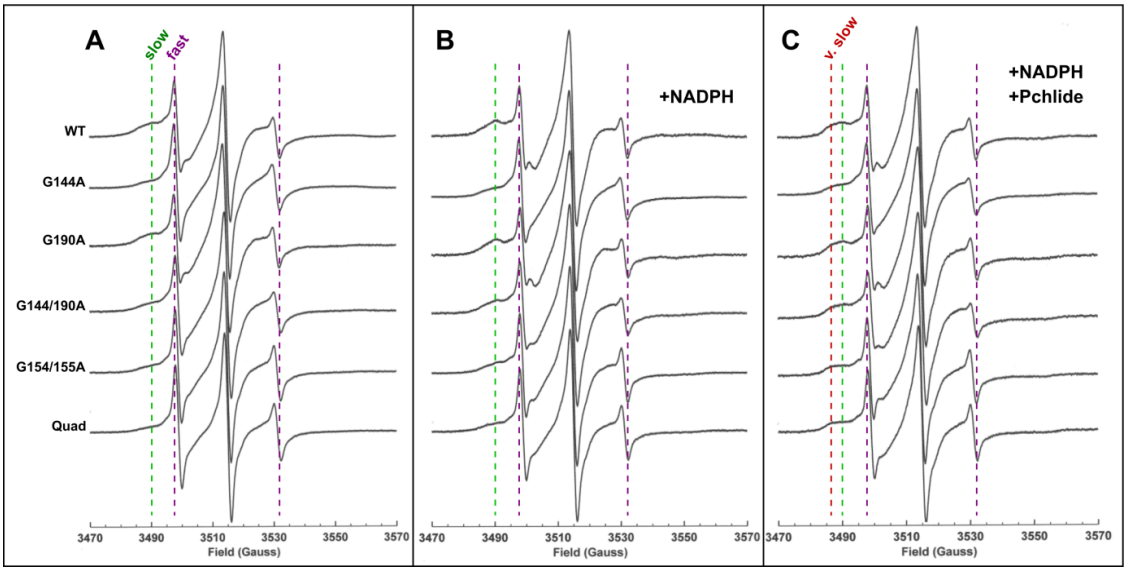

Figure 6.9 EPR spectra of D164C labelled samples. (A) Apo protein. (B) 10 mM NADPH added. (C) 10 mM NADPH and 150 µM Pchlide added. Purple, green and red dotted lines indicated the modes of motion designated as fast, slow and very slow, respectively.

6.6 AUC

In order to analyse the effects of the mutations on the macromolecular structure of the ternary complex, samples were prepared for Analytical Ultracentrifugation (AUC) experiments. Aggregation of POR ternary complex in cyanobacteria has only been reported by (Proudfoot, 2011), however the presence of POR in prolamellar bodies is integral to its complex, ordered structure (Gunning, 2001; Solymosi and Schoefs, 2008). It is possible that the loop could play a part in the protein-protein interactions involved in complex formation and that any mutations that perturb the structure or mobility of the loop could affect the ability for the protein to form these complexes upon substrate binding.

WT, G190A, G144/190A and G144/154/155/190A mutants were purified (detailed 2.10) and the ternary complex, with $NADP^+$ and Pchlide bound, produced and purified (detailed 2.17). Each of these samples were split, with half remaining as a $NADP^+$ bound complex; to the other half 100-fold excess of NADPH and 20 mM βOG was added, the sample was incubated at 55 °C and illuminated with light for 5 min to ensure complete turnover of the enzyme. Apo protein, ternary complex and irradiated ternary complex samples, for each mutant, were run on AUC.

6.6.1 AUC of apo protein samples

Apo protein samples were analysed (Fig. 6.10) at an absorbance of 280 nm, to ensure that only protein components of the sedimentation were observed. All samples exhibited a major peak at 3 S; again, this is expected to be monomeric POR, confirming that the monomeric apo protein form is still the major component in all the mutants. WT and G190A apo POR both had a remarkably similar sedimentation profile which suggests that, although the monomeric form predominates, there are also a number of larger protein complexes in the sample. Though it is known that POR is significantly hydrophobic and associates to form precipitates at high concentrations (> 300 μM), it was not expected that a stable apo POR sample would have quite such a significant number of large complexes. The G144A sample and, particularly, the G144/154/155/190A sample give a more substantial monomeric peak and fewer larger protein complexes, though they are still present. Generally, there is very little difference between the mutant apo protein samples and the WT apo protein.

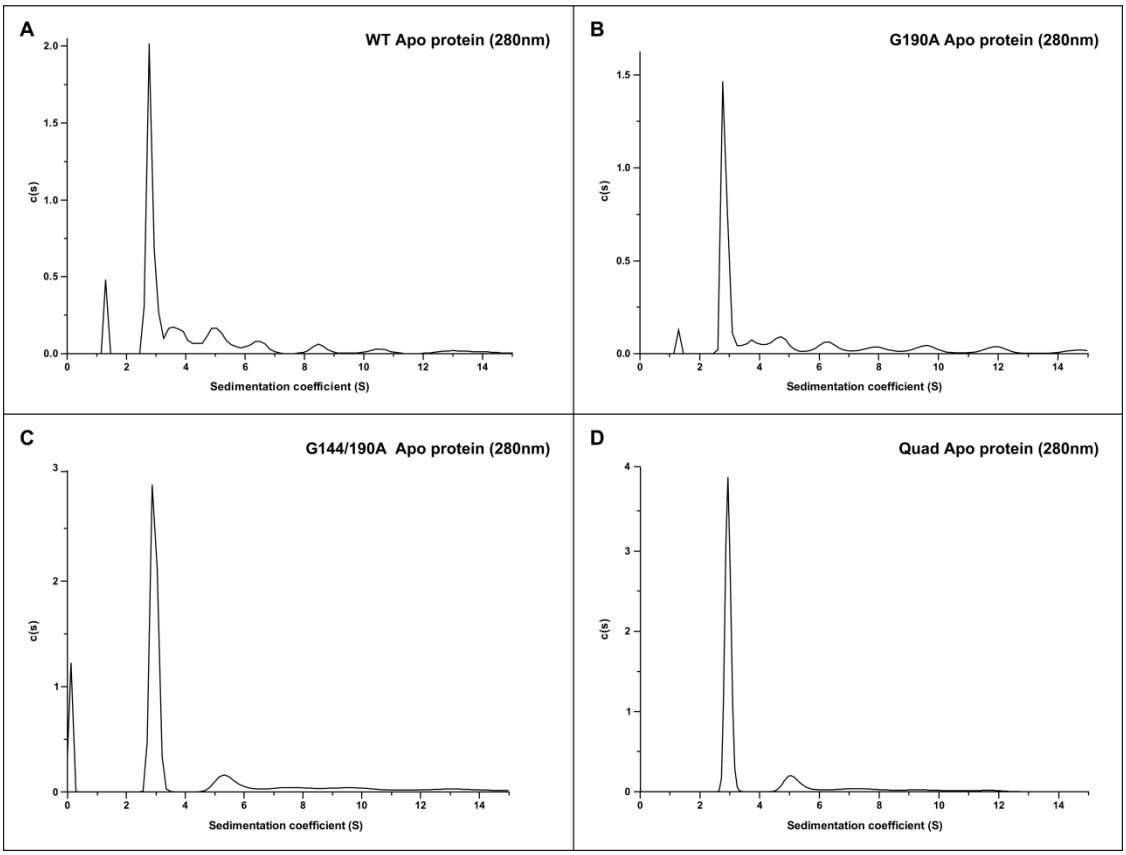

Figure 6.10 AUC spectra of apo protein samples for (A) WT; (B) G190A; (C) G144/190A; (D) G144/154/155/190A. The relative amount of each component is displayed on the y-axis. Sedimentation coefficient (S), representing relative particle size, is displayed on the x-axis.

6.6.2 AUC of ternary complex samples

Ternary complex samples were analysed at 440 nm (Fig. 6.11), to ensure that only protein in complex with substrates was observed. The general trend of all the samples was that the addition of substrates led to a massive increase in the particle size observed. For WT, the peak sedimentation coefficient value is around 16.5 S, estimated to be in the region of 15 - 20 POR monomers (600 - 800 kDa). There are also larger particles which come above the experimental bounds of the experiment at 30 S, which corresponds to approximately 50 POR monomers (2 MDa). In all the mutants, there is still a significant amount of large particles observed, though this seems to have reduced in respect to monomer. The peak observed at 1.8 S, seen clearly in G190A and Quadruple mutant ternary complex samples, is expected to be monomer though it appears slightly smaller than the observed monomer in the apo protein samples. There also appears to be a more regular particle size in the mutants, as seen from the oscillating pattern of the peaks. It is difficult to determine whether this effect is linked to the mutations but it could reveal an underlying structural pattern within what appear generally disordered aggregates.

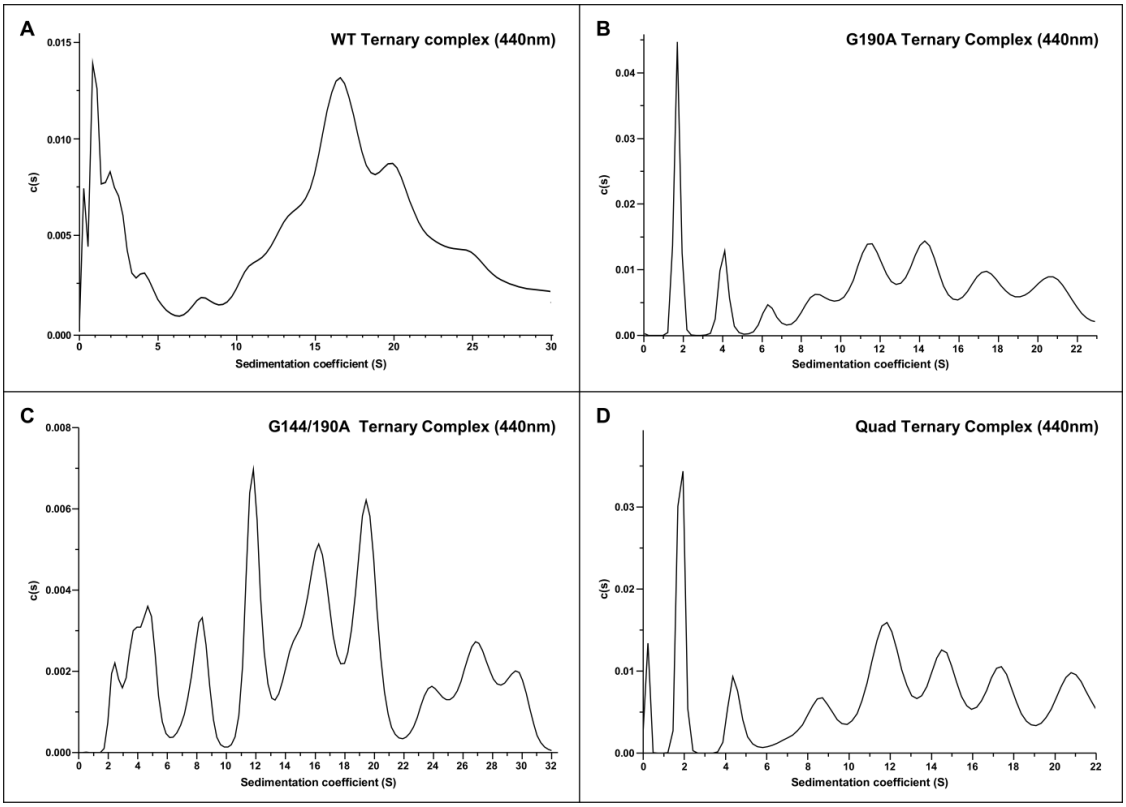

Figure 6.11 AUC spectra of ternary complex samples (with $NADP^+$ and Pchlide bound) for (A) WT; (B) G190A; (C) G144/190A; (D) G144/154/155/190A. The relative amount of sediment is displayed on the y-axis. Sedimentation coefficient (S), representing relative particle size, is displayed on the x-axis.

6.6.3 AUC of irradiated complex samples

The main characteristic of the irradiated complex is the disappearance of the large aggregates observed in the non-irradiated ternary complex (Fig. 6.12). These samples were analysed at the pigment absorbance (440 nm) in order to ensure that only protein in complex with the Chlide product would be observed. The clear distinction of the irradiated samples from the non-irradiated ternary complex samples is the lack of large aggregates in the samples. There is one dominant peak in all but one of the samples at approximately 1.5 S, and a second smaller peak at approximately 3 S. The exception is the quadruple mutant which appears to be an anomalous result, with the same pattern of peaks but the values reduced by approximately 0.8 S; to 0.7 and 2.2 S, respectively. It is clear that the effect of the mutations on the product complex is minimal with no change in sedimentation pattern apparent.

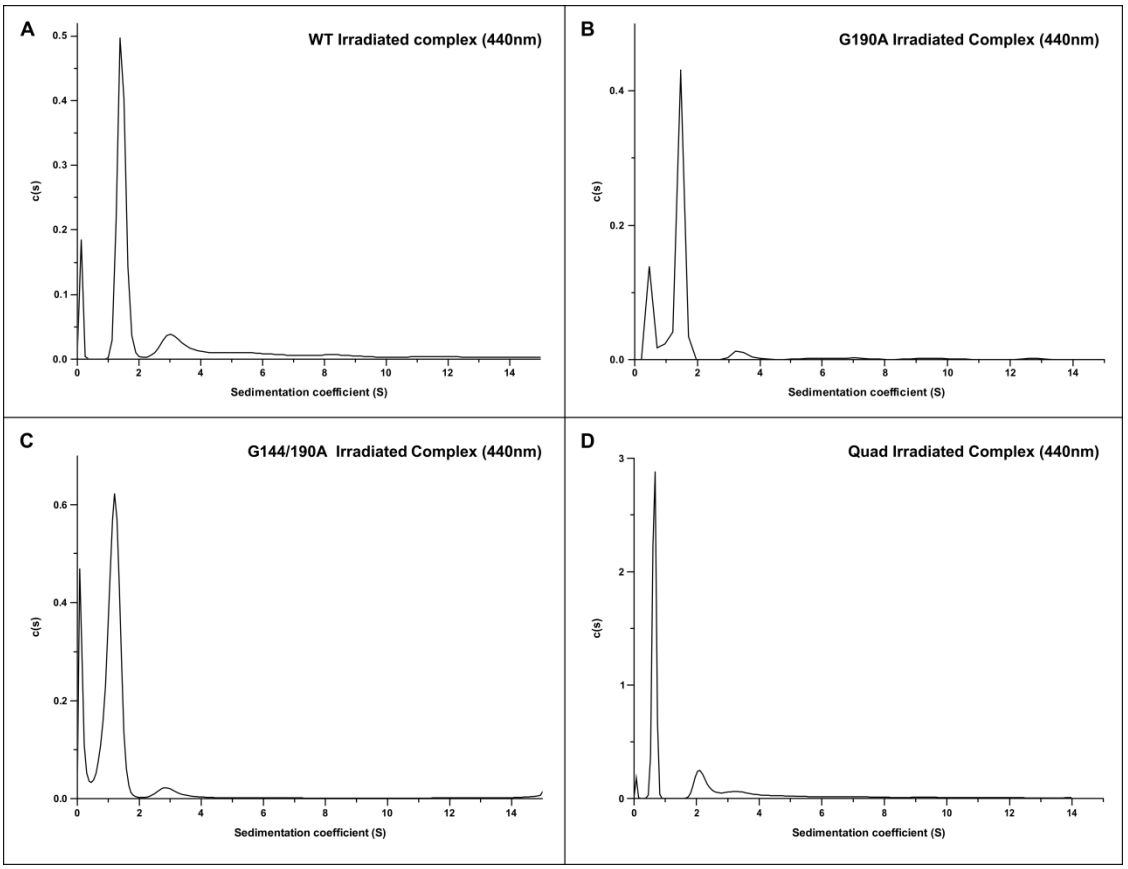

Figure 6.12 AUC spectra of illuminated ternary complex samples (NADPH and Pchlide bound) for (A) WT; (B) G190A; (C) G144/190A; (D) G144/154/155/190A. The relative amount of sediment is displayed on the y-axis. Sedimentation coefficient (S), representing relative particle size, is displayed on the x-axis.

6.7 EM

WT, G190A, G144/190A and Quadruple mutants were purified and prepared as ternary complex (detailed 2.17). Samples were diluted appropriately and adsorbed onto freshly glow-discharged carbon-coated copper grids that were subsequently negatively stained and EM images of each sample were collected, with experiments conducted by Dr Pu Qian (Fig. 6.13) (detailed 2.25). The WT sample exhibits a large amount of aggregation, though it is difficult to determine what proportion of the protein is aggregated and what size the aggregates may be. All of the mutants clearly show some aggregated material however this appears to be more isolated than in the WT images. These also appear to have a large amount of monomeric protein, however the AUC results clearly indicate the significant presence of unbound POR in the complex samples therefore it is likely that these particles represent apo protein.

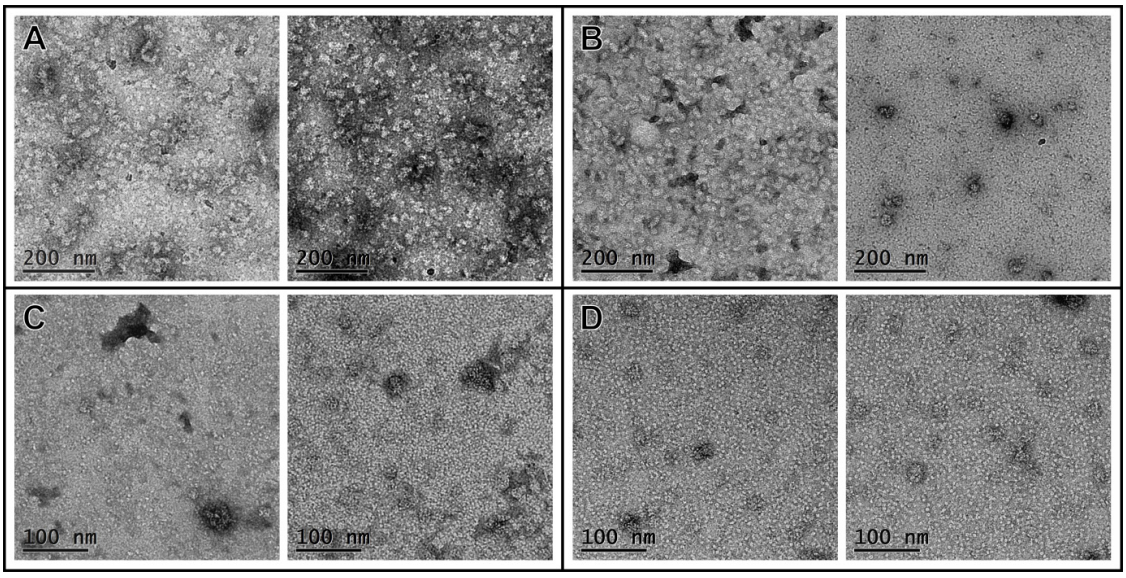

Figure 6.13 Negative stain EM micrographs of the ternary complex samples for (A) WT; (B) G190A; (C) G144/190A; (D) G144/154/155/190A. Scale bars are shown inset.

6.8 Discussion

The unique loop region is thought to be involved in substrate binding or protein-protein interactions. Very little has been reported about the structure and function of this region due to the lack of structural data for the protein and the uniqueness of this region within the SDR superfamily. Mutations of glycines within the loop region clearly reduced activity of POR, suggesting that these perturbed an important function of the protein; it could not be ascertained whether this reduction in activity was related to the disruption of the catalytic mechanism or substrate binding, however. AUC experiments of the apo protein, ternary

complex and irradiated complex did not indicate any substantial difference between the WT and mutant POR samples. There was a slight decrease in the number of larger aggregates observed in the mutant samples yet the mutations clearly did not affect aggregate formation in general, suggesting the decrease in activity is not inherently linked to complex formation. EM images also indicated that, although the mutants appeared to contain more monomeric protein, there were a significant number of large aggregates present in all samples.

EPR spectra confirmed that the loop exhibits motion which is then restricted upon binding of NADPH and more significantly by the binding of Pchlide. The data also suggests that there are differences in the motion of the loop between the WT and the various combinations of mutations. The G144A and G154/5A mutations appear to exhibit an increase in the rate of loop motion compared with WT and G190A mutations. This suggests that none of the glycine mutants act as apparent hinges, suggesting instead that the loop in the apo protein is inherently flexible rather than being a mobile structured region. The apparent increase in motion of some mutations can be put down to a reduction in loop flexibility, reducing the range of possible loop conformations and thus exhibiting EPR spectra which appear to have a faster rate of motion.

This hypothesis ties in with the disorder predictions in Chapter 3, which suggested a tendency for the loop to be unstructured. This unstructured, flexible loop region could then adopt a rigid structural form upon Pchlide binding; either by direct association with the substrate or by global protein structure rearrangement. Reduced mobility of the loop appears to be related to the formation of ternary complex oligomers and thus, linked to this oligomerisation process that occurs upon substrate binding. The number of proteins reported to contain inherently unstructured regions has dramatically increased recently with the advent of new techniques to study these regions (Dunker et al., 2001). It is therefore reasonable to consider that POR does indeed contain an unstructured loop region and that this is involved in either substrate binding, or protein-protein interaction events upon binding of Pchlide.

The structural model produced in Chapter 3, suggests that Pchlide binds in a cleft, enclosed by the α-helix containing the catalytic residues and the C-terminus. The loop does not appear to be in direct contact with the Pchlide molecule, therefore it is unlikely that binding of Pchlide directly effects any change in loop conformation. The start and end of the loop region, however, are both located close to the binding site of the Pchlide model, therefore it is feasible that Pchlide binding could propagate a conformational change within the loop through the residues at either end of the loop region.

Chapter 7

7. Preparation and NMR of ^{2}H, ^{13}C, ^{15}N-labelled POR

7.1 Introduction

Previous attempts to study the structure of POR have been unsuccessful, mainly due to the difficulties in forming crystals for X-ray crystallography; this is likely due to the presence of a number of unstructured regions in POR which would disrupt the crystal packing, in addition to the propensity of the protein to form aggregates. Though much work has been done involving the analysis of the light-dependent catalytic mechanism of POR, the determination of the structure of POR would drive further study of this unique reaction. NMR has been identified as an alternative method for determining the structure of POR as proteins can be studied in solution, this also offers an opportunity to analyse the dynamics of the protein. POR is a 37 kDa protein, which is quite large for structural studies by NMR, compared to the NMR structures that currently exist. Recent advances in NMR hardware, software and labelling techniques, however, have enabled the study of far larger proteins by NMR. This chapter discusses the production of a high quality, ^{2}H, ^{13}C, ^{15}N-labelled sample of POR for NMR experiments; this has enabled the determination of secondary structure and dynamics for much of this protein.

7.2 Optimisation of NMR pH

Previous studies on POR involved the optimisation of the NMR conditions in order for production of the best possible quality spectra (Proudfoot, 2011). The decision to perform experiments at pH 5.5 has since been questioned, due to the suggestion that broad signals, which are observed in the spectra, may be due to the presence of a molten globule state of POR at this pH. In order to determine the optimal pH for NMR of POR, it was decided to carry out both 1D ^{1}H and 2D ^{1}H, ^{15}N-HSQC experiments at a range of pH values.

A ^{15}N-labelled POR sample was prepared for NMR (detailed 2.7 - 2.10) and transferred to NMR buffer (detailed A3.6) with the pH adjusted to 7.5. A 500 µl sample at 150 µM was placed in a NMR tube and sealed; the tube was then transferred to a Bruker Avance I 800 MHz NMR machine set to a temperature of 318 K (45 °C); both 1D ^{1}H experiments and 2D ^{1}H, ^{15}N-HSQC experiments were performed on the sample. After completion of these

experiments, the sample was removed, the pH adjusted to 6.5 and the sample returned to the machine; the experiments were then repeated at this pH, and again at pH 5.5.

The HSQC spectra at pH 5.5, 6.5 and 7.5 are shown in figure 7.1A, 7.1B and 7.1C, respectively. The majority of the peaks in the three spectra overlay well, however it is clear that there is an increase of broadening in signals in the central region of the pH 5.5 and 6.5 spectra. There are also, however, a number of extra peaks present in the pH 5.5 spectrum, compared to the pH 6.5 and 7.5 spectra, indicating that many signals that were not identifiable at higher pH, are observable at pH 5.5. This suggests that the broadening in the centre of the pH 5.5 spectrum has arisen due to the presence of previously unidentified peaks with broad signals.

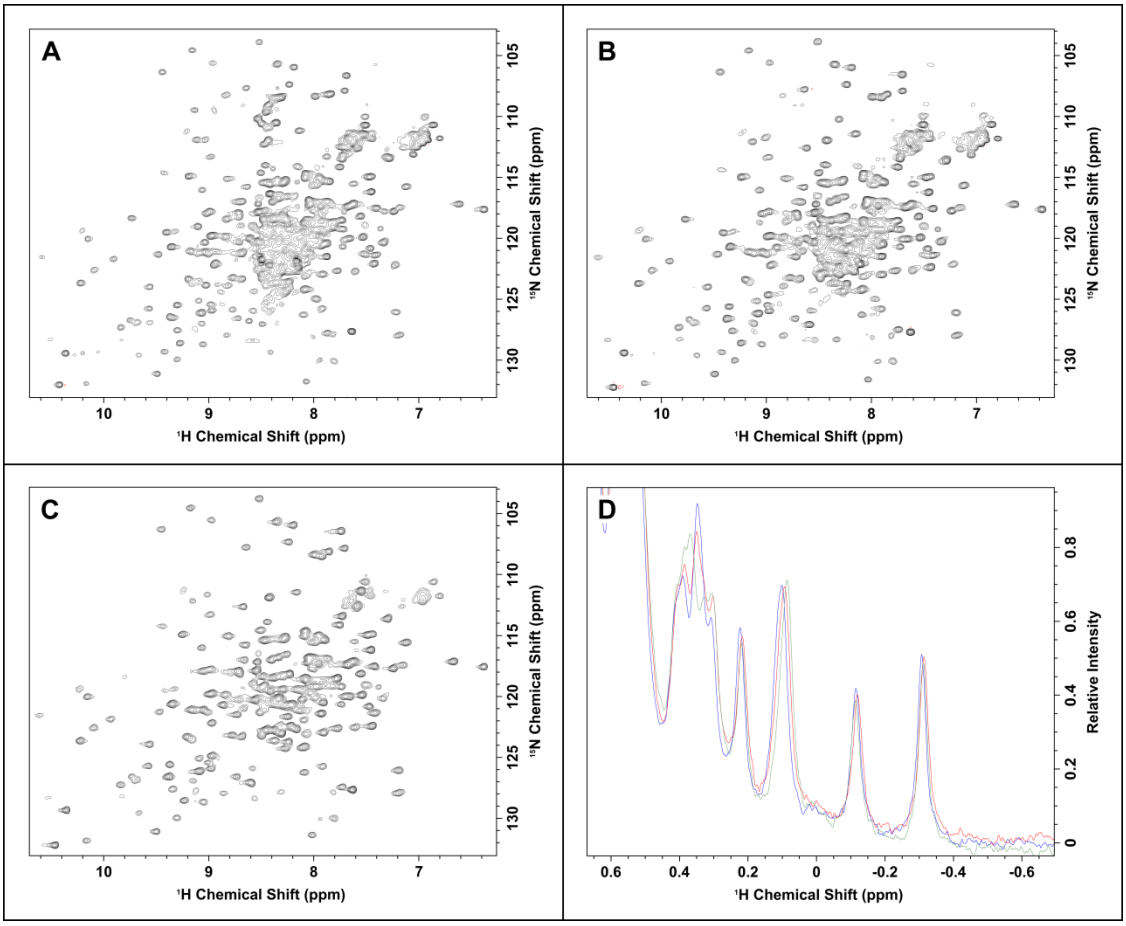

Figure 7.1 Effect of pH on NMR spectra of POR. ^{15}N-HSQC spectra are displayed at (A) pH 5.5; (B) pH 6.5; (C) pH 7.5. The number of signals present in the HSQC decreases as the pH is increased; however, the number of broad signals is reduced at higher pH, suggesting a more structured protein. (D) The upfield shifted methyl proton signals in the one-dimensional spectrum, indicative of methyl groups packed against aromatic side chains, are displayed at: pH 5.5, green; pH 6.5, red; pH 7.5, blue. These peaks are at comparable resolution at all pH values, indicating that the protein structure remains compact in all cases.

The one-dimensional ^{1}H experiments (Figure 7.1D) overlay extremely well, indicating no clear overall change in protein structure. The upfield shifted proton peaks, at chemical shifts

around or below 0 ppm, are representative of protons packed against the faces of aromatic side chains. These peaks are, therefore, good indicators of any disruption in protein structure as this would lead to changes in the proximity of these aromatic groups and thus would cause these peaks to be shifted downfield; it is clear from figure 7.1D that these peaks are not shifted upon changes in pH.

Based on these results, it is evident that the global structure of the protein is not affected by pH. POR does not exist fully in a molten globule state at pH 5.5. NMR experiments performed at pH 5.5 undeniably give an increase in observable signals; some of these are broad peaks and thus contribute to the lack of clarity of the HSQC spectrum. These broad signals are indicative of mobile structures in intermediate exchange, thus are likely to be from unstructured regions of POR; this makes these regions very interesting for studying the dynamic properties of the protein. It was therefore decided to continue to use POR samples at pH 5.5 for any future experiments to ensure that the maximum number of signals could be observed.

7.3 Production a of ^{2}H, ^{13}C, ^{15}N-labelled POR sample

The problems with studying large proteins by NMR are discussed in detail in 1.12; this highlights that one of the most effective techniques for improving NMR signals from large proteins is by the production of deuterated samples. The replacement of the majority of the protons with deuterons greatly reduces the relaxation effects on the other atoms, enabling an increase in observable signal; this is particularly beneficial for multidimensional experiments. It was decided that, in order to obtain useful NMR data for analysis of POR structure, a high quality, triple-labelled (^{2}H, ^{13}C, ^{15}N) sample must be produced. Backbone assignment experiments on this sample should facilitate assignment of signals to all residues in the protein, from which further experiments could enable structural information about POR to be determined.

7.3.1 Growth of *E. coli* in deuterated minimal media

Isotopic labelling of protein required the growth of *E. coli* cells, and subsequent POR expression, to be carried in minimal media containing labelled precursors; M9 media (detailed A1.3) was chosen as the preferred medium for POR expression, with ^{15}N-labelled ammonium sulphate ($[^{15}NH_4]_2SO_4$) and ^{2}H, ^{13}C-labelled glucose the labelled precursors. Growth in deuterated medium causes a large amount of stress to cells, due to the number of

cellular reactions involving the breaking and formation of bonds involving protons; replacing these with deuterons affects the regulation of cellular processes, and thus it takes a number of doubling times for the cells to adjust to the change. Therefore, it was decided to increase the levels of deuteration within the media stepwise, in order to ensure continued growth of the cells.

The pET-9a plasmid containing wild-type POR was transformed (detailed 2.5). A few colonies from the resultant agar plate were used to inoculate a 10 ml starter of M9 media containing $[^{15}NH_4]_2SO_4$ and suitable antibiotics (detailed A2.1-2.2) and left to incubate overnight at 37 °C with agitation. A second starter of 10 ml M9 media was prepared, containing 90 % deuterium oxide (D_2O) and 10 % water, along with $[^{15}NH_4]_2SO_4$, ^{2}H, ^{13}C-glucose, and suitable antibiotics; this starter was inoculated with 500 µl of the overnight starter and again left to incubate overnight at 37°C with agitation. This was repeated for a final starter, containing 100 % D_2O, $[^{15}NH_4]_2SO_4$, ^{2}H, ^{13}C-glucose and suitable antibiotics.

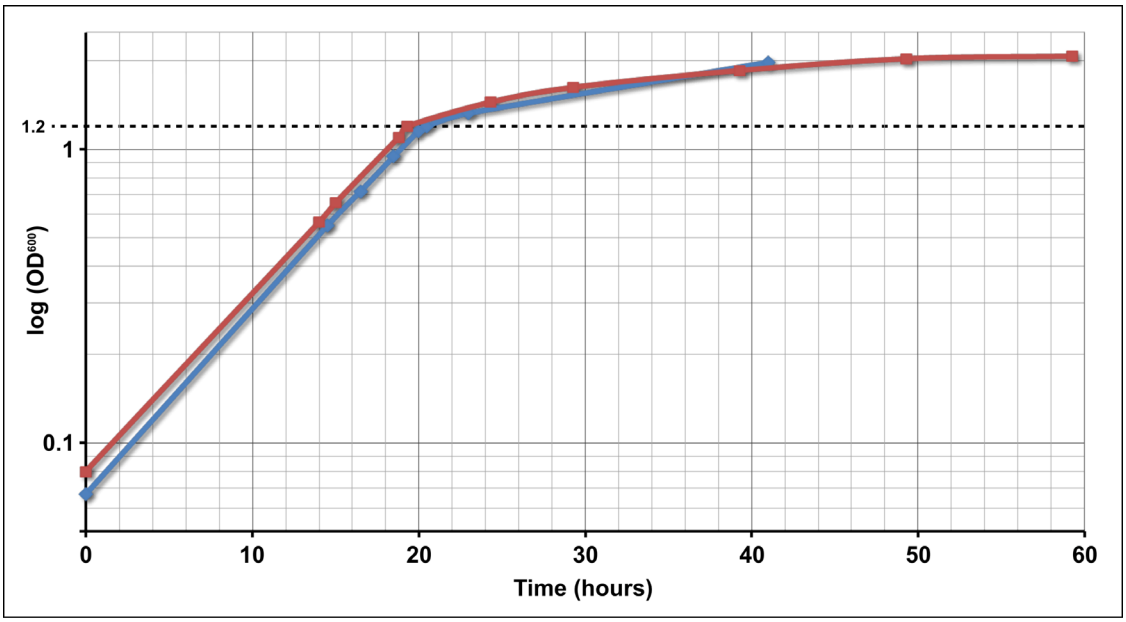

Figure 7.2 The growth of *E. coli* in deuterated M9 media is monitored as the change in optical density, at 600 nm, over time. Two separate growth curves are displayed, with data points represented with blue triangles and red squares.

The final starter was subsequently used to inoculate the final growth media, containing the same components as this starter. This was left to incubate at 37 °C with agitation, with absorbance of the medium, at 600 nm, monitored at regular intervals. Once the cell cultures had reached an optical density (OD) of 1.2, Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used in order to induce the expression of the POR protein; once induced, the cultures were

incubated at 20 °C, with cell density continuing to be monitored for a period of up to 40 hours. Two separate growth curves produced for cells grown in deuterated medium are shown in figure 7.2; these clearly show the exponential growth rate observed at 37 °C, with an OD of 1.2 reached after approximately 20 hours of growth. Subsequent to protein induction, the cells continue to grow, albeit at a slower rate, until approximately 30 hours of induction, at which point the absorbance reaches a maximum at an OD of approximately 2.

7.3.2 Optimisation of POR induction

In order to optimise the amount of protein produced from the deuterated POR preparation, samples were taken at various induction times to be analysed for protein content; cells were pelleted by centrifugation, broken by sonication, and the cell contents analysed by SDS-PAGE (detailed 2.18). The resultant gel (Fig. 7.3) indicates the levels of protein at each point after induction; the levels of POR increase steadily with induction time, while levels of the other proteins remain constant. The final sample, at 40 hours induction, has the highest levels of POR in comparison with the levels of other proteins, indicating a maximum expression level. It was therefore decided to induce POR expression for 40 hours, in order to obtain the maximum labelled POR for NMR experiments.

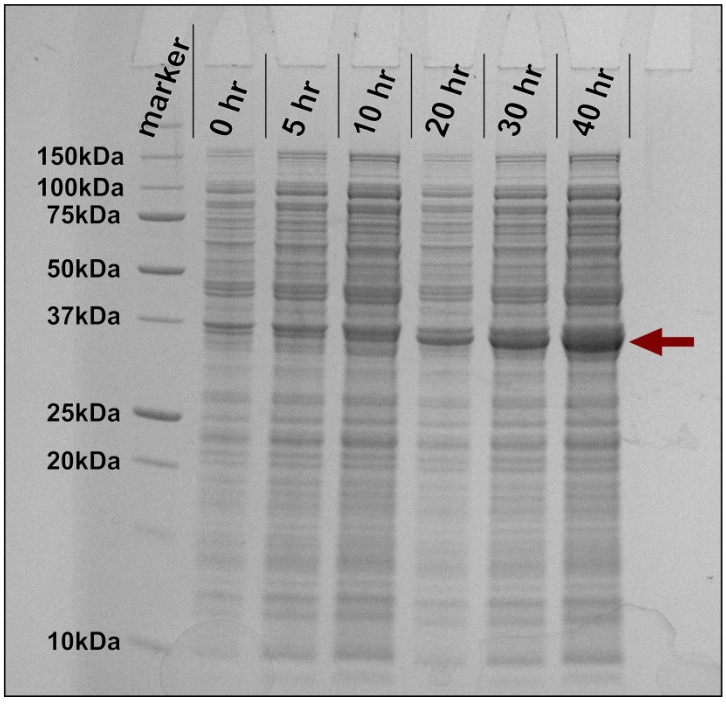

Figure 7.3 Protein gel indicating the protein expression levels over a number of different induction times. Protein standards are run in the first lane, with marker molecular weights indicated on the left. The position of POR in the gel is marked with a red arrow. The gel clearly indicates that the level of POR increases with induction time, while the levels of other proteins remain constant.

7.3.3 Optimisation of POR backbone amide exchange

Proteins grown in deuterated media will contain almost no protons. Though this is beneficial for reducing relaxation effects, the backbone amide groups are required to be protonated for backbone assignment experiments. Protons are significantly more sensitive in NMR than ^{13}C and ^{15}N nuclei and thus are the most common nuclei for initial resonance transfer and detection in NMR experiments; this is particularly key for multidimensional experiments that require long excitation periods. Amide protons exchange with the bulk solvent in favourable conditions, thus it is possible to replace the amide deuterons with protons in order to carry out the backbone assignment experiments.

In order to optimise the amide exchange for deuterated POR, while not wasting valuable deuterated samples, protein was expressed in cells grown in 100 % H_2O M9 media with $[^{15}NH_4]_2SO_4$, for subsequent exchange with D_2O-based buffers. Optimal amide proton exchange should occur at high pH and temperatures, though pH 7.5 is preferable to prevent any unfolding or hydrolysis of the protein; thus, ^{15}N-labelled POR was prepared in NMR buffer adjusted to pH 7.5 (adjusted for difference in water and deuterium oxide pK, see A1.3), placed in an NMR tube and placed in a Bruker Avance I 800 MHz machine at a temperature of 328 K (55 °C). ^{1}H, ^{15}N-HSQC experiments were run sequentially, each for a duration of 5 hours, in order to monitor the disappearance of signals, and thus the exchange of amide protons to deuterons (Fig. 7.4).

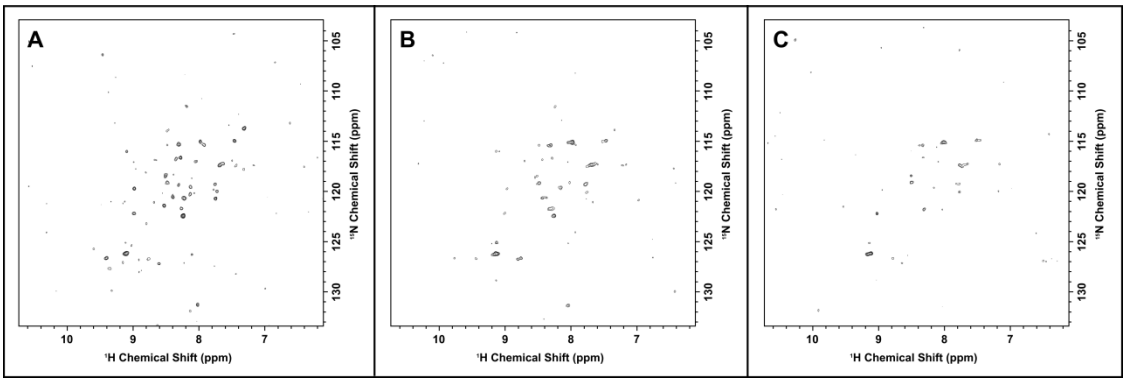

Figure 7.4 HSQC spectra indicating the number of amide signals remaining after deuterium exchange at pH 7.5 and 55 °C for (A) 5 hours, (B) 10 hours and (C) 15 hours. The number and intensity of signals steadily decreases throughout the experiments, however a number of signals are still present after 15 hours incubation.

After the third consecutive HSQC experiment (Fig. 7.4C), representative of 15 hours of incubation at 55 °C, no further loss of signals was observed; there were, however, still a number of signals present. These were likely to be from buried amide groups and, with these exchange conditions, would not be observed in the deuterated sample; therefore a more

rigorous exchange method was required, in order to destabilise the structure to enable these protons to exchange and ensure the observation of maximum signals in the final sample. Exchange conditions of pH 7.5 and 60 °C were also tested, yet a number of signals still remained; POR was observed to precipitate out of solution at temperatures higher than 60 °C, therefore an alternative technique for destabilisation of the protein structure was required.

Previous work, looking at the stability of POR at various denaturant conditions, had shown that POR was stable in Guanidinium Hydrochloride (GuHCl) samples, up to a concentration of 2 M GuHCl at 4 °C (Proudfoot, 2011). It was therefore decided that the best method for exchange would involve a combination of GuHCl and temperature; POR amide exchange was carried out at varying GuHCl concentrations (0.5 M, 1 M, 1.5 M and 2 M) at pH 7.5 and 55 °C. Both the 1.5 M and 2 M GuHCl samples exhibited significant precipitation, with the majority of protein lost, yet the 0.5 M and 1 M samples remained in solution.

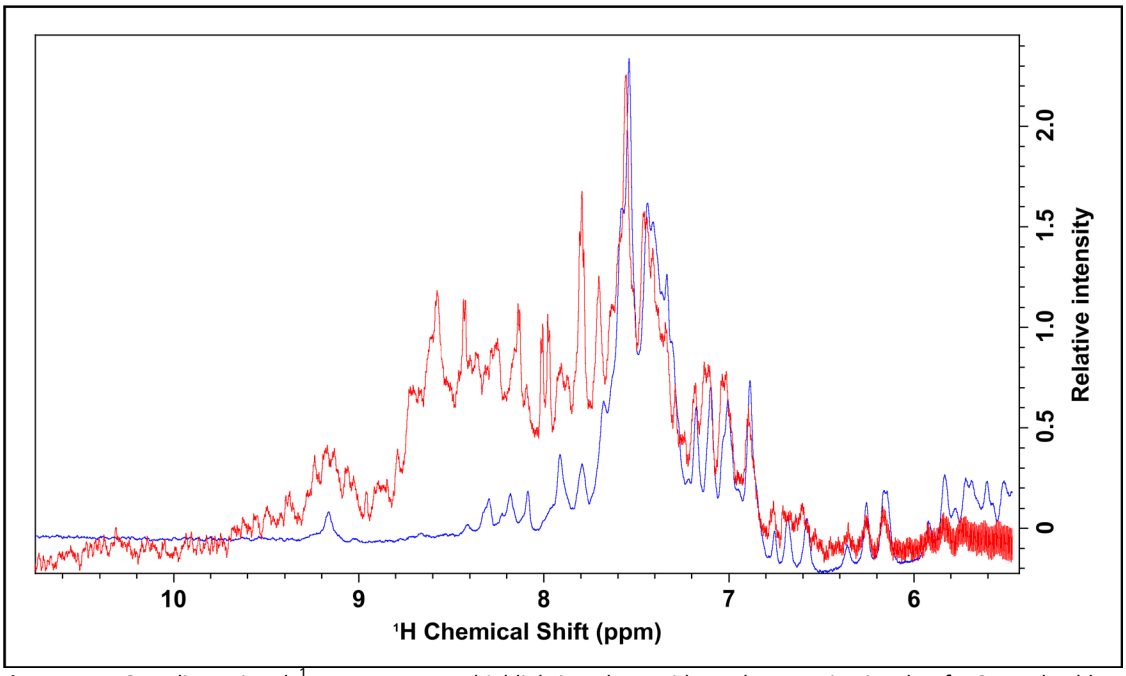

Figure 7.5 One-dimensional ^{1}H NMR spectra highlighting the amide and aromatic signals of POR. The blue spectrum is from the sample incubated at pH 7.5 and 55 °C, with 1 M GuHCl. The red spectrum is from a fully protonated POR sample, as a reference. The samples are scaled to ensure identical intensity of the non-exchanged aromatic protons. The amide signals in the amide exchange sample have almost completely disappeared, save a few small peaks, indicating a high level of exchange to deuterons.

NMR experiments were carried out on the 1 M GuHCl amide exchange sample; HSQC experiments did not give any visible signals, indicating a good level of exchange of the amide groups. The one-dimensional ^{1}H spectrum from this sample was overlaid with a fully protonated sample of POR (Fig. 7.5); the aromatic peaks in both spectra, at 7 - 7.5 ppm,

overlay very well, while almost all of the amide signals, visible at 7.5 - 9.5 ppm, have disappeared, indicating thorough exchange to deuterons. Therefore, it was decided that the exchange step should be carried out in conditions of pH 7.5, 55 °C and 1 M GuHCl to ensure maximum exchange of amide deuterons to protons.

7.3.4 Comparison of sample with ^{15}N-labelled POR

Optimisation of the conditions required for production of ^{2}H, ^{13}C, ^{15}N-labelled POR led to the production of a high quality POR sample for NMR. After exchange of amide deuterons to protons, the sample was transferred to NMR buffer and was concentrated to approximately 300 µl, giving a final concentration of approximately 150 µM. The sample was subsequently placed in a Shigemi NMR tube; this tube contains a glass plunger which mimics extension of the sample volume, enabling smaller sample volumes without compromising the shimming ability, thus ensuring homogeneity of the magnetic field across the sample. Initial experiments were conducted on this sample at 318 K, using a Bruker DRX 600 MHz machine, fitted with a TXI-cryoprobe.

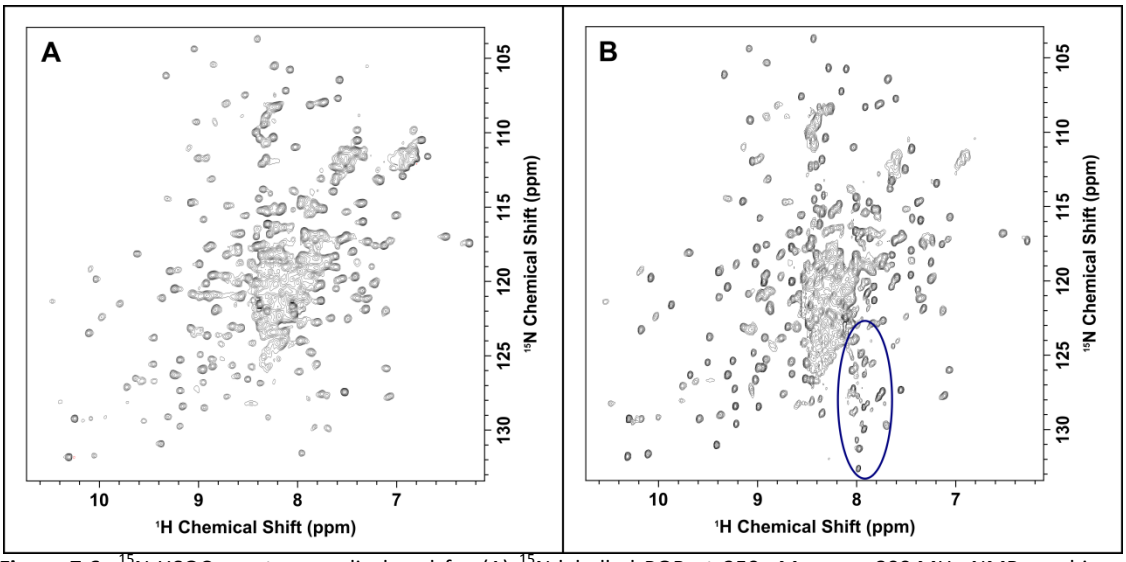

Figure 7.6 ^{15}N-HSQC spectra are displayed for (A) ^{15}N-labelled POR at 250 µM on an 800 MHz NMR machine, (B) ^{2}H, ^{13}C, ^{15}N-labelled POR at 150 µM on a 600 MHz NMR machine. A TROSY form of the HSQC experiment is run for the triple labelled sample. Despite the lower protein concentration and field strength for the triple-labelled sample, the signals in the respective HSQC spectrum are more intense and sharper in the proton dimension, indicating a clear improvement in the sample for NMR experiments. This sample had been subjected to high temperatures for a number of experiments leading to some cleavage products, indicated by a blue circle.

Comparison of HSQC spectra, between the sample described above and a ^{15}N-labelled POR sample (approximately 250 µM concentration, run at 318 K at 800 MHz), indicates thorough amide exchange, with no peaks apparently missing from the triple labelled sample (Fig. 7.6).

Despite the lower concentration and field strength, the HSQC from the triple labelled sample has higher peak intensities, along with a reduced linewidth in the proton dimension. There are, however, a number of new signals in the lower right region of the spectrum, indicating a small number of cleavage products within the sample; it was therefore ensured, for all subsequent samples, that sufficient protease inhibitor was present. It is clear, however, that this protocol has facilitated the production of a high quality, triple labelled sample; the spectra obtained from this sample are noticeably improved upon even the highest quality single labelled sample, indicating that a triple labelled sample represents the best possibility for a full NMR assignment of POR.

7.4 Backbone assignment experiments on ^{2}H, ^{13}C, ^{15}N-labelled POR

7.4.1 Initial backbone assignment experiments at 600 MHz

Backbone assignment experiments enable assignment of NMR chemical shifts to atoms from all residues throughout the protein; these are multidimensional experiments and are often referred to as 'out-and-back' experiments, as magnetisation is transferred from the amide proton out to an observed carbon atom, and back to the proton for detection. Schematic representations of the experiments used in this assignment are presented in Figure 7.7.

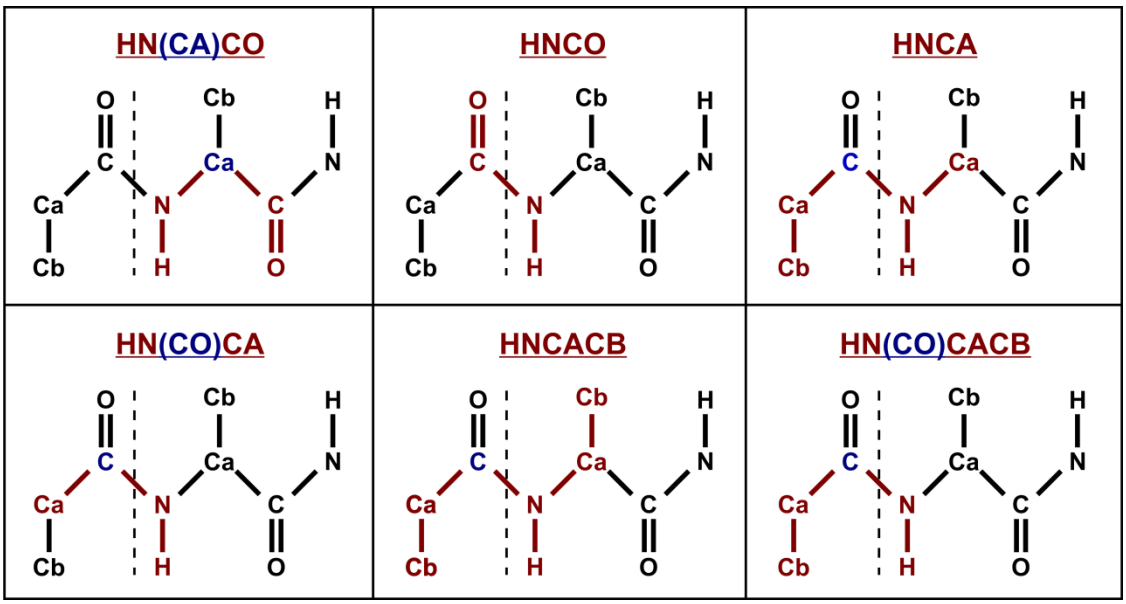

Figure 7.7 Schematic diagrams representing the transfer of magnetisation throughout the protein backbone in each of the backbone assignment experiments. Atoms left and right of the dotted line are part of the previous and current residue, respectively. Each of the experiments starts at the amide proton; transfer of magnetisation is along the red bonds; atoms in the path of resonance transfer, but which are not detected, are coloured blue. HNCA and HNCACB experiments include signals for both the current and previous residue.

A full set of backbone experiments was carried out on the sample prepared in 1.3.4, run at 313 K on the 600 MHz machine; these experiments, however, did not yield high quality spectra, due to the limited concentration of the sample and the relatively low field strength of the magnet. Attempts to assign these spectra were unproductive; therefore it was decided to prepare a new sample, with a higher concentration, and to carry out experiments on a NMR machine with increased field strength.

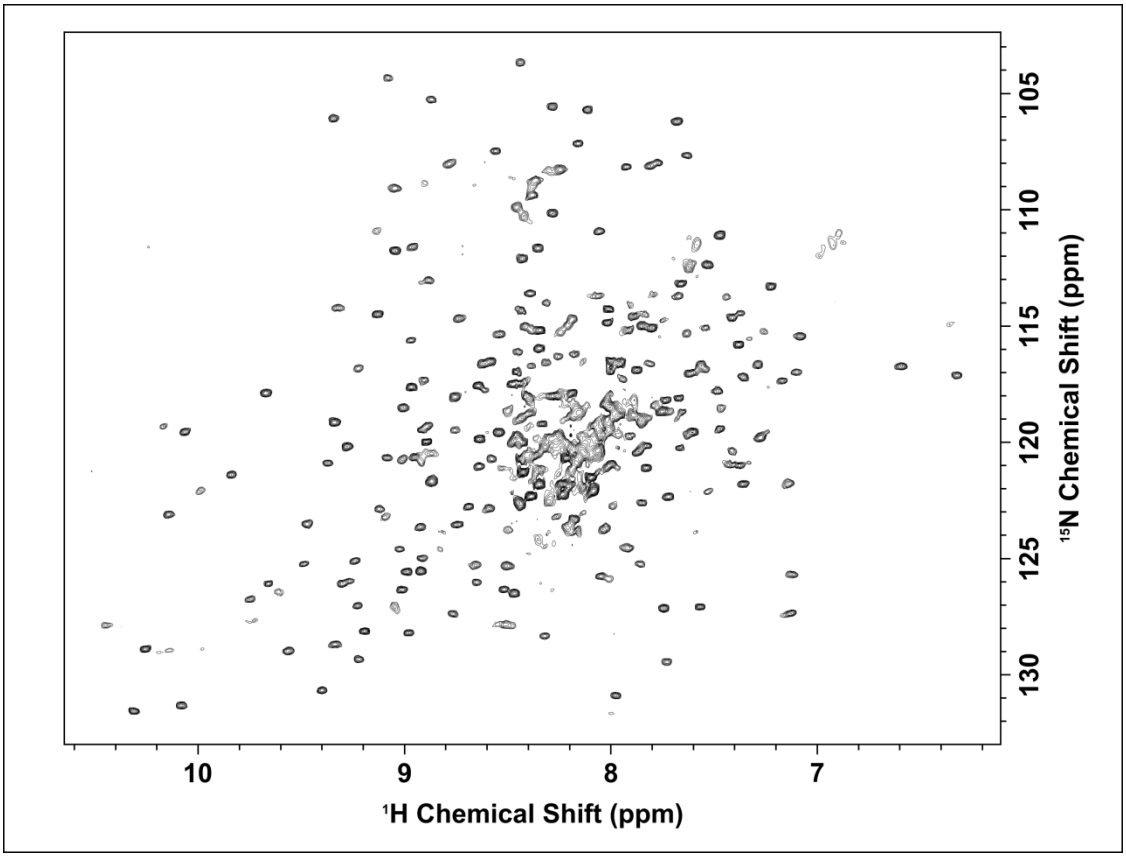

Figure 7.8 TROSY-HSQC spectrum for the triple labelled POR sample, run on a 900 MHz NMR machine. The intensity and distribution of the signals is very good. There are also markedly less broad signals apparent, in the centre of the spectrum, than observed at lower field strengths. Cleavage products apparent in the previous sample (Fig. 7.6) are not present in this sample.

7.4.2 Backbone assignment experiments at 900 MHz

In order for the best possible chance of a full assignment of POR, backbone assignment experiments would need to be carried out at the highest field strength on a single sample; the use of a single sample for all experiments should ensure no variation in chemical shifts and more reliable assignment of peaks. A new triple labelled sample of POR was prepared, with concentration of 300 µM, placed in a Shigemi tube and taken to Utrecht University; a full set of backbone assignment experiments (detailed Table 2.3) were run on this sample by Dr Hans Wienk, using a 900 MHz Bruker Avance II machine, fitted with a TCI-cryoprobe. The HSQC spectrum run at 900 MHz is shown in Figure 7.8; the intensity of peaks is very high and

the amount of overlap in the centre of the spectrum is much reduced compared to the spectra in Figure 7.6. There are also no peaks at the bottom right of the spectrum, which would represent likely cleavage products, therefore indicating a high quality sample.

7.4.3 Peak Picking

All data obtained from the Utrecht experiments was processed in Sheffield, using Felix 2007 and in-house macros. All peaks from the TROSY-HSQC spectrum that could be automatically picked were done so, using the automatic peak picking function of Felix. The nitrogen and proton chemical shift values for these peaks were outputted to an entity in the master database with the peaks assigned a spin system number. This list, in combination with in-house macros, was used to analyse all other backbone assignment spectra and select all peaks throughout the experiments which corresponded to a single spin system. Proton-carbon planes were loaded for all spectra, in the nitrogen plane corresponding to the nitrogen chemical shift of that particular spin system. In heavily overlapped regions, slices were taken through peaks in the carbon plane to ensure that all peaks picked corresponded to the correct spin system.

Spin systems corresponding to peaks on the periphery of the HSQC spectrum were generally picked quite straightforwardly, due to the limited overlap with other signals in this region (Figure 7.9). Peaks within the centre of the HSQC spectrum, however, exhibited much more significant overlap with neighbouring signals in all spectra; thus assigning these peaks to spin systems presented much more difficulty (Figure 7.10). Heavily overlapped peaks in the HSQC spectrum, which had well resolved peaks in the other spectra but had not been selected in the automatic peak picking process, were manually picked in the HSQC spectrum and subsequently added to the entity list.

In many of the experiments, despite being set up for observation of the carbon shift from the current or previous residue, peaks were also present for the alternative atom; for example the HNCA spectra often contained peaks for the Cα of the previous residue (denoted pCα), as well as for the Cα of the current residue (denoted Cα). In these cases, both sets of peaks were picked, labelled depending on their association with the current or previous residue; these could also be helpful in confirmation of peaks in the spin system in crowded and difficult regions. The HNCACB and HN(CO)CACB experiments contain data for both Cα and Cβ shifts, with negative and positive peaks, respectively, corresponding to these atoms. In total, 443 spin systems were produced from the spectra; based on the reliability of shifts from the highest quality spectra, only spin systems containing Cα, pCα and pCO shifts were retained, with 379 spin systems remaining for the subsequent assignment.

Figure 7.9 and Figure 7.10

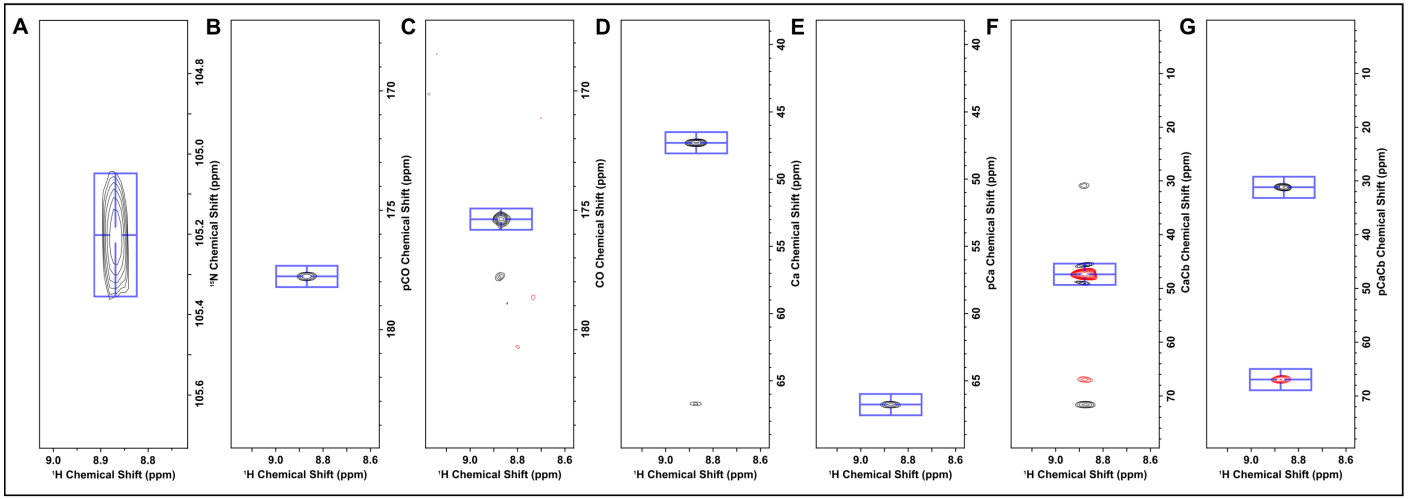

Figure 7.9 Strips of the backbone assignment spectra are displayed for a single spin system: (A) HSQC; (B) HNCO; (C) HN(CA)CO; (D) HNCA; (E) HN(CO)CA; (F) HNCACB; (G) HN(CO)CACB. Peaks picked in the spin system are highlighted with a blue box. Positive and negative signals are displayed in black and red, respectively.

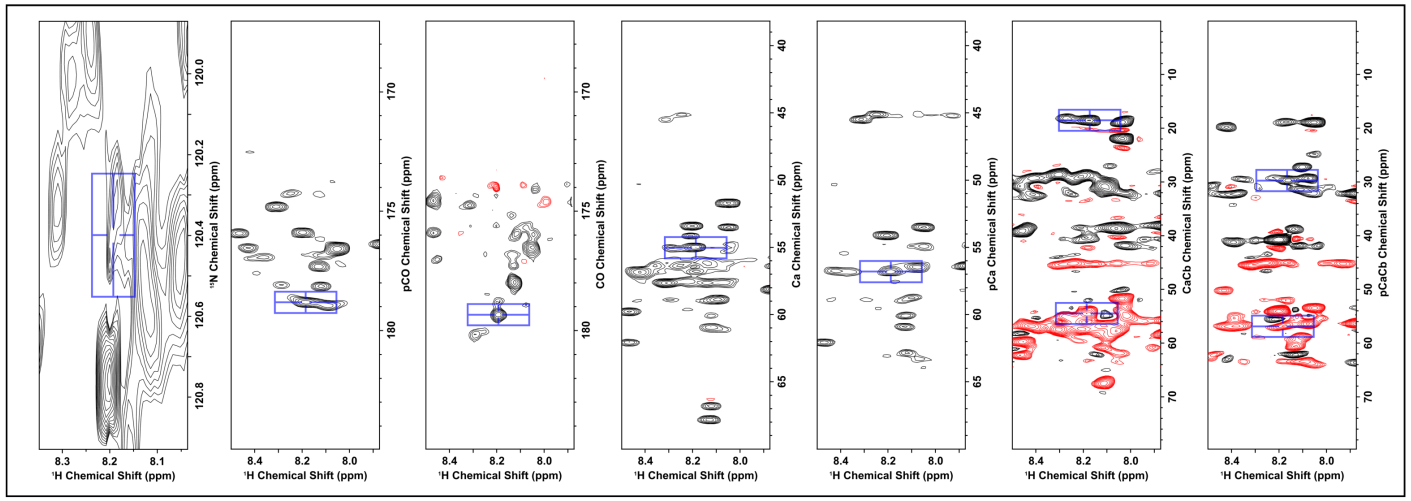

Figure 7.10 Spectra displayed as in Fig. 7.9, but highlighting a spin system in an environment with more significant overlap. The number of signals present at very similar resonance frequencies causes difficulties in assignment of peaks to the correct spin system.

7.4.3 Assignment

The spin system list, using the POR sequence as a template, was run through the Asstools simulated annealing assignment programme (Reed et al., 2003) with 30 iterations; a new randomly generated starting seed would be used each time (detailed 2.22). Asstools uses chemical shift ranges expected for each amino acid, independent of the secondary structure and with data taken from work published by (Grzesiek and Bax, 1993), in addition to the spin system list, in order to assign the spin systems to a suitable amino acid. Asstools uses a simulated annealing Metropolis Monte Carlo algorithm to minimise an energy function; the equation is made up of two components, one referring to the linking energy and the other to the binding energy. The linking energy describes the quality of the match between two sequential chemical shifts, while the binding energy describes how well the chemical shifts match the predicted chemical shifts detailed above.

Asstools uses a power law with a cut off to penalise incorrect matches in the linking energy; this means that as the difference between two incorrect chemical shifts increases, the energy penalty applied to the system increases. This makes it more favourable for the system to place a void spin system in that space rather than an incorrect assignment. The process is repeated whereby a new configuration is generated based on the current configuration, with the new output accepted if it has a lower total energy function than the previous configuration. If, however, the new configuration has a higher total energy function, then it is accepted with a probability based on the difference in energy between the two systems, along with the current temperature of the system. This process is repeated at a number of decreasing temperatures until a configuration with no higher or lower energy is obtained in three successive cycles. At this stage the whole process is repeated starting with a new randomly generated starting seed and it is this process which is repeated 30 times in an Asstools assignment (Reed et al., 2003).

The results output, after completion of the Asstools programme, identifies which spin systems were selected for each residue, along with the number of iterations out of the 30 in which each spin system was selected. This data was analysed manually to determine the reliability of these assignments, as well as identifying any obvious errors in the peaks picked from the original data; the original data was revisited at this point in order to rectify any errors in peak picking that had been identified.

Once the data had been thoroughly analysed, the Asstools programme was run again, with any confident assignments fixed. This process was repeated, until no further confident assignments could be made. Throughout this process, a number of further data

improvement techniques were employed to enable improvements in the assignment: spectra were reprocessed with a more resolution enhancing window function; spectra were analysed in the carbon plane, in order to ensure no carbon signals had been missed; scripts were used which predict spin systems that could feasibly follow and precede a given system, helping identify likely sets of linked residues.

The final assignment (Table 7.1), once all possible improvements to the assignment appeared to have been exhausted, included 217 confident assignments out of a possible 303 (excluding prolines); in addition, the previous residue information gave shifts for carbon atoms of 11 proline residues. The level of assignment in the first 200 residues is good, with almost 90 % of the non-proline residues assigned and with 9 prolines in this region with carbon shifts from the following residue. The assignment within the POR loop region is fair, with most of the latter part of the loop assigned; the assignments within the earlier part of the loop region are more scattered, though there are a number of proline residues in this region which are likely to affect the sequential assignment.

The assignment of the remainder of the protein is quite poor, however, with only around 50 % of assignments in the final 125 residues; this includes two large unassigned regions from residue 226 - 253 and 292 - 313. This difficulty in assignment is surprising, considering the number of spin systems which are present; it could suggest other problems with the data, for example multiple peaks present for certain residues due to intermediate exchange between structures. The first of these large gaps includes a repeating sequence PLFXXXXPLFXXXXPLF, which may hinder the assignment due to the regularity of proline residues and the similarity of the shift patterns.

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
2	S	104	8.34	115.97	58.25	55.66	63.52	32.49	174.24	176.31
3	D	225	8.34	121.89	54.16	58.25	40.86	63.70	175.90	174.26
4	Q	202	8.20	120.75	53.43	54.15	28.51	40.92	174.06	175.90
8	T	137	8.97	117.66	63.34	61.80	70.01	30.96	174.36	175.94
9	V	289	9.30	126.21	56.38	63.33	33.85	70.22	173.20	174.34
10	I	286	8.04	125.88	59.94	56.34	40.71	33.90	174.54	173.18
11	I	288	8.65	126.15	60.27	59.91	40.15	40.85	175.98	174.56
12	T	85	8.41	115.04	60.60	60.29	68.90	40.00	175.62	175.97
13	G	48	9.05	111.72	48.30	60.64	_	69.01	176.25	175.61
14	A	119	6.56	116.76	52.68	48.30	18.79	_	173.82	176.25
15	S	7	7.66	106.10	59.38	52.66	63.61	18.97	172.87	173.81
16	S	35	7.45	111.05	57.02	59.35	65.43	63.74	174.47	172.85
17	G	23	9.05	109.01	46.07	56.97	_	65.45	174.48	174.45
18	V	252	8.74	123.64	66.77	46.11	31.13	_	177.78	174.47
19	G	3	8.87	105.15	47.29	66.79	_	31.29	175.40	177.78
20	L	285	7.10	125.81	57.21	47.31	40.00	_	178.62	175.39
21	Y	130	7.93	117.30	62.91	57.18	37.64	40.21	179.63	178.64
22	A	360	9.28	126.06	54.53	62.92	17.37	37.70	178.70	179.67
23	T	332	7.59	116.98	67.42	54.48	_	17.43	176.05	178.70
24	K	190	7.65	120.31	59.36	67.43	31.92	_	177.17	175.99
25	A	178	8.04	119.79	54.66	59.33	18.49	31.72	181.27	177.17
26	L	158	8.39	118.79	57.93	54.64	39.11	18.63	179.50	181.25
27	A	359	8.98	125.88	55.15	57.91	17.76	39.29	182.81	179.48
28	N	334	8.48	117.56	55.21	55.14	37.84	18.09	177.13	182.81
29	R	139	7.46	117.82	56.15	55.20	31.04	37.96	176.45	177.14
30	G	8	8.15	107.06	45.48	56.10	_	31.22	174.80	176.42
31	W	188	8.31	120.25	57.67	45.47	28.60	_	175.68	174.81
32	H	219	9.85	121.47	54.72	57.66	30.79	28.65	174.16	175.67
33	V	310	8.98	128.33	60.60	54.70	32.69	30.85	175.90	174.16
34	I	290	9.67	126.21	57.78	60.58	34.39	32.92	175.20	175.90
35	M	320	9.23	129.49	54.82	57.77	33.77	34.52	173.30	175.18
36	A	325	7.96	131.07	50.34	54.79	19.76	33.33	176.96	173.31
37	C	204	8.42	120.80	57.00	50.34	32.21	19.90	173.67	176.96
40	L	277	8.65	125.41	58.69	51.92	40.37	39.10	179.06	175.33
41	E	176	8.54	119.63	59.64	58.68	28.54	40.52	179.68	179.00
42	K	162	8.00	119.10	58.47	59.63	31.71	28.64	179.56	179.69
43	A	212	7.81	121.17	55.25	58.42	19.67	31.86	177.84	179.51
44	E	141	8.18	117.94	59.08	55.26	28.52	19.89	179.03	177.80
45	Q	123	7.86	116.92	58.30	58.98	28.10	28.91	178.42	179.01
46	A	378	8.22	122.44	54.80	58.28	18.52	27.97	179.91	178.43
47	A	381	8.25	119.76	55.00	54.77	16.66	18.76	180.12	179.91
48	K	122	7.98	116.92	59.07	54.99	31.65	16.77	181.60	180.15

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
49	N	181	8.45	119.86	55.68	59.06	38.20	32.14	177.09	181.60
50	L	145	7.65	118.14	54.44	55.68	41.76	38.11	175.78	177.08
51	Q	81	8.00	114.86	56.38	54.39	25.16	41.69	175.81	175.79
52	I	186	7.81	120.21	59.85	56.31	37.37	25.23	174.31	175.73
55	E	74	9.13	114.49	57.51	64.68	28.01	31.10	176.94	178.25
56	A	211	8.64	121.11	51.90	57.48	20.94	28.18	174.55	176.94
57	Y	61	7.20	113.29	56.00	51.90	40.78	21.13	173.53	174.57
58	T	110	8.58	116.55	62.55	55.96	72.13	40.86	173.61	173.52
59	I	309	9.20	128.26	61.19	62.55	37.91	72.38	175.17	173.69
60	L	292	8.51	126.46	52.95	61.16	46.86	38.02	175.29	175.15
61	H	183	8.89	120.05	57.05	52.94	30.49	47.08	174.86	175.32
62	L	303	7.73	127.15	54.94	57.02	43.79	30.66	172.68	174.86
63	D	297	8.46	126.63	51.13	54.91	40.39	43.96	175.86	172.68
64	L	354	8.38	122.49	56.78	51.14	42.47	40.20	175.83	175.88
65	S	45	8.35	111.61	59.10	56.76	64.04	42.69	171.77	175.80
66	S	86	7.79	115.08	53.86	59.08	63.85	64.18	175.46	171.76
67	L	326	10.10	131.49	58.54	53.86	38.87	63.93	179.80	175.45
68	A	187	9.28	120.25	55.04	58.52	17.78	39.01	181.70	179.81
69	S	83	7.83	115.00	60.13	55.02	62.15	17.90	178.15	181.69
70	V	351	8.11	120.74	66.80	60.06	30.99	62.30	_	178.15
71	R	113	7.94	116.55	60.47	66.82	29.30	30.86	179.79	178.07
72	G	9	8.55	107.38	46.95	60.45	_	29.32	176.88	179.79
73	F	274	9.24	125.23	60.95	46.96	38.72	_	177.29	176.88
74	V	195	8.13	120.42	67.80	60.93	30.50	38.91	178.06	177.31
75	E	146	7.72	118.24	59.57	67.84	28.69	30.48	179.85	178.01
76	S	134	8.46	117.50	61.84	59.54	62.88	28.71	176.42	179.88
77	F	353	8.87	121.87	60.31	61.80	37.85	62.77	179.19	176.39
78	R	240	8.69	122.86	59.77	60.27	29.07	37.82	178.58	179.19
79	A	172	7.45	119.49	53.25	59.77	17.37	29.19	178.34	178.60
80	L	77	7.39	114.64	55.38	53.25	40.67	17.69	177.56	178.35
81	N	60	7.64	113.16	54.10	55.34	36.92	40.73	_	177.50
82	R	135	8.64	117.60	52.28	54.07	31.60	37.06	173.72	173.69
84	L	283	8.92	125.66	54.94	61.96	43.75	31.49	175.60	177.17
85	R	272	8.91	125.11	57.05	54.92	31.33	43.89	173.86	175.59
86	A	75	7.88	114.60	50.67	57.05	23.46	31.57	_	173.87
87	L	215	8.18	121.24	53.49	50.68	46.13	23.58	173.24	174.88
88	V	275	9.50	125.35	60.10	53.48	32.53	45.82	175.86	173.24
89	C	284	8.99	125.60	57.81	60.08	25.73	32.90	173.19	175.82
90	N	276	7.84	125.36	54.30	57.75	42.30	25.85	173.16	173.17
91	A	356	8.02	123.98	53.56	54.26	18.98	42.47	176.56	173.17
92	A	192	8.05	120.41	51.72	53.52	21.90	19.02	176.98	176.58
93	V	169	10.18	119.36	59.61	51.74	33.79	22.05	174.17	176.94

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
94	Y	153	7.45	118.58	55.30	59.62	39.88	33.96	173.50	174.15
95	Y	302	9.05	127.21	54.44	55.26	38.36	40.09	173.50	173.48
97	L	405	7.46	111.05	57.66	65.15	51.63	30.19	175.07	179.64
98	L	419	7.74	118.77	57.86	56.05	_	_	173.23	175.00
102	I	348	8.46	120.05	59.80	62.09	39.05	30.85	174.61	175.94
103	Y	355	8.43	122.87	57.25	59.77	40.04	39.43	176.66	174.63
104	S	174	10.08	119.61	56.95	57.24	65.89	40.33	177.04	176.65
105	V	165	9.34	119.19	64.43	56.88	30.87	66.00	176.66	177.04
106	D	118	7.54	116.71	54.38	64.43	41.59	31.07	174.71	176.63
107	G	5	8.10	105.59	44.66	54.32	_	41.70	172.99	174.72
108	Y	129	7.34	117.23	53.89	44.67	41.93	_	176.32	172.99
109	E	318	9.57	129.13	60.01	53.87	28.89	42.04	178.99	176.33
110	I	278	8.50	125.44	65.42	59.97	39.52	28.96	176.99	178.95
111	T	299	9.76	126.89	68.09	65.40	66.81	39.61	_	177.01
112	V	251	9.48	123.62	66.99	68.09	31.90	67.05	180.49	177.74
113	A	245	10.16	123.20	55.44	67.00	19.02	32.07	178.71	180.49
114	T	72	8.43	114.32	65.38	55.43	69.11	19.21	173.85	178.70
115	N	79	8.18	114.60	55.96	65.38	37.64	69.17	179.42	173.85
116	H	330	8.37	115.14	57.29	55.94	32.16	37.74	175.32	179.46
117	L	161	8.42	118.86	59.20	57.28	38.86	32.40	178.48	175.32
118	G	6	9.35	105.97	47.88	59.19	_	38.95	176.88	178.50
119	H	127	6.29	117.15	61.21	47.90	28.40	_	176.06	176.87
120	F	207	9.38	120.97	62.77	61.18	38.65	28.37	178.54	176.10
121	L	203	8.57	120.78	58.42	62.76	39.86	38.89	176.46	178.55
129	D	160	7.95	118.82	57.12	59.23	40.51	28.97	179.92	177.59
130	L	170	8.89	119.37	57.47	57.07	42.37	40.33	179.26	179.91
131	K	200	8.95	120.71	59.95	57.46	31.60	42.44	177.98	179.30
132	N	71	8.00	114.27	53.67	59.92	38.51	31.76	_	177.98
133	S	101	7.36	115.81	54.41	53.65	64.15	38.66	174.72	175.59
135	E	175	7.59	119.63	56.59	63.07	29.40	31.00	175.98	176.90
136	S	89	8.34	115.19	58.90	56.58	64.06	29.48	174.65	175.99
137	D	238	8.44	122.54	52.80	58.88	39.91	64.16	174.65	174.63
138	K	233	8.38	122.30	55.20	52.78	33.80	40.10	177.86	174.36
139	R	150	8.02	118.36	54.46	55.14	33.88	33.95	172.79	177.86
140	L	300	9.23	127.15	55.26	54.49	44.45	33.99	174.86	172.78
141	V	268	9.02	124.71	60.24	55.24	34.74	44.56	174.86	174.84
142	I	305	8.76	127.51	60.36	60.23	40.92	34.95	174.65	174.85
143	L	322	7.71	129.60	54.33	60.40	39.92	40.92	177.46	174.62
144	G	31	9.14	110.87	43.84	54.29	_	39.67	172.15	177.47
145	T	54	7.60	112.44	59.83	43.92	69.77	_	172.64	172.14
146	V	385	7.70	118.72	62.21	59.80	31.87	70.17	_	172.68
147	T	392	7.80	122.45	62.92	62.28	70.57	32.21	179.05	175.75

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
154	G	19	8.36	108.68	45.41	55.18	_	41.71	174.81	178.03
155	G	16	8.24	108.19	45.13	45.44	_	_	174.33	174.82
156	K	380	8.26	120.26	56.34	45.13	32.10	_	176.18	174.29
157	I	194	7.85	120.46	58.10	56.39	38.15	32.32	174.63	176.38
161	A	363	8.56	128.07	52.05	61.29	18.79	32.37	176.82	176.12
168	L	57	8.88	113.01	56.17	51.90	37.98	38.03	178.15	175.70
169	E	217	10.53	121.30	61.69	56.15	29.42	38.12	179.41	178.15
170	G	2	9.08	104.20	46.86	61.68	_	29.63	176.95	179.42
171	F	193	7.39	120.44	58.40	46.88	37.94	_	179.49	176.95
172	E	242	9.12	122.96	59.43	58.37	28.57	38.10	178.84	179.50
173	K	65	7.66	113.69	56.12	59.39	32.72	28.95	177.53	178.77
174	G	15	7.91	108.08	45.75	56.12	_	33.02	173.42	177.56
175	F	124	8.45	116.98	58.88	45.76	41.88	_	173.42	173.43
176	K	109	7.99	116.55	53.77	58.83	33.50	42.05	177.23	173.14
177	K	399	8.09	122.38	55.15	53.71	31.96	33.63	175.27	177.21
179	I	324	9.41	130.84	63.63	63.87	36.13	33.03	176.72	176.47
180	A	314	9.34	128.84	51.64	63.56	21.50	36.26	_	176.75
181	M	52	7.51	112.33	52.69	51.63	35.16	21.75	179.00	174.91
182	I	308	10.47	128.01	66.28	52.74	35.52	35.36	_	179.00
183	N	94	8.53	115.36	51.78	66.25	37.04	35.85	176.73	177.27
184	G	4	8.28	105.45	45.44	51.80	_	37.18	173.63	176.72
185	K	210	7.36	121.07	55.31	45.44	29.86	_	173.51	173.63
187	F	182	8.63	119.92	58.49	63.89	39.73	31.87	175.99	176.27
188	K	304	7.10	127.49	54.99	58.48	35.31	39.86	176.08	175.96
189	S	222	8.87	121.64	61.73	54.94	62.91	35.42	177.62	176.12
190	G	21	8.90	108.78	47.19	61.79	_	63.11	176.60	177.60
191	K	223	7.11	121.86	59.02	47.21	31.68	_	176.68	176.62
192	A	241	8.58	122.93	55.97	58.98	17.56	31.80	180.30	176.75
193	Y	88	7.90	115.18	61.71	55.96	37.48	17.69	179.02	180.28
194	K	239	7.98	122.83	60.01	61.71	31.12	37.61	179.27	179.02
195	D	229	10.00	122.17	57.21	59.96	39.78	31.13	179.84	179.30
196	S	335	8.61	117.77	62.60	57.18	_	39.84	172.16	179.84
207	L	197	8.92	120.92	57.59	57.68	40.94	31.64	178.43	178.52
208	H	87	7.52	115.08	60.52	57.55	30.84	40.88	177.83	178.41
209	R	168	8.07	119.32	59.50	60.51	30.07	30.97	_	177.80
210	R	63	8.38	113.58	58.55	59.49	29.68	30.24	179.82	178.59
211	F	32	8.04	110.88	57.90	58.53	41.60	29.74	178.98	179.82
212	H	253	8.92	123.75	61.92	57.89	29.66	41.75	177.15	178.97
213	E	151	9.01	118.56	59.43	61.92	28.75	29.70	178.40	177.15
214	S	28	8.27	110.09	59.63	59.40	63.09	28.66	176.67	178.41
215	T	14	7.79	108.02	62.67	59.63	72.05	63.22	176.30	176.67
216	G	50	8.42	112.06	45.78	62.63	_	72.37	174.31	176.32

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
217	I	248	8.17	123.43	61.75	45.78	37.55	_	175.39	174.29
218	V	312	8.31	128.47	62.74	61.78	32.49	37.65	173.16	175.36
219	F	294	9.01	126.48	56.52	62.74	41.17	32.70	174.98	173.21
220	N	205	9.01	120.80	52.07	56.51	42.27	41.24	173.85	174.97
221	S	44	8.96	111.56	56.96	52.09	66.36	42.49	174.47	173.85
222	L	105	8.45	116.11	54.73	56.97	45.30	66.53	175.28	174.45
223	Y	247	9.10	123.31	53.88	54.72	39.76	45.46	174.84	175.27
225	G	398	8.44	112.24	45.07	62.17	_	32.11	173.86	176.60
Sequence gap										
255	Y	163	7.84	119.12	57.30	45.13	39.26	_	175.98	173.77
256	V	342	8.44	119.13	60.24	57.43	33.34	39.38	174.85	175.97
257	S	157	8.15	118.76	57.13	60.25	64.17	33.60	175.27	174.88
258	Q	196	8.87	120.55	59.19	57.12	27.28	64.29	177.52	175.23
259	E	336	8.24	117.84	59.25	59.16	28.18	27.43	178.68	177.51
260	L	234	7.70	122.43	56.84	59.25	40.38	28.51	178.80	178.69
261	A	191	8.19	120.33	55.01	56.84	18.55	40.63	179.40	178.80
262	G	1	8.43	103.53	47.08	54.97	_	18.61	175.07	179.39
263	E	237	7.83	122.70	58.43	47.09	28.14	_	178.70	175.07
264	R	177	7.90	119.79	59.21	58.42	28.18	28.21	177.88	178.65
265	V	333	7.54	116.90	66.11	59.19	30.23	28.36	176.92	177.90
266	A	156	7.64	118.75	55.02	66.12	18.68	30.37	178.64	176.95
267	M	106	8.16	116.24	58.91	55.00	32.52	18.75	177.40	178.60
268	V	125	7.08	117.01	64.58	58.89	29.91	32.61	177.98	177.42
269	V	345	7.61	119.76	65.57	64.57	31.31	30.16	175.22	177.94
270	A	331	7.94	116.68	53.20	65.58	21.85	30.96	179.15	175.22
271	D	201	9.09	120.74	54.50	53.21	38.85	22.07	178.74	179.18
273	E	140	9.68	117.92	58.63	65.61	27.34	31.49	176.43	178.42
274	F	97	7.06	115.45	57.05	58.58	38.55	27.51	175.29	176.43
275	R	224	7.34	121.88	56.57	57.04	29.43	38.66	174.45	175.28
276	Q	232	8.22	122.32	55.73	56.53	29.75	29.60	173.86	174.46
277	S	144	8.75	118.09	57.45	55.72	62.86	30.15	175.01	173.83
278	G	70	9.33	114.21	45.15	57.41	_	63.01	174.37	175.01
279	V	147	7.87	118.32	57.92	45.17	31.64	_	175.98	174.17
280	H	349	8.12	120.09	56.37	57.93	28.92	31.48	_	176.13
283	W	269	8.83	124.74	59.01	_	31.92	62.02	_	172.86
284	G	13	8.78	107.95	45.22	58.95	_	31.93	173.87	176.89
285	N	221	8.08	121.60	55.04	45.16	41.71	_	177.10	173.91
286	R	230	8.07	122.17	59.53	54.97	32.03	41.78	174.66	177.15
287	Q	208	7.41	121.03	54.64	59.23	41.19	63.15	176.89	174.94
288	K	226	8.22	121.89	53.99	62.78	31.56	31.21	174.11	176.81
289	E	218	8.42	121.36	56.17	62.99	30.18	31.53	177.07	177.09
290	G	27	8.45	109.84	45.15	56.10	_	30.21	174.28	177.08

Residue Number	Residue	Assigned Spin System	Atom Chemical Shifts (δ)							
			HN	N	Cα	pCα	Cβ	pCβ	CO	pCO
291	R	56	7.58	112.87	58.90	45.16	25.20	_	172.57	174.35
Sequence gap										
314	L	131	8.91	117.37	57.21	59.00	42.45	29.95	179.27	180.88
315	S	107	8.24	116.34	63.06	57.16	60.20	42.41	175.27	179.33
316	E	164	8.18	118.87	60.50	63.04	30.21	59.90	178.75	175.23
317	K	132	7.15	117.39	58.45	60.52	31.02	30.22	_	178.71
318	L	383	7.99	117.95	56.93	58.40	40.40	31.48	180.69	180.36
319	V	12	7.76	107.91	60.23	56.95	30.66	40.82	_	180.67
320	G	10	7.61	107.58	46.24	60.20	_	30.71	175.39	175.16
321	L	166	8.32	119.26	54.92	46.26	42.40	_	175.40	175.37
322	A	301	7.55	127.20	52.73	54.86	21.03	42.43	181.19	175.42

Table 7.1 Table of confident NMR assignments, displaying the respective residue number, amino acid, spin system and NMR chemical shifts for each of the protein backbone atoms.

7.4.4 Relaxation experiments

One of the key advantages of NMR as a structural technique, over crystallography for example, is the ability to determine the dynamics of a protein in solution. There are 3 parameters in NMR which can determine the relative dynamics corresponding to each signal in the HSQC: the spin-lattice (T_1) and spin-spin (T_2) relaxation times, and the nuclear overhauser effect (NOE) (Kay et al., 1989). T_2 relaxation describes the return of nuclear spins to equilibrium in the transverse plane, affected by the spins of neighbouring nuclei; this effect is lower in a more mobile environment, therefore the length of T_2 is increased. T_1 relaxation describes the return to equilibrium in the longitudinal plane; the effect of an increase in mobility on this value depends on the overall molecular correlation time, with moderate sized proteins, such as POR, likely to exhibit a negligible increase in T_1. The Nuclear Overhauser Effect (NOE) detects cross-relaxation between nuclei, which is sensitive to motions at low frequency and at twice the Larmor frequency; this can be measured by performing two experiments, one where the NOE is allowed to build up and one where it is not.

A ^{2}H, ^{15}N-labelled POR sample was prepared, at 300 µM in NMR buffer, and taken to Dr Matt Cliff at MIB at the University of Manchester. ^{15}N T_1 and T_2 relaxation and steady state NOE experiments were performed at 313 K in a Bruker 800 MHz spectrometer with a cryoprobe; the relaxation delays in the T_1 and T_2 experiments are detailed in Table 7.2. All spectra were processed in Sheffield, using in-house macros. In the T_1 and T_2 relaxation experiments, the intensities of the assigned peaks in the five spectra acquired were plotted against time (Figures 7.12 and 7.11, respectively) and fitted to an exponential decay using a Levenberg-

Marquardt least-squares algorithm; T_1 and T_2 relaxation times of all assigned residues were calculated from this. The intensities of the peaks in the NOE experiments were evaluated against each other, in order to give a value of the NOE effect for each amide.

Experiment Number	T_1 Relaxation Delay (ms)	T_2 Relaxation Delay (ms)
1	10	0
2	20	3.92
3	40	7.84
4	80	15.68
5	160	31.36
6	320	62.72
7	640	125.44
8	1200	250.88

Table 7.2 Table detailing the relaxation delays used for each of the T_1 and T_2 relaxation experiments. These are used, with the respective intensity of the signals in that experiment, to determine the T_1 and T_2 relaxation rate.

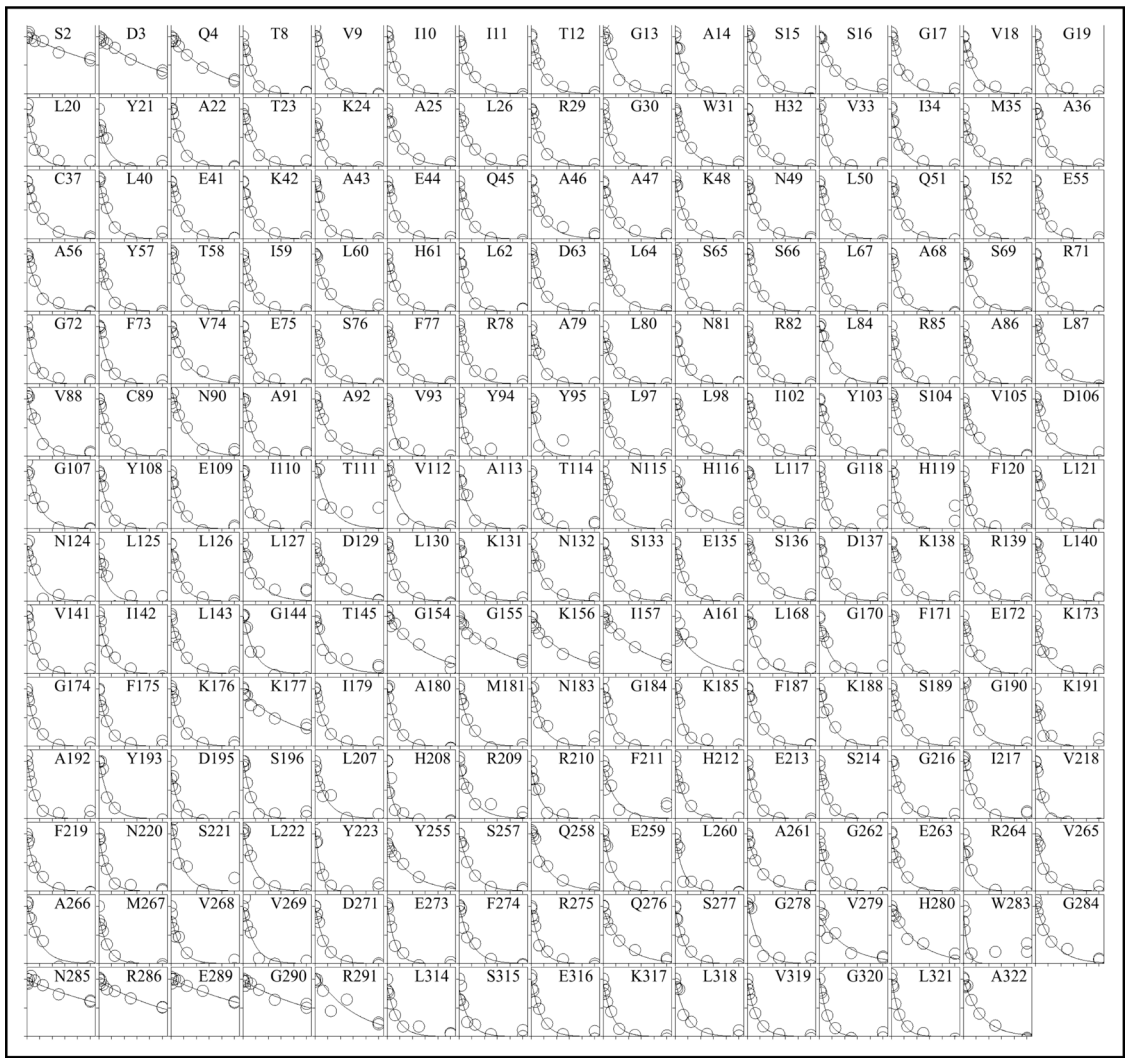

Figure 7.11 T_2 relaxation decay curves, plotted as transverse intensity (with the first value normalised to 1) against delay time (shown in Table 7.2, plotted on scale from 0 to 250 ms). The assigned residue is labelled in the top corner of each plot. There are a total of 209 assigned T_2 relaxation curves, selected on their accuracy of fitting.

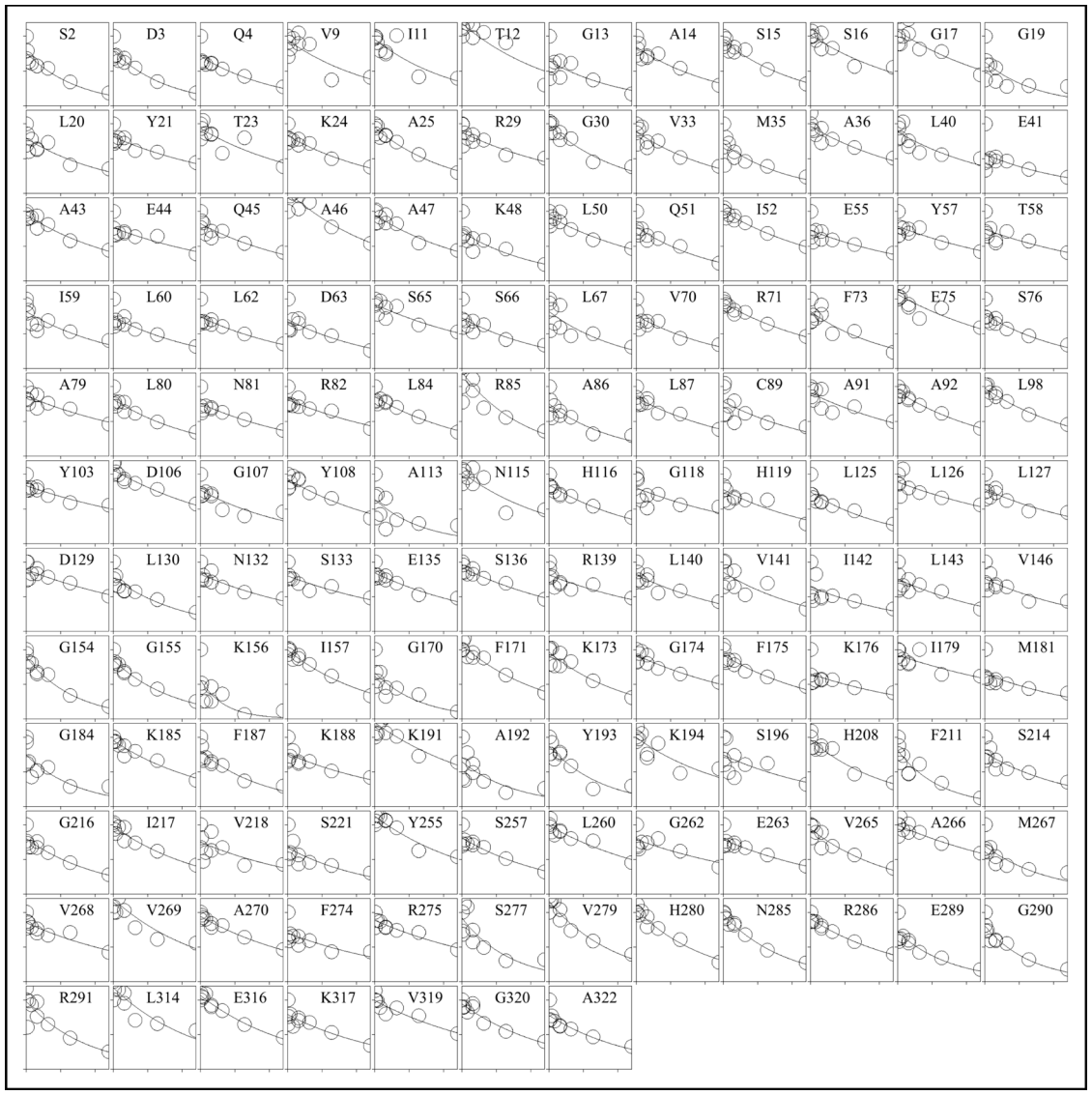

Figure 7.12 T_1 relaxation decay curves, plotted as longitudinal intensity (with the first value normalised to 1) against delay time (shown in Table 7.2, plotted on scale from 0 to 1200 ms). The assigned residue is labelled in the top corner of each plot. There are a total of 139 assigned T_1 relaxation curves, selected on their accuracy of fitting.

Each of the T_1 and T_2 and NOE values for assigned residues are plotted against sequence position in Figure 7.13. T_2 values for each unassigned spin system are also plotted, in order from highest to lowest; unassigned T_1 and NOE values from the same spin systems are plotted in accordance with these values, in the same order. The distribution of these plots indicate the high quality of the T_2 and NOE data, while the T_1 data is significantly more spread; however it is the T_2 and NOE data which provides the key information about protein dynamics, with a larger T_2 value and a lower NOE value indicating greater mobility of the backbone.

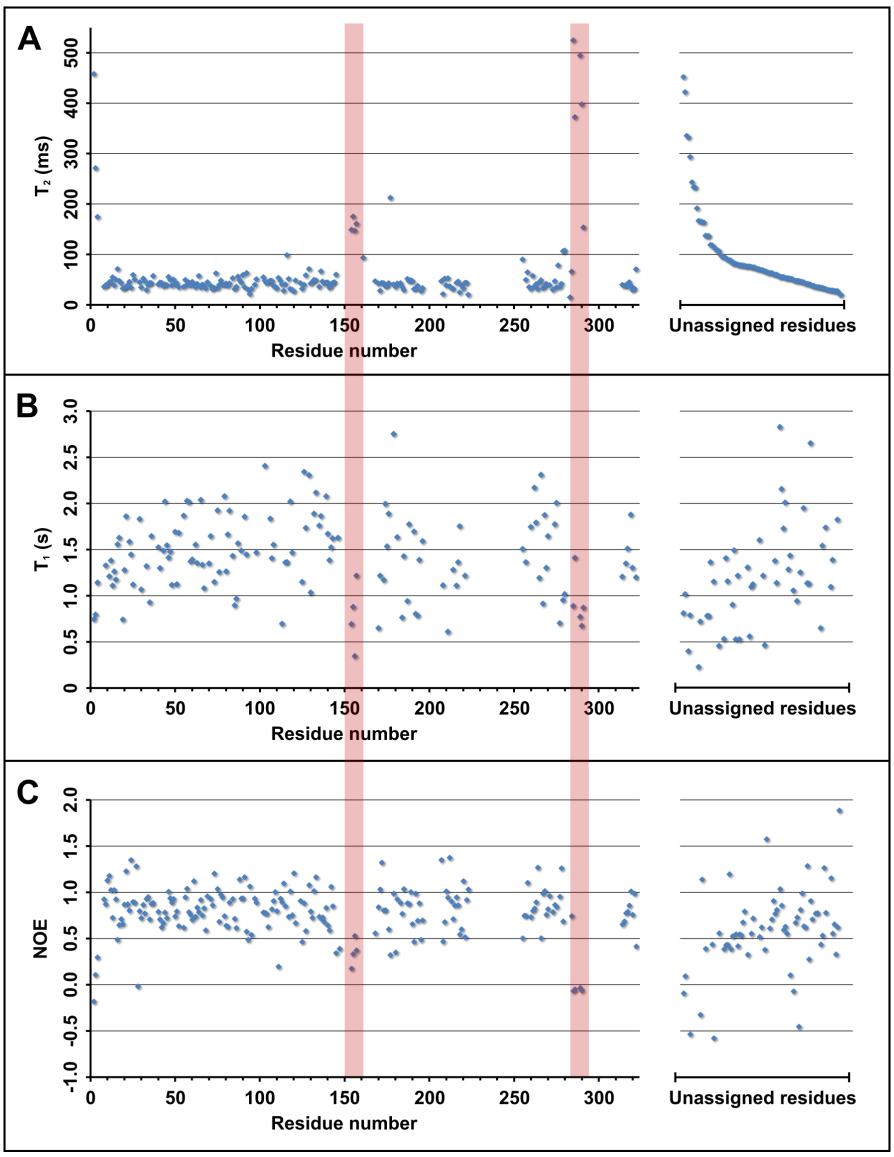

Figure 7.13 (A) T_2 relaxation data for each of the assigned POR residues. Data is also shown for the unassigned spin systems, displayed in order of decreasing T_2 value. An increase in T_2 is indicative of an increase in mobility; this can be observed for the N-terminus, around residues 150 - 160, and residue 285 - 295, indicating that these regions are mobile (highlighted by a pink strip). There are approximately 15 - 20 further unassigned spin systems that exhibit elevated T_2 values. (B) T_1 relaxation data for each of the assigned POR residues. Data is also shown for the unassigned spin systems, corresponding to those in A. There is a much greater spread in the T_1 relaxation data, though the values are broadly similar. T_1 values are unlikely to be altered significantly by changes in mobility in a protein of this size. (C) NOE values for each of the assigned POR residues. Data is also shown for the unassigned spin systems, corresponding to those in A. A decrease in NOE is indicative of an increase in mobility; this corresponds relatively well with the T_2 data, with a decrease in NOE clearly observed for the N-terminus and residue 285 onwards, with a slight decrease in NOE values at residues 150 - 160.

It is clear from this data that the majority of the protein acts as a rigid globular structure, while there are two regions, other than the extreme termini, that clearly indicate mobility; one at around 150 - 160 residues and, particularly, the section from residue 285 onwards; these are the two regions which were previously predicted to be disordered (detailed 3.2.1.2). The number of assignments within these regions is quite poor, suggesting that the dynamics of these residues has an effect on the signals; though, in most cases, an increase in

mobility results in more intense signals in NMR. It may be, however, that these regions are in intermediate exchange, i.e. these sections switch between two conformations during acquisition, thus causing multiple peaks with significantly broader signals. It is assumed that these are the only two mobile regions of the protein, as the number of unassigned residues with elevated T_2 values is no more than 15 - 20; with a number of residues missing around the assigned mobile regions, these are likely to be the missing assignments.

In order to calculate the correlation time (τ_c) of POR, average values for T_1 and T_2 data, for non-mobile regions of the protein, were taken and input into the following equation:

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7}$$

Equation 7.1 Equation to calculate the correlation time of proteins using calculated T_1 and T_2 values.

Where ν_n is the ^{15}N resonance frequency (in Hz). This equation is derived from Eq. 8 in (Farrow et al., 1994). This gave a correlation time of 12.5 ns for POR; this is roughly as expected for a monomeric globular protein at this temperature.

7.4.5 TALOS-N Prediction of Protein Torsion Angles

The chemical shift data from the assignment was submitted to the TALOS-N server (Shen and Bax, 2013) for prediction of backbone and side-chain torsion angles, mobility and secondary structure. The TALOS-N software package takes the chemical shift data and compares it to a structural database, where it matches the shifts for each residue with the existing data for that amino acid, where it is flanked by the same amino acids. The 25 best matches are taken and, if there is a consensus of protein backbone φ/Ψ torsion angles, this is used to predict the likely backbone torsion angles in the POR structure. These backbone torsion angles are used to predict the likely secondary structure of the protein at this region, with a confidence score for each prediction. The programme also produces an estimated S^2 order parameter for each residue, derived from the chemical shifts, based on the Random Coil Index (RCI) described by (Berjanskii and Wishart, 2005); this determines the likelihood of disorder throughout the protein, both in assigned and non-assigned regions.

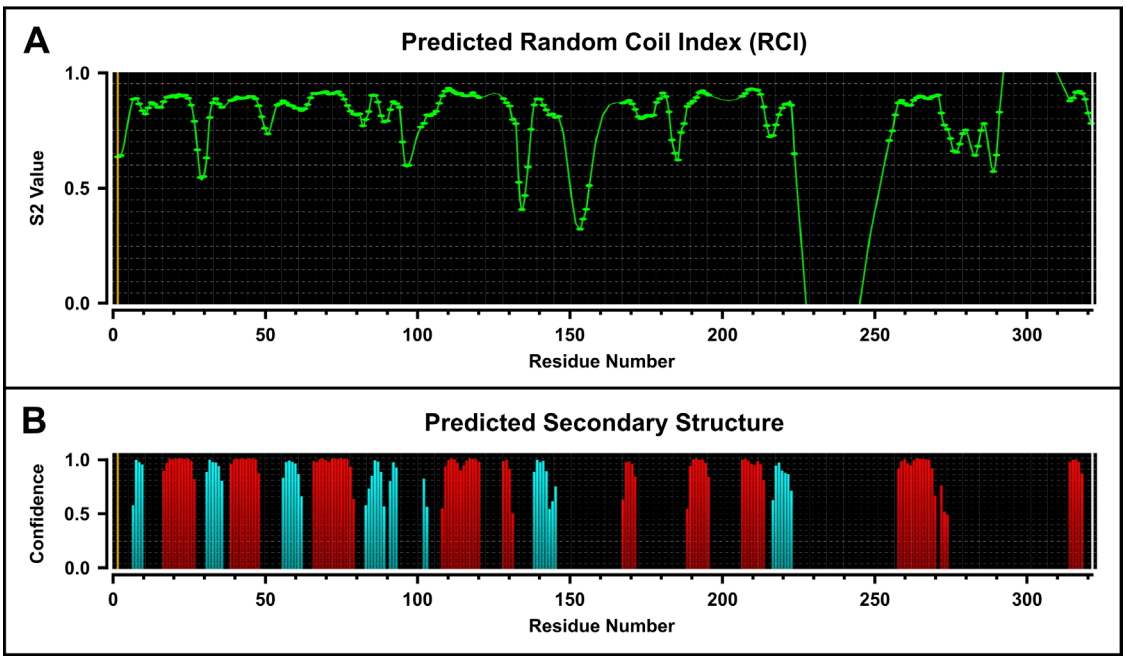

Figure 7.14 (A) Random Coil Index (RCI) data for POR, with predictions based on NMR chemical shift data submitted to TALOS-N. S_2 parameter denotes the relative order in the structure, with 1 representing highly ordered structure, and 0 representing complete disorder. There are a number of small disordered loops predicted throughout the protein. The region at residues 150 - 160 displays clear disorder. The region from residue 220 to 260 displays very high levels of disorder; there are, however, no assignments in the majority of this region (Table 7.1), therefore the prediction is based on the assignment of flanking residues, prior to 226 and after 254. (B) Predicted secondary structure for POR, based on backbone φ/Ψ torsion angles calculated by TALOS-N from the NMR chemical shift data. Much of the N-terminal is well structured, with the C-terminus display significantly less structural character; though this may be due to the limited assignment information in this region.

The predicted RCI for POR (Fig. 7.14A), calculated for each residue from their chemical shift and the shifts of flanking residues, indicates a number of regions where the protein backbone is likely to be quite disordered; these tend to be short regions in between secondary structural elements. The loop region of POR is predicted to disordered at the N-terminal end, at approximately residues 145 - 160, and the C-terminal end predominantly ordered, from approximately residue 160 onwards; this correlates very well with the relaxation data, which indicated mobility at residues 154 – 157, while residues from 168 onward exhibited no mobility. The other region of clear mobility in the relaxation data was from residues 285 onwards, which in the predicted RCI has a moderate level of disorder, followed by a region at which the S^2 value goes above 1; the latter prediction relies on chemical shifts from flanking residues, with no assignments in the majority of this region, therefore is likely to be quite inaccurate. The final region of interest is the region between residues 220 and 260, which indicates an extremely low S^2 value and thus a prediction of high levels of disorder. There are no assignments within this region, suggesting that the prediction of high disorder cannot be fully accurate, however there are assigned residues at either end of this region which do suggest a tendency towards disorder; if this was a region of high disorder, it could explain the lack of assignment due to a number of shifts in intermediate exchange.

The secondary structure predicted for POR by TALOS-N (Fig. 7.14B) indicates a highly structured N-terminal region; this prediction matches well with what is expected for the Rossmann fold region of POR, with alternating β-strands and α-helices. Within the loop region there is a small amount of α-helical structure predicted around residue 170, as previously modelled in chapter 3, however there is no secondary structure predicted for the remainder of the loop region; this would be expected for the early portion of the loop, which is mobile, but is slightly surprising for the latter part of the loop, which is well assigned and known to have no mobility. There are two large regions within the C-terminus where no structure has been predicted; much of this is unassigned and, therefore, is to be expected, while much of the assigned residues around 280 - 290 are known to be mobile so no structure would be expected anyway.

7.4.6 Application of NMR data to the POR Structural Model

The backbone torsion angles produced by TALOS-N were compared with those from the structural model of POR produced in Chapter 3, in order to determine the validity of this as a POR structure. Residues were labelled in the model (Fig. 7.15), depending on various information: which matched well with the TALOS-N data; those that the TALOS-N data disagreed with; residues predicted to be mobile by the relaxation data or TALOS-N; unassigned residues with no structural information. The Pchlide and NADPH substrates are also modelled in the structure for reference to binding sites.

It is clear, as expected, that the well-conserved Rossmann fold region of POR has been modelled correctly; there are only slight errors in the precise length of β-sheets and α-helices, and the conformation of the intervening loops. Within the POR loop region, it is clear that the helical section does not extend as far, C-terminally, as modelled in Fig. 3.9; the residues prior to the known helix region, however, are not assigned and so could form an extension of the helix in the N-terminal direction. The known mobile region of the loop is predicted to involve the end and beginning of two α-helices, suggesting these have been modelled incorrectly.

Much of the C-terminal region of POR is unassigned; therefore conclusions cannot be drawn about the model in these regions. There are, however, still some significant regions of assignment within the C-terminal region, though these all appear to disagree with the proposed structure in the model. The backbone torsion angles from the TALOS-N data could be used, however, in order to identify the correct structure within these regions and refine

the model further; these include regions directly following well modelled structures, which could be confidently remodelled in the structure.

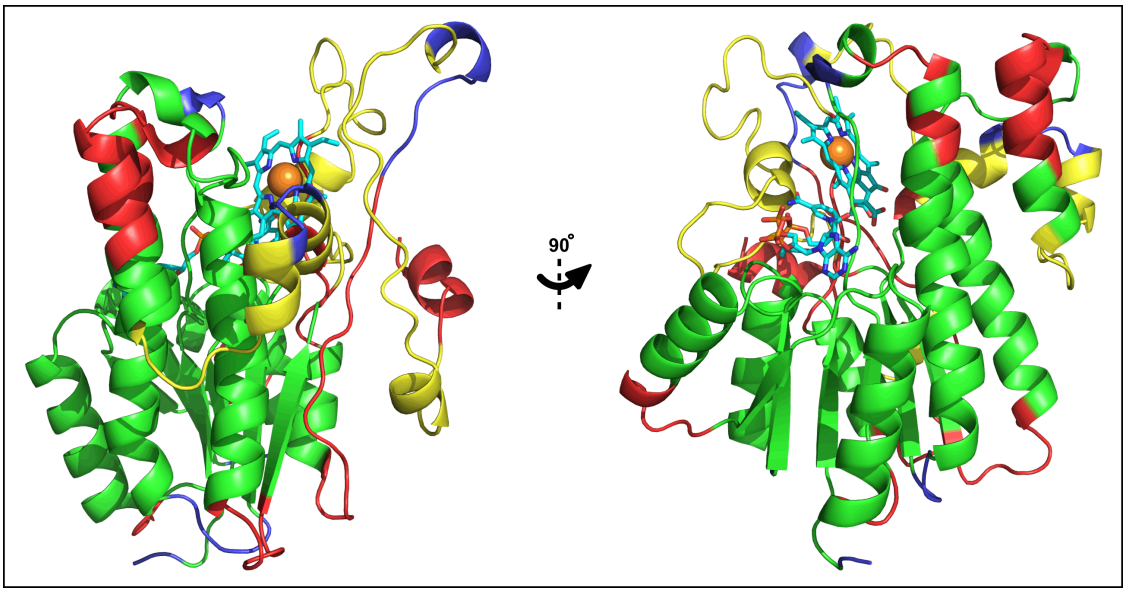

Figure 7.15 Structural model of POR, with the backbone torsion angles compared to those predicted by TALOS-N from the NMR chemical shift data. Residues with backbone torsion angles matching those from TALOS-N are displayed in green. Residues with backbone torsion angles which are contradicted by TALOS-N data are displayed in red. Mobile residues, determined from TALOS-N predictions and NMR relaxation experiments, are displayed in blue. Residues with no assignment data are displayed in yellow. It is clear that the majority of the Rossmann fold region of POR (in the lower half of the structure) is correctly modelled. The C-terminal region of the model, however, displays a large amount of incorrect structure, along with a number of unassigned regions; this region is, therefore, likely to be poorly modelled. The TALOS-N data should be useful for further refinement of the model.

7.5 Discussion

For many years people have attempted to solve the structure of POR, with extensive crystal trials carried out (Professor Neil Hunter FRS, personal communication), yet structural data is yet to be reported for this enzyme. Recent work (Proudfoot, 2011) has turned to NMR as a technique to elucidate the structure of a protein that had been thought to be highly dynamic. The production of a high quality, ^{2}H, ^{13}C, ^{15}N-labelled sample of POR, in this study, has enabled the first structural data to be elucidated for this unique, light-dependent enzyme. In addition to this, relaxation experiments have identified regions of POR which are highly mobile in the apo protein, indicating the dynamic nature of this enzyme.

Production of a triple-labelled sample involved the development of various protocols for cell growth in minimal media, protein labelling and amide exchange; the problems encountered in the amide exchange experiments highlighted the stability within the core of the POR enzyme from *T. elongatus*. Analysis of NMR spectra at a range of pH values indicated that a number of broad signals were only present at lower pH, indicating missing signals for possible mobile residues at the higher pH. The protein was determined to be well-folded at the lower

pH despite suggestion that broad signals may correspond to molten globule states; thus suggesting that, although these signals are likely to correspond to signals from regions exhibiting slow timescale fluctuations, the protein is, on the whole, well folded. Despite the difficulties encountered, a high quality triple-labelled sample was produced which provided significantly more resolved and intense NMR signals than observed with ^{15}N-labelled POR samples.

Backbone assignment experiments were carried out on this sample of POR, in order to assign NMR chemical shifts to residues throughout the protein; these experiments were carried out on a high-field NMR machine, with increased sensitivity provided by the cryo-cooling of the detection probes. The resultant spectra were of high quality with intense signals throughout most of the data; there was still a large amount of overlap in certain regions, however, which led to difficulties in the determination of spin system connectivities in the peak picking process.

Assignment of the spin systems from the backbone assignment experiments involved a number of problems; the assignment process was revisited a number of times, with the spectra analysed further, in order to determine additional assignments. Assignment within the N-terminal region of the protein was very good, with only a few sporadic assignments, and a small number within the loop region, missing. The assignment within the C-terminus was not as successful, with two particular large segments of the protein remaining unassigned. Overall, however, the level of assignment was quite good, far surpassing the previous attempts to assign the ^{15}N-labelled protein in which only around 20 % of the protein could be assigned (Proudfoot, 2011).

A selection of relaxation experiments were carried out on a ^{2}H, ^{15}N-labelled sample of POR in order to analyse the mobility of the protein; these gave values for the T_1, T_2 and NOE parameters, which together describe the dynamics within the protein. Where assignments were available, these mobility parameters were applied to the POR sequence; once plotted, these values gave a clear indication as to the location of the mobile regions of POR. The loop region of POR, discussed in Chapters 3 and 6, was thought to be a likely mobile region and this has proved to be the case for the initial part of the loop; the C-terminal section of the loop does not exhibit an increase in mobility, however, indicating that the mobile region is shorter than originally thought. This region has been suggested to form protein-protein interactions between ternary complexes of POR (Reinbothe et al., 1999); the mobile nature of this region would be likely to be important in positioning this region to enable binding to a neighbouring protein.

The other significantly mobile region of POR, determined from the dynamics data, is from residue 285 onwards; though this region is only assigned up until residue 291, therefore it is not known how large this mobile region could be. The Pchlide binding region of POR must be located in the C-terminus; this mobile region could be important in the binding of the large, hydrophobic Pchlide molecule, by reorganisation of the structure to exclude water from the binding pocket, for instance; alternatively the mobile region could also be involved in protein-protein interactions. The unassigned signals which exhibit increased mobility number approximately 15 - 20 residues; considering the low number of residues assigned within the mobile regions currently, it is likely that these residues form the remainder of these mobile regions, however it cannot be ruled out that they are signals from an alternative mobile region.

The NMR chemical shift data was uploaded to the TALOS-N server, with the protein backbone torsion angles predicted, along with the secondary structure and disorder. Comparison of this data to the POR structural model, produced in Chapter 3, has enabled the confirmation of the structure for much of the N-terminus of this model. There are quite a few regions, particularly within the C-terminus, where the structure does not match with the backbone dihedral angles predicted by TALOS-N. The use of these confident backbone angles, secondary structural information and mobility information could be used in molecular dynamics simulations in order to further improve this model; this would produce a far more accurate reference structure for further work on the catalytic mechanism of POR.

It is hoped that the level of assignment could be improved further, in order to understand even more about the structure of the enzyme. There are a number of other experiments which could aid the improvement of the current model: NMR titrations of NADPH would identify the residues which interact with the cofactor, enabling improved docking of the molecule within the model; paramagnetic relaxation enhancement (PRE) experiments involve the addition of paramagnetic spin labels that broaden signals in a distance-dependent fashion and could be used as a further distance restraint within the model. These techniques would all contribute to the improvement of the current structural model which, in the absence of a crystal structure, presents the best quality POR structure available.

Chapter 8

8. Conclusions and Future Work

8.1 Conclusions

Thorough bioinformatics analysis of POR has indicated that this protein is likely to be significantly dynamic in nature. In addition to highlighting similarities of this protein to other members of the SDR superfamily, two regions of POR, one in the loop region of POR and one near the C-terminus, have been confidently predicted to exhibit a tendency for disorder. The loop region in the POR sequence was previously thought to be quite unique within the SDR superfamily; yet similar regions have also been observed, in two structurally distinct forms, amongst other members of the superfamily. The key difference between these two types of loop structure involves the positioning relative to the two α-helices involved in protein dimerisation within the SDR superfamily. In cases where the loop is positioned over these helices, protein-protein interactions do not occur at this site; whereas in structures facilitating the exposure of these helices, this region acts as a dimerisation face in protein oligomerisation. Carbonyl reductase enzymes, from distinct organisms, in fact contain the two alternative loop structures, thereby forming distinct oligomers despite exhibiting identical functions.

This clearly implicates these loop regions, within the SDR superfamily, in determining protein oligomerisation. It is likely that the loop region in POR could function similarly; with the mobility of this region enabling control of protein-protein interactions, by the ability to translate between structures which obstruct or expose these oligomerisation sites. The bioinformatics analysis has led to the production of a new model of POR, based on the structure of homologous enzymes. A previous model of POR, published by Townley et al., determined the loop region to be positioned across the top of the bound Pchlide molecule; this is unlikely, as there is no evidence to suggest that the loop is involved in substrate binding, with the bioinformatics work here, along with previous work (Reinbothe et al., 2003b), suggesting that the loop is expected to be implicated in protein-protein interactions. In the new model, the loop region of POR is based on a structure which prevents oligomerisation; though, as the loop is expected to be quite mobile, it is possible that this region could exist in a number of conformations.

In order to understand more about the function of the loop region, a number of glycine residues were identified as possible hinge points for loop motion. A number of mutations

were made, converting these glycine residues to alanines, in an attempt to reduce the flexibility of the protein backbone at these regions, therefore restricting mobility of the loop. These mutants were all found to have a clear effect in slowing down catalytic activity of the enzyme, though none completely abolished the activity. EPR experiments, involving attaching a label to the loop to probe its mobility, indicated that the loop was indeed dynamic. The glycine mutations, however, did not appear to have a significant effect on the overall mobility of the loop; though there were indications that the nature of the loop mobility had been altered. This suggested that the loop functions as a disordered region, with the flexibility of the loop altered by the mutations yet, overall, it maintains mobility.

The effect of the glycine mutations upon the formation of large protein complexes was analysed using AUC. This did not indicate any obvious effect of these mutations upon the formation of large aggregates of POR complexes, despite the disruption of the loop's inherent flexibility. This suggests that the effect of the glycine mutations, on the conformation of the loop, was actually quite low, therefore did not affect its function; alternatively, it is conceivable that another region of POR, or even the hydrophobic Pchlide substrate itself, may be involved in the assembly of the high molecular weight structures of POR complex.

NMR experiments in this study have enabled the first structural information to be gathered for POR. The production of a high quality triple-labelled sample of POR led to the assignment of a significant number of residues; the majority of these were in the well-conserved, N-terminal Rossmann fold region of POR; however, assignments were also made within the loop region and the C-terminus of POR. Relaxation experiments were carried out in order to calculate the dynamics of the protein; these indicated two regions of significant mobility, one within the loop region of POR and one within the C-terminus; these matched up well with the regions that were predicted to be disordered in the bioinformatics studies.

Identification of the mobility of the residues in the loop demonstrated that the mobile region is not as extensive as originally thought; the C-terminal end of the loop, from residue 168 onwards, was not mobile, indicating that the dynamic region is only likely to incorporate, at most, 20 residues, with only a small number of these currently assigned. In addition to the assigned residues, there were additional signals which also indicated an increase in mobility; it is not known whether these are likely to represent signals from an alternative mobile region, or solely the missing signals from the currently identified mobile regions; however, the limited number of these suggests the latter is most likely. The NMR chemical shifts have been submitted to the TALOS-N server, predicting the backbone torsion angles and secondary structure for the assigned residues. This structural data has been compared to the model and indicates much of the model is correct, mainly in the N-terminus; the modelling of the C-

terminus was shown to be poorer, though much of this region cannot be compared due to the lack of assignment data.

In order for high quality structural data to be obtained for POR, in complex with its substrates, preparation of pure Pchlide is required, to ensure no artefacts in the complex samples. The previous protocol used for preparation of Pchlide (Proudfoot, 2011), though producing large amounts of pigments, was found to be of very low purity; with the NMR spectrum indicating a large number of contaminants present, that could affect complex formation. Therefore, a new protocol for preparation of Pchlide was devised, involving solvent extraction of Pchlide, directly from the growth medium, and separation of pigment by HPLC. Both absorbance and NMR spectra of this sample indicated its high quality and purity; this protocol was therefore used in the future preparation of substrate for structural studies of the POR complex.

A number of techniques have been used to analyse the structure of the complexes formed with POR and its substrate. Initial formation of POR complex samples, involving the addition of Pchlide from methanol, led to the production of very large complexes, which could be observed by AUC, EM and DLS. These were shown to break down upon turnover of the enzyme, with the product complex appearing to be monomeric. Improvements to the protocol for POR complex formation, involving solubilisation of Pchlide in water-based detergent, led to the formation of complexes with significantly lower levels of aggregation. These complexes were analysed by AUC, DLS, EM and AFM, with all techniques indicating the absence of the extremely high molecular weight complexes observed in the previous protocol.

Instead, the complexes were found to be approximately 4 – 10 POR monomers in size; these were observed, in both the EM and AFM, predominantly as ring structures, some with a central pore and others with a central protein component. These ring structures are likely to be of some functional importance; it is surprising that such a variation in structures exists, suggesting that a single, native ring structure is unable to form in the absence of a membrane surface, or even an additional assembly protein. It is possible that these ring structures further associate, into significantly larger structures, as observed in the initial POR complex preparations. The ring structures could form an arrangement, of both the POR proteins and the pigment molecules, that facilitates more reliable absorption of photons and shuttling of photons to activated POR complexes; alternatively, neighbouring POR complexes could have a photoprotective role, dissipating any excess photons to prevent formation of reactive oxygen species which could be damaging to the cell.

It is clear, from this study, that a number of mobile regions of POR are integral to the unique function of this enzyme; it is likely that some of these are involved in protein-protein interactions, leading to the formation of complexes which may be necessary for function within the cell. Other regions may be involved in the binding of substrate, as it is apparent that much of the C-terminus, implicated in Pchlide binding, does not have a fixed structure in the apo protein; binding of substrate could instigate the appropriate folding of this region, in addition to the propagation of the various other conformational changes throughout the protein.

8.2 Future work

The NMR assignment of such a large number of residues of POR has opened the door for a number of further NMR experiments, in order to understand even more about the structure and function of the enzyme, which were unfortunately not possible in the timescale of this thesis. Initially, attempts should be made to try to improve the assignment further, as, although the number of missing assignments is relatively low, these could give vital structural information for poorly defined regions of POR, particularly in the C-terminus. Arrangements for extending the assignments are currently ongoing. Titration of POR with NADPH could be followed by NMR, indicating the residues involved in the binding of the cofactor; this includes the identification of any conformational changes produced as a result of this binding event.

Paramagnetic relaxation enhancement (PRE) experiments involve the addition of a paramagnetic label to a particular region of the protein, inducing a distance-specific broadening of signals around this label. Application of this technique to POR could enable distance restraints to be determined for certain regions of the protein, particularly the loop region and C-terminus, and help to further refine the current POR structural model; these experiments could also be used to improve the assignment, as it may highlight the locations of some unassigned residues within the POR sequence. Work described in this thesis led to the production of POR with suitable spin labels attached at either of two single cysteines. Thus, almost all of the work required for generating suitable constructs for PRE studies has been done. This is a key target for future research.

The glycine mutants analysed in this thesis were chosen based on the predicted loop region, which has since been determined to be shorter than originally predicted. As the loop region is now known to act as an inherently flexible region, rather than a rigid, hinged structure, mutations can be designed with this in mind; mutations which aim to reduce the flexibility of this region could help to understand how it is important to POR function. Alternatively, due

to the reduced length of the mobile region, it may now be feasible to remove the loop from the sequence altogether, without significantly disrupting protein folding, and analyse any alteration in protein function; the loop could also be replaced with a sequence from an alternative SDR protein.

Organisation of the ternary complex of cyanobacterial POR has not been extensively studied, however this work indicates that it forms some intriguing structures. Further to this work, it would be interesting to prepare complexes of POR in the presence of membrane bilayers, in order to mimic the membrane surface on which POR is proposed to associate; the ability to study structures in near-native conditions using AFM presents an opportunity to study the assembly of these complexes on a membrane surface, under liquid conditions. The samples should also be prepared at varying detergent concentrations, to determine if this has an effect on the structures observed.

Attempts to produce crystals of POR for X-ray crystallography should also be continued, particularly with substrates bound, as a high-resolution structure of POR would improve our understanding of the catalytic mechanism. A more recently developed technique that could be utilised is time-resolved wide angle X-ray scattering (WAXS); this technique enables the study of protein dynamics in solution, with X-ray scattering data collecting on ultrafast timescales to enable conformational changes in the protein to be monitored, albeit at relatively low resolution. Certainly, this thesis has further underlined the unique nature of POR; its light-initiated catalysis, highly dynamic nature and intricate complex structures make it a truly fascinating protein for further study.

References

Addlesee, H.A., Fiedor, L., Hunter, C.N., 2000. Physical Mapping of bchG, orf427, and orf177 in the Photosynthesis Gene Cluster of *Rhodobacter sphaeroides*: Functional Assignment of the Bacteriochlorophyll Synthetase Gene. *J. Bacteriol.* **182**, 3175–3182. doi:10.1128/JB.182.11.3175-3182.2000

Addlesee, H.A., Gibson, L.C.D., Jensen, P.E., Hunter, C.N., 1996. Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, chlP, of Synechocystis sp. PCC 6803. *FEBS Lett.* **389**, 126–130. doi:10.1016/0014-5793(96)00549-2

Addlesee, H.A., Hunter, C.N., 1999. Physical Mapping and Functional Assignment of the Geranylgeranyl-Bacteriochlorophyll Reductase Gene, bchP, of Rhodobacter sphaeroides. *J. Bacteriol.* **181**, 7248–7255.

Adra, A.N., Rebeiz, C.A., 1998. Chloroplast Biogenesis 81: Transient Formation of Divinyl Chlorophyll a Following a 2.5 ms Light Flash Treatment of Etiolated Cucumber Cotyledons. *Photochem. Photobiol.* **68**, 852–856. doi:10.1111/j.1751-1097.1998.tb05295.x

Baker, M.E., 1994. Protochlorophyllide reductase is homologous to human carbonyl reductase and pig 20 beta-hydroxysteroid dehydrogenase. *Biochem. J.* **300**, 605–607.

Barna, J., Laue, E., Mayger, M., Skilling, J., Worrall, S., 1987. Exponential Sampling, an Alternative Method for Sampling in Two-Dimensional NMR Experiments. *J. Magn. Reson.* **73**, 69–77. doi:10.1016/0022-2364(87)90225-3

Battersby, A.R., Fookes, C.J.R., Matcham, G.W.J., McDonald, E., 1979. Order of assembly of the four pyrrole rings during biosynthesis of the natural porphyrins. *J. Chem. Soc. Chem. Commun.* 539–541. doi:10.1039/C39790000539

Bauer, C.E., Bollivar, D.W., Suzuki, J.Y., 1993. Genetic analyses of photopigment biosynthesis in eubacteria: a guiding light for algae and plants. *J. Bacteriol.* **175**, 3919–3925.

Beale, S.I., 1999. Enzymes of chlorophyll biosynthesis. *Photosynth. Res.* **60**, 43–73. doi:10.1023/A:1006297731456

Begley, T.P., Young, H., 1989. Protochlorophyllide reductase. 1. Determination of the regiochemistry and the stereochemistry of the reduction of protochlorophyllide to chlorophyllide. *J. Am. Chem. Soc.* **111**, 3095–3096. doi:10.1021/ja00190a071

Belyaeva, O.B., Timofeev, K.N., Litvin, F.F., 1988. The primary reactions in the protochlorophyll(ide) photoreduction as investigated by optical and ESR spectroscopy. *Photosynth. Res.* **15**, 247–256. doi:10.1007/BF00047356

Benz, J., Wolf, C., Rüdiger, W., 1980. Chlorophyll biosynthesis: Hydrogenation of geranylgeraniol. *Plant Sci. Lett.* **19**, 225–230. doi:10.1016/0304-4211(80)90076-0

Berjanskii, M.V., Wishart, D.S., 2005. A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *J. Am. Chem. Soc.* **127**, 14970–14971. doi:10.1021/ja054842f

Bhattacharya, A., Revington, M., Zuiderweg, E.R.P., 2010. Measurement and interpretation of N-15-H-1 residual dipolar couplings in larger proteins. *J. Magn. Reson.* **203**, 11–28. doi:10.1016/j.jmr.2009.11.014

Birve, S., Selstam, E., Johansson, L., 1996. Secondary structure of NADPH: protochlorophyllide oxidoreductase examined by circular dichroism and prediction methods. *Biochem. J.* **317**, 549–555.

Breckau, D., Mahlitz, E., Sauerwald, A., Layer, G., Jahn, D., 2003. Oxygen-dependent Coproporphyrinogen III Oxidase (HemF) from Escherichia coli Is Stimulated by Manganese. *J. Biol. Chem.* **278**, 46625–46631. doi:10.1074/jbc.M308553200

Bryant, D.A., Frigaard, N.-U., 2006. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol.* **14**, 488–496. doi:10.1016/j.tim.2006.09.001

Buhr, F., Bakkouri, M.E., Valdez, O., Pollmann, S., Lebedev, N., Reinbothe, S., Reinbothe, C., 2008. Photoprotective role of NADPH:protochlorophyllide oxidoreductase A. *Proc. Natl. Acad. Sci.* **105**, 12629–12634. doi:10.1073/pnas.0803950105

Buick, R., 2008. When did oxygenic photosynthesis evolve? *Philos. Trans. R. Soc. B-Biol. Sci.* **363**, 2731–2743. doi:10.1098/rstb.2008.0041

Burke, D.H., Alberti, M., Hearst, J.E., 1993. bchFNBH bacteriochlorophyll synthesis genes of Rhodobacter capsulatus and identification of the third subunit of light-independent protochlorophyllide reductase in bacteria and plants. *J. Bacteriol.* **175**, 2414–2422.

Burton, G., Fagerness, P.E., Hosozawa, S., Jordan, P.M., Scott, A.I., 1979. 13 C n.m.r. evidence for a new intermediate, pre-uroporphyrinogen, in the enzymic transformation of porphobilinogen into uroporphyrinogens I and III. *J. Chem. Soc. Chem. Commun.* 202–204. doi:10.1039/C39790000202

Camadro, J.M., Thome, F., Brouillet, N., Labbe, P., 1994. Purification and properties of protoporphyrinogen oxidase from the yeast Saccharomyces cerevisiae. Mitochondrial location and evidence for a precursor form of the protein. *J. Biol. Chem.* **269**, 32085–32091.

Canniffe, D.P., Chidgey, J.W., Hunter, C.N., 2014. Elucidation of the preferred routes of C8-vinyl reduction in chlorophyll and bacteriochlorophyll biosynthesis. *Biochem. J.* **462**, 433–40. doi:10.1042/BJ20140163

Canniffe, D.P., Jackson, P.J., Hollingshead, S., Dickman, M.J., Hunter, C.N., 2013. Identification of an 8-vinyl reductase involved in bacteriochlorophyll biosynthesis in Rhodobacter sphaeroides and evidence for the existence of a third distinct class of the enzyme. *Biochem. J.* **450**, 397–405. doi:10.1042/BJ20121723

Cavaleiro, J.A.S., Kenner, G.W., Smith, K.M., 1974. Pyrroles and related compounds. Part XXXII. Biosynthesis of protoporphyrin-IX from coproporphyrinogen-III. *J. Chem. Soc. [Perkin 1]* 1188–1194. doi:10.1039/P19740001188

Chew, A.G.M., Bryant, D.A., 2007. Chlorophyll Biosynthesis in Bacteria: The Origins of Structural and Functional Diversity. *Annu. Rev. Microbiol.* **61**, 113–129. doi:10.1146/annurev.micro.61.080706.093242

Chou, P.Y., Fasman, G.D., 1974. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry (Mosc.)* **13**, 211–222. doi:10.1021/bi00699a001

Cole, C., Barber, J.D., Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201. doi:10.1093/nar/gkn238

Coomber, S.A., Chaudhri, M., Connor, A., Britton, G., Hunter, C.N., 1990. Localized transposon Tn5 mutagenesis of the photosynthetic gene cluster of Rhodobacter sphaeroides. *Mol. Microbiol.* **4**, 977–989. doi:10.1111/j.1365-2958.1990.tb00670.x

Cornah, J.E., Roper, J.M., Singh, D.P., Smith, A.G., 2002. Measurement of ferrochelatase activity using a novel assay suggests that plastids are the major site of haem biosynthesis in both photosynthetic and non-photosynthetic cells of pea (Pisum sativum L.). *Biochem. J.* **362**, 423–432.

Crockett, N., Alefounder, P.R., Battersby, A.R., Abell, C., 1991. Uroporphyrinogen III synthase: Studies on its mechanism of action, molecular biology and biochemistry. *Tetrahedron.* **47**, 6003–6014. doi:10.1016/S0040-4020(01)86492-9

Dailey, T.A., Dailey, H.A., 2002. Identification of [2Fe-2S] Clusters in Microbial Ferrochelatases. *J. Bacteriol.* **184**, 2460–2464. doi:10.1128/JB.184.9.2460-2464.2002

Davison, P.A., Schubert, H.L., Reid, J.D., Iorg, C.D., Heroux, A., Hill, C.P., Hunter, C.N., 2005. Structural and Biochemical Characterization of Gun4 Suggests a Mechanism for Its Role in Chlorophyll Biosynthesis. *Biochemistry (Mosc.)* **44**, 7603–7612. doi:10.1021/bi050240x

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z., 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59. doi:10.1016/S1093-3263(00)00138-8

Elder, G.H., Evans, J.O., Jackson, J.R., Jackson, A.H., 1978. Factors determining the sequence of oxidative decarboxylation of the 2- and 4-propionate substituents of coproporphyrinogen III by coproporphyrinogen oxidase in rat liver. *Biochem. J.* **169**, 215–223.

Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F.T., Iii, B.M., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., Steffen, W., 2000. The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science* **290**, 291–296. doi:10.1126/science.290.5490.291

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E., 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry (Mosc.)* **33**, 5984–6003.

Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., Iwata, S., 2004. Architecture of the photosynthetic oxygen-evolving center. *Science.* **303**, 1831–1838. doi:10.1126/science.1093087

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2013. Pfam: the protein families database. *Nucleic Acids Res.* doi:10.1093/nar/gkt1223

Frueh, D.P., Goodrich, A.C., Mishra, S.H., Nichols, S.R., 2013. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* **23**, 734–739. doi:10.1016/j.sbi.2013.06.016

Fujita, Y., Bauer, C.E., 2000. Reconstitution of Light-independent Protochlorophyllide Reductase from Purified Bchl and BchN-BchB Subunits: IN VITRO CONFIRMATION OF NITROGENASE-LIKE FEATURES OF A BACTERIOCHLOROPHYLL BIOSYNTHESIS ENZYME. *J. Biol. Chem.* **275**, 23583–23588. doi:10.1074/jbc.M002904200

Fujita, Y., Takagi, H., Hase, T., 1998. Cloning of the gene encoding a protochlorophyllide reductase: the physiological significance of the co-existence of light-dependent and -independent protochlorophyllide reduction systems in the Cyanobacterium Plectonema boryanum. *Plant Cell Physiol.* **39**, 177–185.

Gerstein, M., Lesk, A.M., Chothia, C., 1994. Structural Mechanisms for Domain Movements in Proteins. *Biochemistry (Mosc.)* **33**, 6739–6749. doi:10.1021/bi00188a001

Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S., Duax, W.L., 2001. Porcine carbonyl reductase - Structural basis for a functional monomer in short chain dehydrogenases/reductases. *J. Biol. Chem.* **276**, 18457–18463. doi:10.1074/jbc.M100538200

Gibson, K., Neuberger, A., Scott, J., 1955. Purification and Properties of Delta-Aminolaevulic Acid Dehydrase. *Biochem. J.* **61**, 618–629.

Gibson, L., Willows, R., Kannangara, C., Von Wettstein, D., Hunter, C., 1995. Magnesium-Protoporphyrin Chelatase of Rhodobacter-Sphaeroides - Reconstitution of Activity by Combining the Products of the Bchh, Bchi, and Bchd Genes Expressed in Escherichia-Coli. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1941–1944. doi:10.1073/pnas.92.6.1941

Gough, S.P., Petersen, B.O., Duus, J.Ø., 2000. Anaerobic chlorophyll isocyclic ring formation in Rhodobacter capsulatus requires a cobalamin cofactor. *Proc. Natl. Acad. Sci.* **97**, 6908–6913. doi:10.1073/pnas.97.12.6908

Granick, S., 1954. Enzymatic Conversion of Delta-Amino Levulinic Acid to Porphobilinogen. *Science* **120**, 1105–1106. doi:10.1126/science.120.3131.1105

Griffiths, W., 1978. Reconstitution of Chlorophyllide Formation by Isolated Etioplast Membranes. *Biochem. J.* **174**, 681–692.

Grzesiek, S., Bax, A., 1993. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR* **3**, 185–204. doi:10.1007/BF00178261

Grzyb, J.M., Solymosi, K., Strzałka, K., Mysliwa-Kurdziel, B., 2013. Visualization and characterization of prolamellar bodies with atomic force microscopy. *J. Plant Physiol.* **170**, 1217–1227. doi:10.1016/j.jplph.2013.04.017

Grzybowska, E., Góra, M., Plochocka, D., Rytka, J., 2002. Saccharomyces cerevisiae Ferrochelatase Forms a Homodimer. *Arch. Biochem. Biophys.* **398**, 170–178. doi:10.1006/abbi.2001.2730

Gunning, B.E.S., 2001. Membrane geometry of "open" prolamellar bodies. *Protoplasma* **215**, 4–15. doi:10.1007/BF01280299

Heyes, D.J., Heathcote, P., Rigby, S.E.J., Palacios, M.A., Grondelle, R. van, Hunter, C.N., 2006. The First Catalytic Step of the Light-driven Enzyme Protochlorophyllide Oxidoreductase Proceeds via a Charge Transfer Complex. *J. Biol. Chem.* **281**, 26847–26853. doi:10.1074/jbc.M602943200

Heyes, D.J., Hunter, C.N., 2004. Identification and Characterization of the Product Release Steps within the Catalytic Cycle of Protochlorophyllide Oxidoreductase. *Biochemistry (Mosc.)* **43**, 8265–8271. doi:10.1021/bi049576h

Heyes, D.J., Hunter, C.N., 2005. Making light work of enzyme catalysis: protochlorophyllide oxidoreductase. *Trends Biochem. Sci.* **30**, 642–649. doi:10.1016/j.tibs.2005.09.001

Heyes, D.J., Kruk, J., Hunter, C.N., 2006. Spectroscopic and kinetic characterization of the light-dependent enzyme protochlorophyllide oxidoreductase (POR) using monovinyl and divinyl substrates. *Biochem. J.* **394**, 243. doi:10.1042/BJ20051635

Heyes, D.J., Levy, C., Sakuma, M., Robertson, D.L., Scrutton, N.S., 2011. A Twin-track Approach Has Optimized Proton and Hydride Transfer by Dynamically Coupled Tunneling during the Evolution of Protochlorophyllide Oxidoreductase. *J. Biol. Chem.* **286**, 11849–11854. doi:10.1074/jbc.M111.219626

Heyes, D.J., Martin, G.E.M., Reid, R.J., Hunter, C.N., Wilks, H.M., 2000. NADPH:protochlorophyllide oxidoreductase from Synechocystis: overexpression, purification and preliminary characterisation. *FEBS Lett.* **483**, 47–51. doi:10.1016/S0014-5793(00)02081-0

Heyes, D.J., Menon, B.R.K., Sakuma, M., Scrutton, N.S., 2008. Conformational Events during Ternary Enzyme-Substrate Complex Formation Are Rate Limiting in the Catalytic Cycle of the Light-Driven Enzyme Protochlorophyllide Oxidoreductase. *Biochemistry (Mosc.)* **47**, 10991–10998. doi:10.1021/bi801521c

Heyes, D.J., Ruban, A.V., Hunter, C.N., 2003. Protochlorophyllide Oxidoreductase: "Dark" Reactions of a Light-Driven Enzyme. *Biochemistry (Mosc.)* **42**, 523–528. doi:10.1021/bi0268448

Heyes, D.J., Ruban, A.V., Wilks, H.M., Hunter, C.N., 2002. Enzymology below 200 K: The kinetics and thermodynamics of the photochemistry catalyzed by protochlorophyllide oxidoreductase. *Proc. Natl. Acad. Sci.* **99**, 11145–11150. doi:10.1073/pnas.182274199

Heyes, D.J., Sakuma, M., Scrutton, N.S., 2007. Laser Excitation Studies of the Product Release Steps in the Catalytic Cycle of the Light-driven Enzyme, Protochlorophyllide Oxidoreductase. *J. Biol. Chem.* **282**, 32015–32020. doi:10.1074/jbc.M706098200

Heyes, D.J., Sakuma, M., Visser, S.P. de, Scrutton, N.S., 2009. Nuclear Quantum Tunneling in the Light-activated Enzyme Protochlorophyllide Oxidoreductase. *J. Biol. Chem.* **284**, 3762–3767. doi:10.1074/jbc.M808548200

Hinchigeri, S.B., Hundle, B., Richards, W.R., 1997. Demonstration that the BchH protein of Rhodobacter capsulatus activates S-adenosyl-l-methionine:magnesium protoporphyrin IX methyltransferase. *FEBS Lett.* **407**, 337–342. doi:10.1016/S0014-5793(97)00371-2

Hollingshead, S., Kopecna, J., Jackson, P.J., Canniffe, D.P., Davison, P.A., Dickman, M.J., Sobotka, R., Hunter, C.N., 2012. Conserved Chloroplast Open-reading Frame ycf54 Is Required for Activity of the Magnesium Protoporphyrin Monomethylester Oxidative Cyclase in Synechocystis PCC 6803. *J. Biol. Chem.* **287**, 27823–27833. doi:10.1074/jbc.M112.352526

Hsueh, H.T., Chu, H., Chang, C.C., 2007. Identification and characteristics of a cyanobacterium isolated from a hot spring with dissolved inorganic carbon. *Environ. Sci. Technol.* **41**, 1909–1914. doi:10.1021/es0620639

Ikeda, T., 1968. Analytical Studies on Structure of Prolamellar Body. *Bot. Mag.-Tokyo* **81**, 517–&.

Islam, M.R., Aikawa, S., Midorikawa, T., Kashino, Y., Satoh, K., Koike, H., 2008. slr1923 of Synechocystis sp PCC6803 is essential for conversion of 3,8-divinyl(proto)chlorophyll(ide) to 3-monovinyl(proto)chlorophyll(ide). *Plant Physiol.* **148**, 1068–1081. doi:10.1104/pp.108.123117

Ito, H., Yokono, M., Tanaka, R., Tanaka, A., 2008. Identification of a novel vinyl reductase gene essential for the biosynthesis of monovinyl chlorophyll in Synechocystis sp PCC6803. *J. Biol. Chem.* **283**, 9002–9011. doi:10.1074/jbc.M708369200

Jacobs, N.J., Jacobs, J.M., 1975. Fumarate as alternate electron acceptor for the late steps of anaerobic heme synthesis in Escherichia coli. Biochem. *Biophys. Res. Commun.* **65**, 435–441. doi:10.1016/S0006-291X(75)80112-4

Jacobs, N.J., Jacobs, J.M., 1976. Nitrate, fumarate, and oxygen as electron acceptors for a late step in microbial heme synthesis. *Biochim. Biophys. Acta BBA - Bioenerg.* **449**, 1–9. doi:10.1016/0005-2728(76)90002-5

Jaffe, E., Abrams, W., Kaempfen, H., Harris, K., 1992. 5-Chlorolevulinate Modification of Porphobilinogen Synthase Identifies a Potential Role for the Catalytic Zinc. *Biochemistry (Mosc.)* **31**, 2113–2123. doi:10.1021/bi00122a032

Jensen, P.E., Gibson, L.C.D., Henningsen, K.W., Hunter, C.N., 1996. Expression of the chlI, chlD, and chlH Genes from the Cyanobacterium Synechocystis PCC6803 in Escherichia coli and Demonstration That the Three Cognate Proteins Are Required for Magnesium-protoporphyrin Chelatase Activity. *J. Biol. Chem.* **271**, 16662–16667.

Jensen, P.E., Gibson, L.C., Hunter, C.N., 1998. Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphyrin IX chelatase from Synechocystis PCC68031. *Biochem. J.* **334**, 335–344.

Jensen, P.E., Gibson, L.C., Hunter, C.N., 1999. ATPase activity associated with the magnesium-protoporphyrin IX chelatase enzyme of Synechocystis PCC6803: evidence for ATP hydrolysis during Mg2+ insertion, and the MgATP-dependent interaction of the ChlI and ChlD subunits. *Biochem. J.* **339**, 127–134.

Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., Krauss, N., 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution. *Nature* **411**, 909–917. doi:10.1038/35082000

Jordan, P.M., Burton, G., Nordlöv, H., Schneider, M.M., Pryde, L., Scott, A.I., 1979. Pre-uroporphyrinogen: a substrate for uroporphyrinogen III cosynthetase. *J. Chem. Soc. Chem. Commun.* 204–205. doi:10.1039/C39790000204

Jordan, P.M., Seehra, J.S., 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. doi:10.1016/0014-5793(79)80853-4

Jordan, P.M., Warren, M.J., 1987. Evidence for a dipyrromethane cofactor at the catalytic site of E. coli porphobilinogen deaminase. *FEBS Lett.* **225**, 87–92. doi:10.1016/0014-5793(87)81136-5

Jornvall, H., Hoog, J.O., Persson, B., 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *Febs Lett.* **445**, 261–264. doi:10.1016/S0014-5793(99)00130-1

Kaneko, T., Tabata, S., 1997. Complete genome structure of the unicellular cyanobacterium Synechocystis sp. PCC6803. *Plant Cell Physiol.* **38**, 1171–1176.

Kaschner, M., Loeschcke, A., Krause, J., Minh, B.Q., Heck, A., Endres, S., Svensson, V., Wirtz, A., von Haeseler, A., Jaeger, K.-E., Drepper, T., Krauss, U., 2014. Discovery of the first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria. *Mol. Microbiol.* **93**, 1066–1078. doi:10.1111/mmi.12719

Kavanagh, K.L., Jornvall, H., Persson, B., Oppermann, U., 2008. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906. doi:10.1007/s00018-008-8588-y

Kay, L.E., Torchia, D.A., Bax, A., 1989. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry (Mosc.)* **28**, 8972–8979. doi:10.1021/bi00449a003

Klement, H., Helfrich, M., Oster, U., Schoch, S., Rüdiger, W., 1999. Pigment-free NADPH:protochlorophyllide oxidoreductase from Avena sativa L. *Eur. J. Biochem.* **265**, 862–874. doi:10.1046/j.1432-1327.1999.00627.x

Koumanov, A., Benach, J., Atrian, S., Gonzalez-Duarte, R., Karshikoff, A., Ladenstein, R., 2003. The catalytic mechanism of Drosophila alcohol dehydrogenase: Evidence for a proton relay modulated by the coupled ionization of the active site lysine/tyrosine pair and a NAD(+) ribose OH switch. *Proteins-Struct. Funct. Genet.* **51**, 289–298. doi:10.1002/prot.10354

Kovacevic, D., Dewez, D., Popovic, R., 2007. Irradiation-induced in vivo re-localization of NADPH-protochlorophyllide oxidoreductase from prolamellar body to stroma of barley etioplast. *Photosynthetica* **45**, 105–109. doi:10.1007/s11099-007-0016-2

Kozlowski, L.P., Bujnicki, J.M., 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC *Bioinformatics* **13**, 111. doi:10.1186/1471-2105-13-111

Krogh, A., Brown, M., Mian, I.S., Sjölander, K., Haussler, D., 1994. Hidden Markov Models in Computational Biology: Applications to Protein Modeling. *J. Mol. Biol.* **235**, 1501–1531. doi:10.1006/jmbi.1994.1104

Laemmli, U., 1970. Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature* **227**, 680–&. doi:10.1038/227680a0

Lake, V., Olsson, U., Willows, R.D., Hansson, M., 2004. ATPase activity of magnesium chelatase subunit I is required to maintain subunit D in vivo. *Eur. J. Biochem.* **271**, 2182–2188. doi:10.1111/j.1432-1033.2004.04143.x

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. doi:10.1093/bioinformatics/btm404

Larkin, R.M., Alonso, J.M., Ecker, J.R., Chory, J., 2003. GUN4, a Regulator of Chlorophyll Synthesis and Intracellular Signaling. *Science* **299**, 902–906. doi:10.1126/science.1079978

Lebedev, N., Karginova, O., McIvor, W., Timko, M.P., 2001. Tyr275 and Lys279 Stabilize NADPH within the Catalytic Site of NADPH:Protochlorophyllide Oxidoreductase and Are Involved in the Formation of the Enzyme Photoactive State. *Biochemistry (Mosc.)* **40**, 12562–12574. doi:10.1021/bi0105025

Lebedev, N., Timko, M.P., 1998. Protochlorophyllide photoreduction. *Photosynth. Res.* **58**, 5–23. doi:10.1023/A:1006082119102

Lebedev, N., Timko, M.P., 1999. Protochlorophyllide oxidoreductase B-catalyzed protochlorophyllide photoreduction in vitro: Insight into the mechanism of chlorophyll formation in light-adapted plants. *Proc. Natl. Acad. Sci.* **96**, 9954–9959. doi:10.1073/pnas.96.17.9954

Lebedev, N., Van Cleve, B., Armstrong, G., Apel, K., 1995. Chlorophyll Synthesis in a Deetiolated (det340) Mutant of Arabidopsis without NADPH-Protochlorophyllide (PChlide) Oxidoreductase (POR) A and Photoactive PChlide-F655. *Plant Cell* **7**, 2081–2090. doi:10.1105/tpc.7.12.2081

LeMaster, D.M., 1990. Deuterium Labeling in Nmr Structural-Analysis of Larger Proteins. *Q. Rev. Biophys.* **23**, 133–174.

Lin, K., Simossis, V.A., Taylor, W.R., Heringa, J., 2005. A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* **21**, 152–159. doi:10.1093/bioinformatics/bth487

Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Malashkevich, V.N., Hadener, A., Warren, M.J., ShoolinginJordan, P.M., 1996. The three-dimensional structure of Escherichia coli porphobilinogen deaminase at 1.76-angstrom resolution. *Proteins-Struct. Funct. Genet.* **25**, 48–78. doi:10.1002/(SICI)1097-0134(199605)25:13.0.CO;2-G

Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C., Jordan, P.M., 1992. Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature* **359**, 33–39. doi:10.1038/359033a0

Luo, J., Lim, C.K., 1993. Order of uroporphyrinogen III decarboxylation on incubation of porphobilinogen and uroporphyrinogen III with erythrocyte uroporphyrinogen decarboxylase. *Biochem. J.* **289**, 529–532.

Masuda, S., Ikeda, R., Masuda, T., Hashimoto, H., Tsuchiya, T., Kojima, H., Nomata, J., Fujita, Y., Mimuro, M., Ohta, H., Takamiya, K.-I., 2009. Prolamellar bodies formed by cyanobacterial protochlorophyllide oxidoreductase in Arabidopsis. *Plant J. Cell Mol. Biol.* **58**, 952–960. doi:10.1111/j.1365-313X.2009.03833.x

Masuda, T., 2008. Recent overview of the Mg branch of the tetrapyrrole biosynthesis leading to chlorophylls. *Photosynth. Res.* **96**, 121–143. doi:10.1007/s11120-008-9291-4

Masuda, T., Takamiya, K., 2004. Novel Insights into the Enzymology, Regulation and Physiological Functions of Light-dependent Protochlorophyllide Oxidoreductase in Angiosperms. *Photosynth. Res.* **81**, 1–29. doi:10.1023/B:PRES.0000028392.80354.7c

Mathews, M.A.A., Schubert, H.L., Whitby, F.G., Alexander, K.J., Schadick, K., Bergonia, H.A., Phillips, J.D., Hill, C.P., 2001. Crystal structure of human uroporphyrinogen III synthase. *EMBO J.* **20**, 5832–5839. doi:10.1093/emboj/20.21.5832

Mathis, P., Sauer, K., 1972. Circular dichroism studies on the structure and the photochemistry of protochlorophyllide and chlorophyllide holochrome. *Biochim. Biophys. Acta BBA - Bioenerg.* **267**, 498–511. doi:10.1016/0005-2728(72)90178-8

Mathis, P., Sauer, K., 1973. Chlorophyll Formation in Greening Bean Leaves during the Early Stages. *Plant Physiol.* **51**, 115–119. doi:10.1104/pp.51.1.115

Menon, B.R.K., Davison, P.A., Hunter, C.N., Scrutton, N.S., Heyes, D.J., 2010. Mutagenesis Alters the Catalytic Mechanism of the Light-driven Enzyme Protochlorophyllide Oxidoreductase. *J. Biol. Chem.* **285**, 2113–2119. doi:10.1074/jbc.M109.071522

Menon, B.R.K., Waltho, J.P., Scrutton, N.S., Heyes, D.J., 2009. Cryogenic and Laser Photoexcitation Studies Identify Multiple Roles for Active Site Residues in the Light-driven Enzyme Protochlorophyllide Oxidoreductase. *J. Biol. Chem.* **284**, 18160–18166. doi:10.1074/jbc.M109.020719

Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., Fujita, Y., 2010. X-ray crystal structure of the light-independent protochlorophyllide reductase. *Nature* **465**, 110–114. doi:10.1038/nature08950

Nagata, N., Tanaka, R., Satoh, S., Tanaka, A., 2005. Identification of a Vinyl Reductase Gene for Chlorophyll Synthesis in Arabidopsis thaliana and Implications for the Evolution of Prochlorococcus Species. *Plant Cell Online* **17**, 233–240. doi:10.1105/tpc.104.027276

Nagata, N., Tanaka, R., Tanaka, A., 2007. The Major Route for Chlorophyll Synthesis Includes [3,8-divinyl]-chlorophyllide a Reduction in Arabidopsis thaliana. *Plant Cell Physiol.* **48**, 1803–1808. doi:10.1093/pcp/pcm153

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., Tabata, S., 2002. Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. *Dna Res.* **9**, 123–130. doi:10.1093/dnares/9.4.123

Nomata, J., Kondo, T., Mizoguchi, T., Tamiaki, H., Itoh, S., Fujita, Y., 2014. Dark-operative protochlorophyllide oxidoreductase generates substrate radicals by an iron-sulphur cluster in bacteriochlorophyll biosynthesis. *Sci. Rep.* **4**. doi:10.1038/srep05455

Oliver, R.P., Griffiths, W.T., 1982. Pigment-Protein Complexes of Illuminated Etiolated Leaves. *Plant Physiol.* **70**, 1019–1025. doi:10.1104/pp.70.4.1019

Oster, U., Bauer, C.E., Rüdiger, W., 1997. Characterization of Chlorophyll a and Bacteriochlorophyll a Synthases by Heterologous Expression in Escherichia coli. *J. Biol. Chem.* **272**, 9671–9676. doi:10.1074/jbc.272.15.9671

Pinta, V., Picaud, M., Reiss-Husson, F., Astier, C., 2002. Rubrivivax gelatinosus acsF (Previously orf358) Codes for a Conserved, Putative Binuclear-Iron-Cluster-Containing Protein Involved in Aerobic Oxidative Cyclization of Mg-Protoporphyrin IX Monomethylester. *J. Bacteriol.* **184**, 746–753. doi:10.1128/JB.184.3.746-753.2002

Pontier, D., Albrieux, C., Joyard, J., Lagrange, T., Block, M.A., 2007. Knock-out of the Magnesium Protoporphyrin IX Methyltransferase Gene in Arabidopsis: EFFECTS ON CHLOROPLAST DEVELOPMENT AND ON CHLOROPLAST-TO-NUCLEUS SIGNALING. *J. Biol. Chem.* **282**, 2297–2304. doi:10.1074/jbc.M610286200

Porra, R.J., Falk, J.E., 1961. Protein-bound porphyrins associated with protoporphyrin biosynthesis. *Biochem. Biophys. Res. Commun.* **5**, 179–184. doi:10.1016/0006-291X(61)90106-1

Porra, R.J., Falk, J.E., 1964. The enzymic conversion of coproporphyrinogen III into protoporphyrin IX. *Biochem. J.* **90**, 69–75.

Proudfoot, A.W., 2011. Structural studies of NADPH-protochlorophyllide oxidoreductase (Ph.D.). University of Sheffield.

Raghava, G.P.S., 2002. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning.

Rebeiz, C.A., Parham, R., Fasoula, D.A., Ioannides, I.M., 2007. Chlorophyll a Biosynthetic Heterogeneity, in: Director, D.J.C.O., Ackrill, K. (Eds.), *Ciba Foundation Symposium 180 - The Biosynthesis of the Tetrapyrrole Pigments.* John Wiley & Sons, Ltd., pp. 177–193.

Reed, M.A.C., Hounslow, A.M., Sze, K.H., Barsukov, I.G., Hosszu, L.L.P., Clarke, A.R., Craven, C.J., Waltho, J.P., 2003. Effects of Domain Dissection on the Folding and Stability of the 43 kDa Protein PGK Probed by NMR. *J. Mol. Biol.* **330**, 1189–1201. doi:10.1016/S0022-2836(03)00625-9

Reid, J.D., Hunter, C.N., 2002. Current understanding of the function of magnesium chelatase. *Biochem. Soc. Trans.* **30**, 643. doi:10.1042/BST0300643

Reid, J.D., Hunter, C.N., 2004. Magnesium-dependent ATPase Activity and Cooperativity of Magnesium Chelatase from Synechocystis sp. PCC6803. *J. Biol. Chem.* **279**, 26893–26899. doi:10.1074/jbc.M400958200

Reinbothe, C., Bakkouri, M.E., Buhr, F., Muraki, N., Nomata, J., Kurisu, G., Fujita, Y., Reinbothe, S., 2010. Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction. *Trends Plant Sci.* **15**, 614–624. doi:10.1016/j.tplants.2010.07.002

Reinbothe, C., Buhr, F., Pollmann, S., Reinbothe, S., 2003a. In Vitro Reconstitution of Light-harvesting POR-Protochlorophyllide Complex with Protochlorophyllides a and b. *J. Biol. Chem.* **278**, 807–815. doi:10.1074/jbc.M209738200

Reinbothe, C., Lebedev, N., Reinbothe, S., 1999. A protochlorophyllide light-harvesting complex involved in de-etiolation of higher plants. *Nature* **397**, 80–84. doi:10.1038/16283

Reinbothe, C., Lepinat, A., Deckers, M., Beck, E., Reinbothe, S., 2003b. The Extra Loop Distinguishing POR from the Structurally Related Short-chain Alcohol Dehydrogenases Is Dispensable for Pigment Binding but Needed for the Assembly of Light-harvesting POR-Protochlorophyllide Complex. *J. Biol. Chem.* **278**, 816–822. doi:10.1074/jbc.M209739200

Rüdiger, W., 1987. Chlorophyll Synthetase and its Implication for Regulation of Chlorophyll Biosynthesis, in: Biggins, J. (Ed.), *Progress in Photosynthesis Research.* Springer Netherlands, pp. 461–467.

Sahni-Arya, B., Flynn, M.J., Bergeron, L., Salyan, M.E.K., Pedicord, D.L., Golla, R., Ma, Z., Wang, H., Seethala, R., Wu, S.C., Li, J.J., Nayeem, A., Gates, C., Hamann, L.G., Gordon, D.A., Blat, Y., 2007. Cofactor-specific modulation of 11 beta-hydroxysteroid dehydrogenase 1 inhibitor potency. *Biochim. Biophys. Acta-Proteins Proteomics* **1774**, 1184–1191. doi:10.1016/j.bbapap.2007.07.005

Salzmann, M., Wider, G., Pervushin, K., Senn, H., Wüthrich, K., 1999. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844–848. doi:10.1021/ja9834226

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

Sano, S., Granick, S., 1961. Mitochondrial Coproporphyrinogen Oxidase and Protoporphyrin Formation. *J. Biol. Chem.* **236**, 1173–1180.

Sarma, R., Barney, B.M., Hamilton, T.L., Jones, A., Seefeldt, L.C., Peters, J.W., 2008. Crystal Structure of the L Protein of Rhodobacter sphaeroides Light-Independent Protochlorophyllide Reductase with MgADP Bound: A Homologue of the Nitrogenase Fe Protein. *Biochemistry (Mosc.)* **47**, 13004–13015. doi:10.1021/bi801058r

Scheer, H., 1991. Structure and occurence of chlorophylls, in: *Chlorophylls.* CRC Press, Cleveland, Ohio, pp. 3–30.

Schmid, R., Shemin, D., 1955. The Enzymatic Formation of Porphobilinogen from Delta-Aminolevulinic Acid and Its Conversion to Protoporphyrin. *J. Am. Chem. Soc.* **77**, 506–507. doi:10.1021/ja01607a103

Schoch, S., 1978. The esterification of chlorophyllide a in greening bean leaves. *Z. Fuer Naturforschung Sect. C Biosci.*

Schwarzinger, S., Wright, P.E., Dyson, H.J., 2002. Molecular Hinges in Protein Folding: the Urea-Denatured State of Apomyoglobin. *Biochemistry (Mosc.)* **41**, 12681–12686. doi:10.1021/bi020381o

Shen, Y., Bax, A., 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241. doi:10.1007/s10858-013-9741-y

Siepker, L.J., Ford, M., de Kock, R., Kramer, S., 1987. Purification of bovine protoporphyrinogen oxidase: immunological cross-reactivity and structural relationship to ferrochelatase. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **913**, 349–358. doi:10.1016/0167-4838(87)90146-4

Sofia, H.J., Chen, G., Hetzler, B.G., Reyes-Spindola, J.F., Miller, N.E., 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. doi:10.1093/nar/29.5.1097

Soll, J., Schultz, G., Rüdiger, W., Benz, J., 1983. Hydrogenation of Geranylgeraniol Two Pathways Exist in Spinach Chloroplasts. *Plant Physiol.* **71**, 849–854. doi:10.1104/pp.71.4.849

Solymosi, K., Schoefs, B., 2008. Prolamellar body: a unique plastid compartment, which does not only occur in dark-grown leaves. *Plant Cell Compart. Sel. Top.* 151–202.

Solymosi, K., Schoefs, B., 2010. Etioplast and etio-chloroplast formation under natural conditions: the dark side of chlorophyll biosynthesis in angiosperms. *Photosynth. Res.* **105**, 143–166. doi:10.1007/s11120-010-9568-2

Stark, W.M., Hawker, C.J., Hart, G.J., Philippides, A., Petersen, P.M., Lewis, J.D., Leeper, F.J., Battersby, A.R., 1993. Biosynthesis of porphyrins and related macrocycles. Part 40. Synthesis of a spiro-lactam related to the proposed spiro-intermediate for porphyrin biosynthesis: inhibition of cosynthetase. *J. Chem. Soc. [Perkin 1]* 2875–2892. doi:10.1039/P19930002875

Stott, K., Stonehouse, J., Keeler, J., Hwang, T.-L., Shaka, A.J., 1995. Excitation Sculpting in High-Resolution Nuclear Magnetic Resonance Spectroscopy: Application to Selective NOE Experiments. *J. Am. Chem. Soc.* **117**, 4199–4200. doi:10.1021/ja00119a048

Sturgis, J.N., Tucker, J.D., Olsen, J.D., Hunter, C.N., Niederman, R.A., 2009. Atomic Force Microscopy Studies of Native Photosynthetic Membranes. *Biochemistry (Mosc.)* **48**, 3679–3698. doi:10.1021/bi900045x

Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288. doi:10.1093/bioinformatics/btm098

Suzuki, J.Y., Bollivar, D.W., Bauer, C.E., 1997. Genetic Analysis of Chlorophyll Biosynthesis. *Annu. Rev. Genet.* **31**, 61–89. doi:10.1146/annurev.genet.31.1.61

Sytina, O.A., Heyes, D.J., Hunter, C.N., Alexandre, M.T., van Stokkum, I.H.M., van Grondelle, R., Groot, M.L., 2008. Conformational changes in an ultrafast light-driven enzyme determine catalytic activity. *Nature* **456**, 1001–1004. doi:10.1038/nature07354

Sytina, O.A., van Stokkum, I.H.M., Heyes, D.J., Hunter, C.N., Groot, M.L., 2012. Spectroscopic characterization of the first ultrafast catalytic intermediate in protochlorophyllide oxidoreductase. *Phys. Chem. Chem. Phys.* **14**, 616. doi:10.1039/c1cp21713e

Tait, G.H., 1969. Coproporphyrinogenase activity in extracts from rhodopseudomonas spheroides. *Biochem. Biophys. Res. Commun.* **37**, 116–122. doi:10.1016/0006-291X(69)90888-2

Tait, G.H., 1972. Coproporphyrinogenase activities in extracts of Rhodopseudomonas spheroides and Chromatium strain D. *Biochem J* **128**, 1159–1169.

Tanaka, R., Tanaka, A., 2007. Tetrapyrrole Biosynthesis in Higher Plants. *Annu. Rev. Plant Biol.* **58**, 321–346. doi:10.1146/annurev.arplant.57.032905.105448

Townley, H.E., Sessions, R.B., Clarke, A.R., Dafforn, T.R., Griffiths, W.T., 2001. Protochlorophyllide oxidoreductase: A homology model examined by site-directed mutagenesis. *Proteins-Struct. Funct. Genet.* **44**, 329–335. doi:10.1002/prot.1098

Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E., Kay, L.E., 2003. Cross-correlated relaxation enhanced H-1-C-13 NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428. doi:10.1021/ja030153x

Valera, V., Fung, M., Wessler, A.N., Richards, W.R., 1987. Synthesis of 4R- and 4S-tritium labeled NADPH for the determination of the coenzyme stereospecificity of NADPH: Protochlorophyllide oxidoreductase. *Biochem. Biophys. Res. Commun.* **148**, 515–520. doi:10.1016/0006-291X(87)91141-7

Vaughan, G.D., Sauer, K., 1974. Energy transfer from protochlorophyllide to chlorophyllide during photoconversion of etiolated bean holochrome. *Biochim. Biophys. Acta BBA - Bioenerg.* **347**, 383–394. doi:10.1016/0005-2728(74)90077-2

Walker, C., Willows, R., 1997. Mechanism and regulation of Mg-chelatase. *Biochem. J.* **327**, (321-333).

Waterbury, J., Watson, S., Guillard, R., Brand, L., 1979. Widespread Occurrence of a Unicellular, Marine, Planktonic, Cyanobacterium. *Nature* **277**, 293–294. doi:10.1038/277293a0

Wilks, H.M., Timko, M.P., 1995. A light-dependent complementation system for analysis of NADPH:protochlorophyllide oxidoreductase: identification and mutagenesis of two conserved residues that are essential for enzyme activity. *Proc. Natl. Acad. Sci.* **92**, 724–728.

Willows, R.D., Hansson, A., Birch, D., Al-Karadaghi, S., Hansson, M., 2004. EM single particle analysis of the ATP-dependent BchI complex of magnesium chelatase: an AAA+ hexamer. J. Struct. Biol., AAA+ Proteins: Based on Contributions to the 5th International Conference on AAA+ Proteins at Airlie House, Virginia **146**, 227–233. doi:10.1016/j.jsb.2003.11.019

Willows, R.D., Kriegel, A.M., 2009a. *Biosynthesis of Bacteriochlorophylls in Purple Bacteria*.

Willows, R.D., Kriegel, A.M., 2009b. Biosynthesis of Bacteriochlorophylls in Purple Bacteria, in: Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty, J.T. (Eds.), *The Purple Phototrophic Bacteria, Advances in Photosynthesis and Respiration.* Springer Netherlands, pp. 57–79.

Wu, C.-K., Dailey, H.A., Rose, J.P., Burden, A., Sellers, V.M., Wang, B.-C., 2001. The 2.0 Å structure of human ferrochelatase, the terminal enzyme of heme biosynthesis. *Nat. Struct. Mol. Biol.* **8**, 156–160. doi:10.1038/84152

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40. doi:10.1186/1471-2105-9-40

Zouni, A., Witt, H.T., Kern, J., Fromme, P., Krauss, N., Saenger, W., Orth, P., 2001. Crystal structure of photosystem II from Synechococcus elongatus at 3.8 angstrom resolution. *Nature* **409**, 739–743. doi:10.1038/35055589

Appendix A

Media and Buffer Recipes

A1 Growth Media

A1.1 LB Agar

For each litre, the following compounds were added to 950 ml Milli-Q water:

40 g LB-Agar Miller

This was made up to to 1 litre with Milli-Q water, prior to sterilising by autoclaving.

A1.2 LB Media

For each litre, the following compounds were added to 950 ml Milli-Q water:

25 g LB-Broth Miller

The pH was adjusted to 7.0 and the volume was made up to 1 litre with Milli-Q water, prior to sterilising by autoclaving.

A1.3 M9 Media

For each litre, the following compounds were added to 950 ml Milli-Q water[1]:

6 g Na_2HPO_4 3 g KH_2PO_4

0.5 g NaCl

The pH was adjusted to 7.4 and the volume was made up to 1 litre with Milli-Q water, prior to sterilising by autoclaving.

After cooling the following chemicals were added to the media:

650 µl Trace elements [2] 3 g Glucose [3]

1 ml Thiamine (1 mg ml^{-1}) [4] 2 ml $(^{15}NH_4)_2SO_4$ (0.5 $mg.ml^{-1}$) [4]

1 ml 1M $MgSO_4$ (autoclaved) 0.1 ml 1M $CaCl_2$ (autoclaved) [5]

Notes:

[1] Deuterium Oxide (99.8 % purity) was used instead of Milli-Q water when conducting triple labelled growths. After adjusting the pH to 7.4 (corrected to pH 7.8 for D_2O, as increase in 0.4

pH units corresponds to pD) the medium was filtered through a sterile 0.2 µm filter, rather than autoclaving. Empty flasks were autoclaved separately and dried under UV light, prior to the media being transferred.

[2] Trace elements. For 100 ml final volume the following compounds were added to 70 ml Milli-Q water:

550 mg $CaCl_2.2H_2O$	140 mg $MnSO_4.H_2O$
40 mg $CuSO_4.5H_2O$	220 mg $ZnSO_4.7H_2O$
45 mg $CoCl_2.6H_2O$	26 mg $Na_2MoO_4.2H_2O$
40 mg H_3BO_4	26 mg KI

The pH was adjusted to 8.0 before adding:

500 mg EDTA

The pH was then readjusted to 8.0 before adding:

375 mg $FeSO_4.7H_2O$

The solution was then made up to 100 ml volume with Milli-Q water prior to autoclaving.

[3] Labelled glucose (^{13}C 99 % and/or 1,2,3,4,5,6,6-D_7, 98 %) was used here if required, at 1.5 g L^{-1}.

[4] 0.2 µm syringe-filter sterilised.

[5] $CaCl_2$ was added last and the flask was immediately agitated to disperse the precipitate. The preparation was abandoned if the precipitate was not successfully dispersed.

A1.4 M22 (10 x Stock) (4 L)

For 4 litres, the following compounds were added to 3 L Milli-Q water:

122.4 g KH_2PO_4	120 g K_2HPO_4
100 g Na lactate solution	20 g NaCl
20 g $(NH_4)_2SO_4$	173.7 g Sodium Succinate
10.8 g Sodium Glutamate	1.6 g Aspartic acid

800 ml Solution C[1]

The pH was adjusted to 6.8 and the volume was made up to 4 L with Milli-Q water, prior to sterilising by autoclaving.

Notes:

[1] Solution C. For 4 litres, the following compounds were added to 3 L Milli-Q water:

40 g	Nitrilotriacetic acid (NTA)	96 g	$MgCl_2$
13.36 g	$CaCl_2$	0.5 g	EDTA
1.044 g	$ZnCl_2$	1.0 g	$FeCl_2$
0.36 g	$MnCl_2$	0.037 g	$(NH_4)_6Mo_7O_{24}(H_2O)_4$
0.031 g	$CuCl_2$	0.0496 g	$Co(NO_3)_2$
0.0228 g	Boric acid		

The solution was made up to 4 L and frozen at -20 °C in 800 ml aliquots.

A1.5 Vitamins (10, 000 x)

The following compounds were made up to 100 ml:

1 g	Nicotimic acid	0.5 g	Thiamine
0.1 g	4-aminobenzoic acid	10 mg	Biotin

The solution was filter sterilised through a 0.2 µm syringe filter, dispensed into 10 ml aliquots and stored in the freezer.

A1.6 M22 Agar (200 ml)

The following components were combined to a final 200 ml:

200 ml	1 x M22	3 g	Bacto Agar

This was subsequently sterilised by autoclaving.

A1.7 M22 (1 x) Media (4 L)

For 4 litres, the following were added to 3.52 L Milli-Q water:

400 ml	10 x M22 (A1.5)	80 ml	5% CAA[1]

[1]CAA. 50 g Casein Hydrosylate acid in 1 L Milli-Q water, aliquotted into 200 ml.

The media was sterilising by autoclaving, with Vitamins added to the media in a 1 in 10,000 ratio, immediately prior to use.

A2 Antibiotics and IPTG

A2.1 Kanamycin

Kanamycin was dissolved in Milli-Q water to a concentration of 30 mg ml^{-1}. This stock solution was sterilised by filtration through a 0.2 µm syringe-filter and was stored at -20 °C in aliquots. When required, the aliquots were thawed on ice and added to a final concentration of 30 mg L^{-1}.

A2.2 Chloramphenicol

Chloramphenicol was dissolved in ethanol to a concentration of 35 mg ml^{-1}. This stock solution did not require sterilisation due to the presence of ethanol, but was still stored at -20 °C. When required, the stock solution was added to a final concentration of 25 mg L^{-1}.

A2.3 Rifampicin

Rifampicin was dissolved in methanol to a concentration of 25 mg ml^{-1}. This stock solution did not require sterilisation due to the presence of methanol, but was stored at -20 °C in a light resistant container. When required, the stock solution was added to a final concentration as detailed in 2.3.

A2.4 IPTG (Isopropyl-β-D-galactosidase)

IPTG was dissolved in Milli-Q water to a concentration of 238 mg ml^{-1} (1 M). This stock solution was sterilised by filtration through a 0.2 µm syringe-filter and stored at -20 °C in 1 ml aliquots. When required, the aliquots were thawed and added to the media to a final concentration as detailed in 2.6.4.

A3 Buffers and Reagents

A3.1 LEW Binding Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

7.10 g	Na_2HPO_4	35.06 g	NaCl

The pH was adjusted to 8.0 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.2 LEW Wash Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

7.10 g	Na_2HPO_4	35.06 g	NaCl
0.68 g	Imidazole		

The pH was adjusted to 8.0 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.3 LEW Elution Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

7.10 g	Na_2HPO_4	35.06 g	NaCl
34.04 g	Imidazole		

The pH was adjusted to 8.0 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.4 Low Salt SP Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

1.06 g	Na_2HPO_4	2.73 g	NaH_2PO_4
11.69 g	NaCl	0.15 g	DTT
Roche Complete protease inhibitor			

The pH was adjusted to 6.5 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.5 High Salt SP Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

1.06 g	Na_2HPO_4	2.73 g	NaH_2PO_4
11.69 g	NaCl	0.15 g	DTT
Roche Complete protease inhibitor			

The pH was adjusted to 6.5 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.6 NMR Buffer

For each litre, the following compounds were added to 900 ml Milli-Q water:

2.8 ml	500 mM Na_2HPO_4	97.2 ml	500 mM NaH_2PO_4
5.84 g	NaCl	0.77 g	DTT

Roche Complete protease inhibitor

The pH was checked to be 5.5, prior to vacuum sterilisation through a 0.2 µm filter and de-gassing.

A3.7 Assay Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

6.06 g	TRIS	5.84 g	NaCl
0.1 %	β-mercaptoethanol	0.1 %	Triton X-100

The pH was adjusted to 7.4 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.8 Complex Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

6.06 g	TRIS	5.84 g	NaCl
14.5 g	BOG	0.77 g	DTT

Roche Complete Protease Inhibitor

The pH was adjusted to 7.4 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.9 HPLC Buffer

For each litre, the following components were added to 650 ml Milli-Q water:

300 ml	Methanol	23.12 g	Ammonium Acetate

The pH was adjusted to 6.9 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter and de-gassing.

A3.10 Low Salt Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

6.06 g	TRIS	11.69 g	NaCl
0.15 g	DTT	Roche Complete Protease Inhibitor	

The pH was adjusted to 7.4 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

A3.10 Solubilisation Buffer

For each litre, the following compounds were added to 950 ml Milli-Q water:

6.06 g	TRIS	5.84 g	NaCl
36.25 g	BOG	0.77 g	DTT
Roche Complete Protease Inhibitor			

The pH was adjusted to 7.4 and the volume was made up to 1 litre with Milli-Q water, prior to vacuum sterilisation through a 0.2 µm filter.

Appendix B

NMR parameters and pulseprograms

```
::::::::::::::
hncacb_listpars
::::::::::::::
Listpar.py output for dataset 18/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncacbetgp2h3d.ffm
      NS( )  = 56
      DS( )  = 128
      RG( )  = 905
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = -5.0   dB
      P1(not_determined)  = 9.050  u (apparently at PL1)
      P2(not_determined)  : 18.10  u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 29.2   dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.3664353 Mhz
      O2( )  = 9054.30  hz( ~  40.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = -2.9   dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = -2.9   dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 171.0   u
         SP3(not_determined)  = -3.1   dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = -3.1   dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : 30331.880
Hz
      BF2( )  : 226.357381 Mhz
  CNST21( )  = 174.0
f3 ----------------
    SFO3( )  = 91.22829627 Mhz
      O3( )  = 10809.270 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.188268751 Mhz
      O4( )  = 552.751 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 1024
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs =  512; aq =  35.5 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 4.00029565178
    IN10( )  : 91.35  u
    IN30( )  : 91.35  u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 80.0   ppm
     ND0( )  = 1.00040597027
     IN0( )  : 55.20  u
    IN20( )  : 100.0  u
     TD1( )  = 106
swh =  18108.6 Hz; nprs =   53; aq =   2.9 m;
true_dwell_time =  55.2 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
    GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 2.0
    GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 2.0
    GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = -80.0
    GPNAM4( )  = SMSQ10.100     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 2.0
    GPNAM5( )  = SMSQ10.100     GPX5( )  = 0.0
GPY5( )  = 0.0          GPZ5( )  = 30.0
    GPNAM6( )  = SMSQ10.100     GPX6( )  = 0.0
GPY6( )  = 0.0          GPZ6( )  = 45.0
    GPNAM7( )  = SMSQ10.100     GPX7( )  = 0.0
GPY7( )  = 0.0          GPZ7( )  = 30.13
Miscellaneous
.. set by user
      D1( )  = 1000.0  m
     D16(not_determined)  = 200.0  u
    INF1( )  = 55.2202007497
    INF2( )  = 365.383709829
     P21(not_determined)  = 36.0   u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  : 3.00   u
     D10( )  : 3.00  u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D20( )  : 0.00   u
     D23( )  : 11.00  m
     D25( )  : 2.30   m
     D26( )  : 2.30   m
     D27( )  : 0.00   u
     D28( )  : 3.60   m
     D30( )  : 3.66   m
      D4( )  : 1.70   m
   DATE( )  : 1351873261.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 72.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
hncaco_listpars
::::::::::::::
Listpar.py output for dataset
/ld10/davida/utrecht_data_4/4/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncacoetgp2h3d
      NS( )  = 32
      DS( )  = 128
      RG( )  = 905
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 8.750  u (apparently at PL1)
      P2(not_determined)  : 17.50  u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.3972199 Mhz
      O2( )  = 39838.90  hz( ~ 176.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = 120.0  dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = 120.0  dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 171.0   u
         SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = 120.0  dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : 27615.60
Hz
         SP7(not_determined)  = 120.0  dB
SPNAM7(not_determined)  = Q3.1000  SPOFFS7( )  : -27615.60
Hz
     P24(not_determined)  = 213.0   u
         SP9(not_determined)  = 120.0  dB
SPNAM9(not_determined)  = Q3.1000  SPOFFS9( )  : 0.000  Hz
      BF2( )  : 226.357381 Mhz
  CNST21( )  = 176.0    (equivalent sfo= 226.3972199)
  CNST22( )  = 54.0     (equivalent sfo= 226.3696043)
f3 ----------------
    SFO3( )  = 91.22829627 Mhz
      O3( )  = 10809.270 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.18826875 Mhz
      O4( )  = 552.750 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 2048
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs = 1024; aq =  71.0 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 4.00029565178
    IN10( )  : 91.35  u
    IN30( )  : 91.35  u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 18.0   ppm
     ND0( )  = 4.00053677863
     IN0( )  : 61.35  u
     TD1( )  = 128
swh =   4074.4 Hz; nprs =   64; aq =  15.7 m;
true_dwell_time = 245.4 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
    GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 2.0
    GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 2.0
    GPNAM3( )  = SMSQ10.100     GPX3( )  = -80.0
GPY3( )  = 0.0          GPZ3( )  = -80.0
    GPNAM4( )  = SMSQ10.100     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 2.0
    GPNAM5( )  = SMSQ10.50      GPX5( )  = 0.0
GPY5( )  = 0.0          GPZ5( )  = 30.0
    GPNAM6( )  = SMSQ10.100     GPX6( )  = 0.0
GPY6( )  = 0.0          GPZ6( )  = 45.0
    GPNAM7( )  = SMSQ10.100     GPX7( )  = 0.0
GPY7( )  = 0.0          GPZ7( )  = 30.13
Miscellaneous
.. set by user
      D1( )  = 1000.0  m
     D16(not_determined)  = 200.0  u
    INF1( )  = 245.389742773
    INF2( )  = 365.383709829
     P21(not_determined)  = 37.50  u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  : 3.00   u
     D10( )  : 3.00  u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D22( )  : 4.00   m
     D23( )  : 12.00  m
     D25( )  : 2.30   m
     D26( )  : 2.30   m
     D30( )  : 4.16   m
   DATE( )  : 1350656599.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 75.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
hnca_listpars
::::::::::::::
Listpar.py output for dataset
/ld10/davida/utrecht_data_4/10/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncagp2h3d2
      NS( )  = 16
      DS( )  = 128
      RG( )  = 108
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 8.970  u (apparently at PL1)
      P2(not_determined)  : 17.940 u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.369604299 Mhz
      O2( )  = 12223.299 hz( ~  54.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = 120.0  dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = 120.0  dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 171.0   u
         SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = 120.0  dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : 27615.60
Hz
      BF2( )  : 226.357381 Mhz
  CNST21( )  = 176.0
f3 ----------------
    SFO3( )  = 91.2282962722 Mhz
      O3( )  = 10809.272 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.188268751 Mhz
      O4( )  = 552.751 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 1024
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs =  512; aq =  35.5 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 2.00014782594
    IN10( )  : 182.70 u
    IN30( )  : 182.70 u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 32.0   ppm
     ND0( )  = 4.001623881
     IN0( )  : 34.50  u
     TD1( )  = 120
swh =   7243.4 Hz; nprs =   60; aq =   8.3 m;
true_dwell_time = 138.1 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
    GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 30.0
    GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 18.0
    GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 44.0
Miscellaneous
.. set by user
      D1( )  = 1.50   s
     D16(not_determined)  = 200.0  u
    INF1( )  = 138.048569272
    INF2( )  = 365.383709829
     P21(not_determined)  = 37.50  u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  : 3.00   u
     D10( )  : 3.00   u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D23( )  : 12.00  m
     D26( )  : 2.30   m
     D30( )  : 10.98  m
   DATE( )  : 1351095511.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 75.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
hncocacb_listpars
::::::::::::::
Listpar.py output for dataset 15/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncocacbetgp2h3d.ffm
      NS( )  = 24
      DS( )  = 128
      RG( )  = 905
f1 ----------------
    SFO1( )  = 900.214233 Mhz
```

```
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 9.150  u (apparently at PL1)
      P2(not_determined)  : 18.30  u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.3664353 Mhz
      O2( )  = 9054.30  hz( ~  40.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = 120.0  dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = 120.0  dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 171.0   u
         SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = 120.0  dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : -27162.880
Hz
         SP7(not_determined)  = 120.0  dB
SPNAM7(not_determined)  = Q3.1000  SPOFFS7( )  : 30331.890
Hz
     BF2( )  : 226.357381 Mhz
  CNST21( )  = 174.0    (equivalent sfo= 226.3967672)
  CNST22( )  = 54.0
  CNST23( )  = 40.0     (equivalent sfo= 226.3664353)
f3 ----------------
    SFO3( )  = 91.22829627 Mhz
      O3( )  = 10809.270 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.18826875 Mhz
      O4( )  = 552.750 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 1024
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs =  512; aq =  35.5 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 4.00029565178
    IN10( )  : 91.35  u
    IN30( )  : 91.35  u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 80.0   ppm
    ND0( )  = 1.00040597027
    IN0( )  : 55.20  u
   IN20( )  : 0.00   u
     TD1( )  = 128
swh =  18108.6 Hz; nprs =   64; aq =   3.5 m;
true_dwell_time =  55.2 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
     GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 2.0
     GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 2.0
     GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = -80.0
     GPNAM4( )  = SMSQ10.100     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 2.0
     GPNAM5( )  = SMSQ10.50      GPX5( )  = 0.0
GPY5( )  = 0.0          GPZ5( )  = 30.0
     GPNAM6( )  = SMSQ10.100     GPX6( )  = 0.0
GPY6( )  = 0.0          GPZ6( )  = 45.0
     GPNAM7( )  = SMSQ10.100     GPX7( )  = 0.0
GPY7( )  = 0.0          GPZ7( )  = 30.13
Miscellaneous
.. set by user
      D1( )  = 1000.0  m
     D16(not_determined)  = 200.0  u
    INF1( )  = 55.2202007497
    INF2( )  = 365.383709829
     P21(not_determined)  = 35.0   u
 ZGOPTNS( )  =
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  = 3.00   u
     D10( )  : 3.00   u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D20( )  : 0.00   u
     D22( )  : 4.00   m
     D23( )  : 11.00  m
     D25( )  : 2.30   m
     D26( )  : 2.30   m
     D27( )  : 0.00   u
     D28( )  : 3.60   m
     D30( )  : 3.66   m
      D4( )  : 1.70   m
   DATE( )  : 1351507475.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 70.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
hncoca_listpars
::::::::::::::
Listpar.py output for dataset
/ld10/davida/utrecht_data_4/13/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncocagp2h3d
      NS( )  = 16
      DS( )  = 128
      RG( )  = 108
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 8.930  u (apparently at PL1)
      P2(not_determined)  : 17.860 u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.3696043 Mhz
      O2( )  = 12223.30  hz( ~  54.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = 120.0  dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = 120.0  dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 171.0   u
         SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = 120.0  dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : -27615.60
Hz
         SP7(not_determined)  = 120.0  dB
SPNAM7(not_determined)  = Q3.1000  SPOFFS7( )  : 27615.60
Hz
     BF2( )  : 226.357381 Mhz
  CNST21( )  = 176.0    (equivalent sfo= 226.3972199)
  CNST22( )  = 54.0     (equivalent sfo= 226.3696043)
f3 ----------------
    SFO3( )  = 91.22829627 Mhz
      O3( )  = 10809.270 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.18826875 Mhz
      O4( )  = 552.750 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 1024
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs =  512; aq =  35.5 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 2.00014552599
    IN10( )  : 182.70 u
    IN30( )  : 182.70 u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 32.0   ppm
     ND0( )  = 4.0016238886
     IN0( )  : 34.50  u
     TD1( )  = 120
swh =   7243.4 Hz; nprs =   60; aq =   8.3 m;
true_dwell_time = 138.1 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
     GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 30.0
     GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 18.0
     GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 44.0
Miscellaneous
.. set by user
      D1( )  = 1000.0  m
     D16(not_determined)  = 200.0  u
    INF1( )  = 138.048569533
    INF2( )  = 365.383289688
     P21(not_determined)  = 37.50  u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  = 3.00   u
     D10( )  : 3.00   u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D22( )  : 4.00   m
     D23( )  : 12.00  m
     D26( )  : 2.30   m
     D30( )  : 10.98  m
   DATE( )  : 1351262894.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 75.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
hnco_listpars
::::::::::::::
Listpar.py output for dataset
/ld10/davida/utrecht_data_4/7/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trhncogp2h3d
      NS( )  = 16
      DS( )  = 128
      RG( )  = 82
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL0( )  = 120.0  dB
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 8.870  u (apparently at PL1)
      P2(not_determined)  : 17.740 u (apparently at PL1)
     P11(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = Sinc1.1000  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.397219899 Mhz
      O2( )  = 39838.899 hz( ~ 176.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P13(not_determined)  = 213.0   u
         SP2(not_determined)  = 120.0  dB
SPNAM2(not_determined)  = Q5.1000  SPOFFS2( )  : 0.000  Hz
         SP8(not_determined)  = 120.0  dB
SPNAM8(not_determined)  = Q5tr.1000  SPOFFS8( )  : 0.000
Hz
     P14(not_determined)  = 170.0   u
         SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Q3.1000  SPOFFS3( )  : 0.000  Hz
         SP5(not_determined)  = 120.0  dB
SPNAM5(not_determined)  = Q3.1000  SPOFFS5( )  : -27615.60
Hz
     BF2( )  : 226.357381 Mhz
  CNST22( )  = 54.0
f3 ----------------
    SFO3( )  = 91.2282962722 Mhz
      O3( )  = 10809.272 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
f4 ----------------
    SFO4( )  = 138.188268751 Mhz
      O4( )  = 552.751 hz( ~   4.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL17(?)  = 120.0  dB
 CPDPRG4( )  = waltz16  (apparently at PL17)
   PCPD4(?)  = 110.0  u (apparently at PL17)
     BF4( )  : 138.187716 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 2048
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs = 1024; aq =  71.0 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=321)
Fdim_2 ----------------
 NUCDIM2( )  = 15N
  BFDIM2( )  = 91.217487
 FNMODE2( )  = EA
     SW2( )  = 30.0   ppm
    ND10( )  = 2.00014782594
    IN10( )  : 182.70 u
    IN30( )  : 182.70 u
     TD2( )  = 48
swh =   2736.5 Hz; nprs =   24; aq =   8.8 m;
true_dwell_time = 365.4 u
Fdim_1 ----------------
 NUCDIM1( )  = 13C
  BFDIM1( )  = 226.357381
 FNMODE1( )  = States-TPPI
     SW1( )  = 18.0   ppm
     ND0( )  = 2.00026838931
     IN0( )  : 122.70 u
     TD1( )  = 128
swh =   4074.4 Hz; nprs =   64; aq =  15.7 m;
true_dwell_time = 245.4 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
     GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 30.0
     GPNAM2( )  = SMSQ10.100     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 18.0
     GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 44.0
Miscellaneous
.. set by user
      D1( )  = 1.40   s
     D16(not_determined)  = 200.0  u
    INF1( )  = 245.389742773
    INF2( )  = 365.383709829
     P21(not_determined)  = 37.50  u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
      D0( )  : 3.00   u
     D10( )  : 3.00   u
     D11( )  : 30.00  m
     D12( )  : 20.00  u
     D23( )  : 12.00  m
     D26( )  : 2.30   m
     D30( )  : 10.98  m
   DATE( )  : 1350907493.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 75.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002


::::::::::::::
t1_listpar_single_expt
::::::::::::::
Listpar.py output for dataset 876
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[['l21', []], ['l22', []]]

 PULPROG( )  : mjc_n15_r1_pss_tr_phasecycle_2
      NS( )  = 16
      DS( )  = 32
      RG( )  = 128
f1 ----------------
    SFO1( )  = 800.31376946 Mhz
      O1( )  = 3769.460 hz( ~   4.71 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 14.30  u (apparently at PL1)
     PL11(not_determined)  = 120.0  dB
      P27(not_determined)  = 14.30  u (apparently at PL11)
      P12(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = <Sinc1.1000>  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 800.31 Mhz
f3 ----------------
    SFO3( )  = 81.1042771755 Mhz
      O3( )  = 9569.176 hz( ~ 118.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P21(not_determined)  = 29.0   u (apparently at PL3)
     BF3( )  : 81.094708 Mhz
Direct dimension -----------
      SW( )  = 16.019 ppm
      TD( )  = 3072
   DECIM( )  : 1560.0
swh =  12820.5 Hz; nprs = 1536; aq = 119.8 m;
true_dwell_time =  78.0 u
Sequence uses mc syntax (aqseq=None)
Fdim_1 ----------------
 NUCDIM1( )  = 15N
  BFDIM1( )  = 81.094708
 FNMODE1( )  = EA
```

```
     SW1( )  = 30.0   ppm
     ND0( )  = 2.000236
     IN0( )  : 205.50 u
     TD1( )  = 128
swh =   2432.8 Hz; nprs =   64; aq =  26.3 m;
true_dwell_time = 411.0 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
     GPNAM12( )  = <SMSQ10.100>    GPX12( )  = 0.0
GPY12( )  = 0.0         GPZ12( )  = 20.0
     GPNAM13( )  = <SMSQ10.100>    GPX13( )  = 0.0
GPY13( )  = 0.0         GPZ13( )  = 10.0
     GPNAM1( )  = <SMSQ10.100>     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 80.0
     GPNAM3( )  = <SMSQ10.100>     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 60.0
     GPNAM4( )  = <SMSQ10.100>     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 80.0
Miscellaneous
.. set by user
      D0( )  = 6.00   u
      D1( )  = 2.00   s
     D11( )  = 30.00  m
     D14( )  = 2.30   m
     D16(not_determined)  = 200.0  u
     D22( )  = 5.00   m
      D4( )  = 2.30   m
    INF1( )  = 411.0
     L20( )  = 121.0
     L22( )  = 1.0
 ZGOPTNS( )  = _
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 10.0   u
  DIGMOD( )  = digital
.. predefined
    DATE( )  : 1381801945.0
  DSPFVS( )  : 20.0
     L21( )  : 120.0
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z106229/0006


::::::::::::::
t2_listpar_single_expt
::::::::::::::
Listpar.py output for dataset 813
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[['l14', []], ['l15', []]]

PULPROG( )  : mjc_n15_r2_cpmg_pss_tr_3
      NS( )  = 16
      DS( )  = 32
      RG( )  = 128
f1 ----------------
    SFO1( )  = 800.31376946 Mhz
      O1( )  = 3769.460 hz( ~   4.71 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 14.0   u (apparently at PL1)
    PL14(?)  = 120.0  dB
     P14(not_determined)  = 14.0   u (apparently at PL14)
     P12(not_determined)  = 1000.0   u
         SP1(not_determined)  = 120.0  dB
SPNAM1(not_determined)  = <Sinc1.1000>  SPOFFS1( )  = 0.000
Hz
     BF1( )  : 800.31 Mhz
f3 ----------------
    SFO3( )  = 81.1042771755 Mhz
      O3( )  = 9569.176 hz( ~ 118.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
    PL23(not_determined)  = 120.0  dB
     P30(not_determined)  = 40.0   u (apparently at PL23)
     PL3(not_determined)  = 120.0  dB
     P21(not_determined)  = 29.0   u (apparently at PL3)
     BF3( )  : 81.094708 Mhz
Direct dimension -----------
      SW( )  = 16.019 ppm
      TD( )  = 3072
   DECIM( )  : 1560.0
swh =  12820.5 Hz; nprs = 1536; aq = 119.8 m;
true_dwell_time =  78.0 u
Sequence uses mc syntax (aqseq=None)
Fdim_1 ----------------
 NUCDIM1( )  = 15N
  BFDIM1( )  = 81.094708
 FNMODE1( )  = EA
     SW1( )  = 30.0   ppm
     ND0( )  = 2.000236
     IN0( )  : 205.50 u
     TD1( )  = 128
swh =   2432.8 Hz; nprs =   64; aq =  26.3 m;
true_dwell_time = 411.0 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
     GPNAM12( )  = <SMSQ10.100>    GPX12( )  = 0.0
GPY12( )  = 0.0         GPZ12( )  = 20.0
     GPNAM13( )  = <SMSQ10.100>    GPX13( )  = 0.0
GPY13( )  = 0.0         GPZ13( )  = 10.0
     GPNAM1( )  = <SMSQ10.100>     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 80.0
     GPNAM3( )  = <SMSQ10.100>     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 60.0
     GPNAM4( )  = <SMSQ10.100>     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 80.0
Miscellaneous
.. set by user
      D0( )  = 3.00   u
      D1( )  = 2.00   s
     D11( )  = 30.00  m
     D14( )  = 2.30   m
     D15(?)  = 490.0  u
     D16(not_determined)  = 200.0  u
      D4( )  = 2.30   m
    INF1( )  = 411.0
     L15( )  = 0.0
     L16( )  = 64.0
 ZGOPTNS( )  = _
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 10.0   u
  DIGMOD( )  = digital
.. predefined
    DATE( )  : 1381359544.0
  DSPFVS( )  : 20.0
     L14( )  : 64.0
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z106229/0006


::::::::::::::
trosy_listpars
::::::::::::::
Listpar.py output for dataset
/ld10/davida/utrecht_data_4/3/
Using pulse program from directory
/u4/exp/stan/nmr/lists/pp
[]
 PULPROG( )  : trosyetf3gpsi
      NS( )  = 16
      DS( )  = 32
      RG( )  = 439
f1 ----------------
    SFO1( )  = 900.214233 Mhz
      O1( )  = 4233.0   hz( ~   4.70 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     PL1(not_determined)  = 120.0  dB
      P1(not_determined)  = 8.750  u (apparently at PL1)
      P2(not_determined)  : 17.50  u (apparently at PL1)
     BF1( )  : 900.21 Mhz
f2 ----------------
    SFO2( )  = 226.3664353 Mhz
      O2( )  = 9054.30  hz( ~  40.00 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     P14(not_determined)  = 500.0   u
        SP3(not_determined)  = 120.0  dB
SPNAM3(not_determined)  = Crp80,0.5,20.1  SPOFFS3( )  =
0.000  Hz
     BF2( )  : 226.357381 Mhz
f3 ----------------
    SFO3( )  = 91.22829627 Mhz
      O3( )  = 10009.270 hz( ~ 118.50 ppm in BRUKER
convention; for 13C add 2.8ppm to get to Wishart+Sykes )
     BF3( )  : 91.217487 Mhz
Direct dimension -----------
      SW( )  = 16.022 ppm
      TD( )  = 2048
   DECIM( )  : 1386.66666667
swh =  14423.0 Hz; nprs = 1024; aq =  71.0 m;
true_dwell_time =  69.3 u
Sequence uses mc syntax (aqseq=None)
Fdim_1 ----------------
 NUCDIM1( )  = 15N
  BFDIM1( )  = 91.217487
 FNMODE1( )  = EA
     SW1( )  = 30.0   ppm
     ND0( )  = 2.00014782589
     IN0( )  : 182.70 u
     TD1( )  = 128
swh =   2736.5 Hz; nprs =   64; aq =  23.4 m;
true_dwell_time = 365.4 u
gradients ----------------
     P16(not_determined)  = 1000.0   u
      GPNAM1( )  = SMSQ10.100     GPX1( )  = 0.0
GPY1( )  = 0.0          GPZ1( )  = 30.0
      GPNAM3( )  = SMSQ10.100     GPX3( )  = 0.0
GPY3( )  = 0.0          GPZ3( )  = 45.0
      GPNAM4( )  = SMSQ10.100     GPX4( )  = 0.0
GPY4( )  = 0.0          GPZ4( )  = 50.0
     P19(not_determined)  = 500.0   u
      GPNAM2( )  = SMSQ10.50     GPX2( )  = 0.0
GPY2( )  = 0.0          GPZ2( )  = 80.0
      GPNAM5( )  = SMSQ10.50     GPX5( )  = 0.0
GPY5( )  = 0.0          GPZ5( )  = 16.2
Miscellaneous
.. set by user
    CNST4( )  = 90.0
      D1( )  = 1000.0  m
     D16(not_determined)  = 200.0  u
    INF1( )  = 365.383709829
     P21(not_determined)  = 37.50  u
 ZGOPTNS( )  = -DLABEL_CN
.. rarely changed
  AQ_MOD( )  = DQD
      DE( )  = 6.5    u
  DIGMOD( )  = digital
.. predefined
     D0( )  : 6.00   u
     D11( )  : 30.00  m
     D26( )  : 2.78   m
   DATE( )  : 1350654082.0
  DSPFVS( )  : 20.0
     P22(not_determined)  : 75.0   u
  PROBHD( )  : 5 mm CPTCI 1H-13C/15N/D Z-GRD Z44910/0002
```

```
::::::::::::::
hncacb_pulseprogram
::::::::::::::
# 1
"/opt/topspin3.0/exp/stan/nmr/lists/pp/user/trhncacbetgp2h3
d.ffm"
;trhncacbetgp2h3d
;avance-version (07/08/13)
;TROSY-HNCACB
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(Caliph.,t1) -> F3(N,t2) ->
F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho gradient selection
(t2)
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;with suppression of protonated carbons
;uncompensated version d25=d26
;
;T. Schulte-Herbrueggen & O.W. Sorensen, J. Magn. Reson.
144,
;   123 - 128 (2000)
;A. Eletsky, A. Kienhoefer & K. Pervushin,
;   J. Biomol. NMR 20, 188-180 (2001)
;(M. Salzmann, G. Wider, K. Pervushin, H. Senn & K.
Wuethrich,
;   J. Am. Chem. Soc. 121, 844-848 (1999))


::::::::::::::
hncaco_pulseprogram
::::::::::::::
# 1
"/opt/topspin3.0/exp/stan/nmr/lists/pp/trhncacoetgp2h3d"
;trhncacoetgp2h3d
;avance-version (10/02/12)
;TROSY-HNCACO
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(Ca) -> F2(C=O,t1)
;            -> F2(Ca) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho gradient selection
(t2)
;using real time in t1
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;uncompensated version d25=d26
;(use parameterset TRHNCACOETGP2H3D)
;
;T. Schulte-Herbrueggen & O.W. Sorensen, J. Magn. Reson.
144,
;   123 - 128 (2000)
;A. Eletsky, A. Kienhoefer & K. Pervushin,
;   J. Biomol. NMR 20, 188-180 (2001)
;(M. Salzmann, G. Wider, K. Pervushin, H. Senn & K.
Wuethrich,
;   J. Am. Chem. Soc. 121, 844-848 (1999))


::::::::::::::
hnca_pulseprogram
::::::::::::::
# 1 "/opt/topspin3.0/exp/stan/nmr/lists/pp/trhncagp2h3d2"
;trhncagp2h3d2
;avance-version (10/02/12)
;TROSY-HNCA
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(Ca,t1) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho (t2)
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;with H-1 180degree pulses in t1
;(use parameterset TRHNCAGP2H3D2)
;
;(M. Salzmann, K. Pervushin, G. Wider, H. Senn & K.
Wuethrich, Proc.
;   Natl. Acad. Sci. USA 95, 13585-13590 (1998))
;A. Eletsky, A. Kienhoefer & K. Pervushin,
;   J. Biomol. NMR 20, 188-180 (2001)
```

```
::::::::::::::
hncocacb_pulseprogram
::::::::::::::
# 1
"/opt/topspin3.0/exp/stan/nmr/lists/pp/user/trhncocacbetgp2
h3d.ffm"
;trhncocacbetgp2h3d.ffm
;avance-version (07/08/13)
;TROSY-HNCOCACB
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(C=O) -> F2(Ca -> Cb,t1)
;            -> F2(C=O) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho gradient selection
(t2)
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;with suppression of protonated carbons
;uncompensated version d25=d26
;(use parameterset TRHNCOCACBETGP2H3D)
;
;T. Schulte-Herbrueggen & O.W. Sorensen, J. Magn. Reson.
144,
;   123 - 128 (2000)
;A. Eletsky, A. Kienhoefer & K. Pervushin,
;   J. Biomol. NMR 20, 188-180 (2001)
;(M. Salzmann, G. Wider, K. Pervushin, H. Senn & K.
Wuethrich,
;   J. Am. Chem. Soc. 121, 844-848 (1999))


::::::::::::::
hncoca_pulseprogram
::::::::::::::
# 1 "/opt/topspin3.0/exp/stan/nmr/lists/pp/trhncocagp2h3d"
;trhncocagp2h3d
;avance-version (10/02/12)
;TROSY-HNCOCA
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(C=O) -> F2(Ca,t1)
;            -> F2(C=O) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho (t2)
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;with H-1 180degree pulses in t1
;(use parameterset TRHNCOCAGP2H3D)
;
;M. Salzmann, G. Wider, K. Pervushin, H. Senn & K.
Wuethrich,
;   J. Am. Chem. Soc. 121, 844-848 (1999)
;(A. Eletsky, A. Kienhoefer & K. Pervushin,
;   J. Biomol. NMR 20, 188-180 (2001))


::::::::::::::
hnco_pulseprogram
::::::::::::::
# 1 "/opt/topspin3.0/exp/stan/nmr/lists/pp/trhncogp2h3d"
;trhncogp2h3d
;avance-version (10/02/12)
;TROSY-HNCO
;3D sequence with
;   inverse correlation for triple resonance
;      via TROSY and inept transfer steps
;
;      F1(H) -> F3(N) -> F2(C=O,t1) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho (t2)
;using constant time in t2
;with H-2 decoupling - requires lockswitch unit or BSMS 2H-
TX board
;(use parameterset TRHNCOGP2H3D)
;
;M. Salzmann, K. Pervushin, G. Wider, H. Senn & K.
Wuethrich, Proc.
;   Natl. Acad. Sci. USA 95, 13585-13590 (1998)


::::::::::::::
trosy_pulseprogram
::::::::::::::
# 1 "/opt/topspin3.0/exp/stan/nmr/lists/pp/trosyetf3gpsi"
;trosyetf3gpsi
;avance-version (09/04/17)
;2D H-1/X correlation via TROSY
;   using sensitivity improvement
;phase sensitive using Echo/Antiecho gradient selection
;using f3 - channel
;(use parameterset TROSYETF3GPSI)
;
;M. Czisch & R. Boelens, J. Magn. Reson. 134, 158-160
(1998)
;K. Pervushin, G. Wider & K. Wuethrich, J. Biomol. NMR 12,
;   345-348 (1998)
;A. Meissner, T. Schulte-Herbrueggen, J. Briand & O.W.
Sorensen, Mol. Phys. 96,
;   1137-1142 (1998)
;J. Weigelt, J. Am. Chem. Soc. 120, 10778-10779 (1998)
;M. Rance, J.P. Loria & A.G. Palmer III, J. Magn. Reson.
136, 91-101 (1999)
;G. Zhu, X.M. Kong & K.H. Sze, J. Biomol. NMR 13, 77-81
(1999)


::::::::::::::
t1_pp_single_expt_pulseprogram
::::::::::::::
# 1
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
; mjc_n15_r1_pss_tr_phasecycle based_on(09Mar10):
mjc_n15_r1_pss_tr (adding phasecycling to get decays to
zero.)
; mjc_n15_r1_pss_tr based_on(23Feb10):
mjc_n15_r1_pss_02_2010 (matt edited to have trosy sequence
at end.)
; mjc_n15_r1_pss_02_2010 based_on(23Feb10):
mjp_n15_r1_pss_02_2005_3 (Matt edited to center pulses)
; mjp_n15_r1_pss_02_2005_3 based_on(04May07):
mjp_n15_r1_pss_02_2005_2 (mjp; insertion of delay between
shaped pulses and power level changes for Mag4)
; mjp_n15_r1_pss_02_2005_2 based_on(04Apr07):
mjp_n15_r1_pss_02_2005_1 (;Maya - delays, gradients and
pulses renamed to be consistent, shaped pulse for
Watergate)
; mjp_n15_r1_pss_02_2005_1 based_on(27Mar07):
n15_r1_pss_02_2005 (mjp; change p21 to p28, p22 to p28x0.5,
p31 to p16, p3 to p21, pl13 to pl16, d18 to d1, define
DELTA9)
;N15_R1
;avance-version
;Inclusion of Pre-Saturation block (PSS)
;And compensation for constant heating (DCC)
; Grover Yip and Erik Zuiderweg, 2005

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (07/12/14)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 165 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9 2009/07/02 16:40:43 ber Exp $
# 13
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2" 2

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>


# 31 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.13 2009/07/02 16:40:43 ber Exp $
# 2
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2" 2

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version 00/02/07
;
;$CLASS=HighRes Incl
;$COMMENT=

# 9 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8
define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
# 29 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2

# 48 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

# 58 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;$Id: Delay.incl,v 1.13 2009/07/02 16:40:43 ber Exp $
# 15
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2" 2


"l21=l20-l22" ; l20 is constant
"l0=1"
"in0=inf1/2"
"d0=6u"
;"DELTA=d0*2+2u"
"DELTA4=d14-p16-p12-d16- larger(p1,p21) -10u"
" DELTA5= DELTA4-p21"
# 25
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
define delay DELTA9
"DELTA9=d22-p27" ; 2*d22 time/loop (d22=5ms for 10ms/loop)

;l31 = 1 1D l31=2 2D

;aqseq 321
;d11: disk delay
;d1: relaxation delay
;pl1: proton power
;pl0: watergate power (shaped pulse)
;pl3: nitrogen power
;pl16: nitrogen aq dec power
;pl11: proton CC power (probably same as pl1)
;p27: proton CC pulse length
;p16: gradient pulse length
;p1: proton 90
;p12: shaped pulse for proton watergate
;p20: proton purge for H2O supp
;p21: nitrogen 90
;d14: 1/4JNH
;d4: less than 1/4JNH
;sp1: shaped pulse for water
;l20: Cnst total CC heating (l20=l21+l22)
;l21: Heating compensation Loop (l21=l20-l22)
;l22: T1 mix loop = 2(d22) (total = 2*d22*l22)
;d22: 2*(d22-p27) = proton CC pulse to pulse delay
# 1 "mc_line 54 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_pss
_tr_phasecycle_2 exp. def. part of mc cmd. before ze"
define delay MCWRK
define delay MCREST
define delay d0orig
"d0orig=d0"
define loopcounter l0orig
"l0orig=l0"
define loopcounter t1loop
"t1loop=0"
define loopcounter ph1loop
"ph1loop=0"
define loopcounter ST1CNT
"ST1CNT = td1/(2)"
"MCWRK = 0.142857*d11"
"MCREST = d11 - d11"

    dccorr
# 54
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
# 54
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
1 d11 ze
# 1 "mc_line 54 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_pss
_tr_phasecycle_2 exp. def. of mc cmd. after ze"
# 55
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
# 1 "mc_line 55 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_pss
_tr_phasecycle_2 exp. start label for mc cmd."
2 MCWRK  * 3 do:f3
LBLF1, MCWRK  * 3
LBLST1, MCWRK
  MCREST
# 56
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
3 d11*2 pl1:f1 pl3:f3

4 20u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  20u setnmr3|0

;Duty Cycle Compensation
# 63
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
   20u pl11:f1 ; set p1 CC power for p27 90 degree pulse
21 DELTA9
   (p27*2 ph10):f1
   DELTA9
   lo to 21 times l21


  20u pl1:f1 pl3:f3  ; pre-scan saturation sequence
  p1 ph10
  p16:gp12
  40u
  p1 ph12
  p16:gp13
  40u
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  4u setnmr3^0
# 81
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
  d1        ; recovery delay ~ 1H T1

  20u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  20u setnmr3|0
  (p21 ph10):f3
  p16:gp1*0.3 ; z-spoil
  40u

# 91
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
  (p1 ph10):f1
  d4    ;d4: less than 1/4JNH
  (center(p1*2 ph10):f1 (p21*2 ph10):f3)
  d4
;  (p20 ph10):f1 ;leave this out
  (p1 ph11):f1

  p16:gp1
  40u

  (p21 ph5):f3
   d14  ;d14: 1/4JNH
  (center(p1*2 ph10):f1  (p21*2 ph10):f3)
   d14
  (p21 ph6):f3
# 109
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
  20u pl11:f1 ; set p1 CC power for p27 90 degree pulse
22 DELTA9
   (p27*2 ph10):f1
   DELTA9
   lo to 22 times l22
   20u pl1:f1 ; reset original p1 power


;   (p28 ph21:r):f1   ; water suppression
;   (p28*0.5 ph22:r):f1
;   p16:gp2
;   40u
# 122
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"

 if "l0 %2 == 1"
         {
         (p21 ph1):f3
                  }
  else
         {
         (p21 ph4):f3
                  }
  d0 gron0
```

```
  d0 gron0*-1
  2u groff
;  (p21*2 ph10):f3
;  DELTA
;  (p21*2 ph10):f3
  (p1 ph2):f1
  10u pl0:f1
  (p12:sp1 ph2:r):f1
  4u
  p16:gp3
  d16 pl1:f1
  DELTA4
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  DELTA4
  p16:gp3
  d16 pl0:f1
  (p12:sp1 ph12):f1
  10u pl1:f1
  (center (p1 ph10):f1 (p21 ph11):f3 )
  DELTA4
  p16:gp4
  d16 pl0:f1
  (p12:sp1 ph12:r):f1
  10u pl1:f1
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  10u pl0:f1
  (p12:sp1 ph12:r):f1
  DELTA5
  p16:gp4
  d16 pl1:f1
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  (p21 ph3):f3
  go=2 ph31
# 1 "mc_line 165 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_pss
_tr_phasecycle_2 exp. mc cmd. in line"
  MCWRK  do:f3 wr #0 if #0 zd

  "ph1loop+=1"
      "d0=d0orig + t1loop * in0 "
      "phval2 = (ph1loop % 2) * 180"
      MCWRK ip2 + phval2
      "phval3 = (ph1loop % 2) * 180"
      MCWRK ip3 + phval3
      "l0 = l0orig + ph1loop * 1"

  lo to LBLF1 times 2
  MCWRK

  "t1loop+=1"
      "d0=d0orig + t1loop * in0 "
      "phval2 = (ph1loop % 2) * 180"
      MCWRK ip2 + phval2
      "phval3 = (ph1loop % 2) * 180"
      MCWRK ip3 + phval3
      "l0 = l0orig + ph1loop * 1"

  lo to LBLST1 times ST1CNT
  MCWRK
  "t1loop=0"
  "ph1loop=0"
  MCWRK
# 166
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"

exit




ph10=0
ph11=1
ph12=2
ph13=3
# 178
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
ph1=1 3 2 0
ph2=1
ph3=0
ph4=3 1 2 0
ph31=0 2 3 1 2 0 1 3


# 186
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r1_ps
s_tr_phasecycle_2"
ph5=0 ;2
ph6=1 1 1 1 3 3 3 3
ph7=1

ph21=0
ph22=1
ph25=2
;ph31=0 2 2 0
```

```
::::::::::::::
t2_pp_single_expt_pulseprogram
::::::::::::::
# 1
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
; mjc_n15_r2_cpmg_pss_tr_2 based_on(19Feb10):
mjc_n15_r2_cpmg_pss_tr_1 (matt cliff; increasing phase
cycle.)
; mjc_n15_r2_cpmg_pss_tr_1 based_on(18Feb10):
mjc_n15_r2_cpmg_pss_tr (trying to include echo for d0)
; mjc_n15_r2_cpmg_pss_tr based_on(11Feb10):
mjc_n15_r2_cpmg_pss_02_2010 (added trosy element (from
ah_troesy_15n1hhetnoe))
; mjc_n15_r2_cpmg_pss_02_2010 based_on(11Feb10):
mjp_n15_r2_cpmg_pss_02_2005_6 (matt cliff modified to use
centre terminology;)
; mjp_n15_r2_cpmg_pss_02_2005_6 based_on(22Jun07):
mjp_n15_r2_cpmg_pss_02_2005_3 (mjp; alter CPMG phase cycle
to exclude offset dependence compensation)
; mjp_n15_r2_cpmg_pss_02_2005_3 based_on(04May07):
mjp_n15_r2_cpmg_pss_02_2005_2 (mjp; insertion of delay
between shaped pulses and power level changes for Mag4)
; mjp_n15_r2_cpmg_pss_02_2005_2 based_on(05Apr07):
mjp_n15_r2_cpmg_pss_02_2005_1 (mpw; delays, gradients and
pulses renamed to be consistent, shaped pulse for
Watergate)
; mjp_n15_r2_cpmg_pss_02_2005_1 based_on(27Mar07):
n15_r2_cpmg_pss_02_2005 (mjp; change p21 to p28, p22 to
p28*0.5, p3 to p21, p15 to p30, pl13 to pl16, pl15 to pl23,
define delay DELTA14, DELTA15)
;N15-R2_CPMG  with DCC & PSS
;CPMG phase cycle of [0,0,0,0]
;Inclusion of Pre-Saturation block (PSS)
; And compensation for constant heating (DCC)
; Grover Yip and Erik Zuiderweg, 2005
; The end is a block structure which allows 1D or 2D
acquisition,
; depending on l31

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (07/12/14)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 165 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9 2009/07/02 16:40:43 ber Exp $
# 17
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3" 2

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>


# 31 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.13 2009/07/02 16:40:43 ber Exp $
# 18
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3" 2

# 1 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version 00/02/07
;
;$CLASS=HighRes Incl
;$COMMENT=

# 9 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
# 29 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
```

```
;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2

# 48 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

# 58 "/opt/topspin3.1.7/exp/stan/nmr/lists/pp/Delay.incl"
;$Id: Delay.incl,v 1.13 2009/07/02 16:40:43 ber Exp $
# 19
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3" 2


# 22
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
"in0=inf1/2"
"l14=l16-l15"
;l31 = 1 1D l31=2 2D
" DELTA4= d14 - p16 - d16 - p12 - larger(p1,p21)-10u"
" DELTA5= DELTA4-p21"

define delay DELTA14
"DELTA14=d15-p14-p30"
define delay DELTA15
"DELTA15=d15-p30"
"DELTA=d0*2+p21+2u"
"l0=1"
;aqseq 321
;d0: 6u; initial t1/2
;d11: disk delay
;d1: relaxation/recovery delay
;pl1: proton power
;p1: proton 90
;pl3: nitrogen power
;p21: nitrogen 90
;pl0: watergate power (shaped pulse)
;p11: shaped pulse for proton watergate
;sp0: shaped pulse power
;p123: 15N CPMG power
;p30: 15N 90 CPMG pulse
;pl14: 1H cc suppression power
;p14: 1H 90 cc suppression pulse
;p16: gradient pulse length (1ms)
;d16: gradient recovery delay
;pl16: nitrogen aq dec power
;p20: proton H2O purge
;p28: proton H2O scramble
;d14: 1/4JNH
;d4: less than 1/4JNH
;d15: tcp/2, 8*d15 = time of one CPMG loop
;l14: pre-heating loop counter l14=l16-l15
;l15: CPMG loop counter (8*d15)
;l0: hardwired?
;l16: l15+l14=l16=constant (be sure l16>l15)

# 1 "mc_line 63 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cpm
g_pss_tr_3 exp. def. part of mc cmd. before ze"
define delay MCWRK
define delay MCREST
define delay d0orig
"d0orig=d0"
define loopcounter l0orig
"l0orig=l0"
define loopcounter t1loop
"t1loop=0"
define loopcounter ph1loop
"ph1loop=0"
define loopcounter ST1CNT
"ST1CNT = td1/(2)"
"MCWRK = 0.142857*d11"
"MCREST = d11 - d11"

    dccorr
# 63
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
# 63
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
```

```
1 d11 ze
# 1 "mc_line 63 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cpm
g_pss_tr_3 exp. def. of mc cmd. after ze"
# 64
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
# 1 "mc_line 64 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cpm
g_pss_tr_3 exp. start label for mc cmd."
2 MCWRK  * 3 do:f3
LBLF1, MCWRK  * 3
LBLST1, MCWRK
  MCREST
# 65
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
3 d11*2 do:f3
4 10u do:f3
  10u pl1:f1 pl3:f3
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  4u setnmr3^0

;*******************************
;
; Duty-Cycle Compensation (DCC)
; Pre-HEATING BLOCK

  20u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  20u setnmr3|0
  10u pl23:f3 pl14:f1 ; power levels for CMPG sequence
# 79
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
14 DELTA15
    (p30*2 ph10):f3
    DELTA15
    DELTA15
    (p30*2 ph10):f3
    DELTA14
    (p14*2 ph10):f1
    DELTA14
    (p30*2 ph10):f3
    DELTA15
    DELTA15
    (p30*2 ph10):f3
    DELTA15
    lo to 14 times l14

; END HEATING BLOCK
;**********************

;*****************************
;
; Pre-Scan Saturation (PSS)
# 100
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
  10u pl1:f1 pl3:f3
  p1 ph10
  p16:gp12
  d16
  p1 ph12
  p16:gp13
  d16
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  4u setnmr3^0

;**********************
; Begin Real sequence

  d1 ; recovery delay ~ 1H T1
  20u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  20u setnmr3|0
# 117
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
  (p21 ph10):f3
  p16:gp1*0.3 ;? why 0.3
  d16
  (p1 ph14):f1
  d4 ;d4: less than 1/4JNH
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  d4
;  (p20 ph10):f1 ; water purge: is this necessary?)
  (p1 ph11):f1

  p16:gp1
  d16
  (p21 ph15):f3
  d14 ;d14: 1/4JNH
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  d14

;*************************
; RELAXATION BLOCK

   10u pl23:f3 pl14:f1 ; power levels for CMPG sequence

15 DELTA15
   (p30*2 ph10):f3
   DELTA15
   DELTA15
   (p30*2 ph10):f3
```

```
   DELTA14
   (p14*2 ph10):f1
   DELTA14
   (p30*2 ph10):f3
   DELTA15
   DELTA15
   (p30*2 ph10):f3
   DELTA15
   lo to 15 times l15
# 154
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
   10u pl1:f1 pl3:f3

; END RELAXATION BLOCK
;*************************
;

   (p21 ph6):f3
# 161
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
;**********************
; Water Suppression (leave out?)

;   (p28 ph21:r):f1
;   (p28*0.5 ph22:r):f1
   p16:gp2
   d16

  if "l0 %2 == 1"
          {
          (p21 ph4):f3
                  }
  else
          {
          (p21 ph1):f3
                  }
  d0 gron0
  d0 gron0*-1
  2u groff
  (p21*2 ph10):f3
  DELTA
  (p1 ph2):f1
  10u pl0:f1
  (p12:sp0 ph2:r):f1
  4u
  p16:gp3
  d16 pl1:f1
  DELTA4
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  DELTA4
  p16:gp3
  d16 pl0:f1
  (p12:sp0 ph12):f1
  10u pl1:f1
  (center (p1 ph10):f1 (p21 ph11):f3 )
  DELTA4
  p16:gp4
  d16 pl0:f1
  (p12:sp0 ph12:r):f1
  10u pl1:f1
  (center (p1*2 ph10):f1 (p21*2 ph10):f3)
  10u pl0:f1
  (p12:sp0 ph12:r):f1
  DELTA5
  p16:gp4
  d16 pl1:f1
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  (p21 ph3):f3
  go=2 ph31
# 1 "mc_line 210 file
/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cpm
g_pss_tr_3 exp. mc cmd. in line"
  MCWRK  wr #0 if #0 zd

  "ph1loop+=1"
      "d0=d0orig + t1loop * in0 "
      "phval2 = (ph1loop % 2) * 180"
      MCWRK ip2 + phval2
      "phval3 = (ph1loop % 2) * 180"
      MCWRK ip3 + phval3
      "l0 = l0orig + ph1loop * 1"

  lo to LBLF1 times 2
  MCWRK

  "t1loop+=1"
      "d0=d0orig + t1loop * in0 "
      "phval2 = (ph1loop % 2) * 180"
      MCWRK ip2 + phval2
      "phval3 = (ph1loop % 2) * 180"
      MCWRK ip3 + phval3
      "l0 = l0orig + ph1loop * 1"

  lo to LBLST1 times ST1CNT
  MCWRK
  "t1loop=0"
  "ph1loop=0"
  MCWRK
# 211
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
```

```
exit


ph10=0
ph11=1
ph12=2
ph13=3
# 221
"/opt/topspin3.1.7/exp/stan/nmr/lists/pp/user/mjc_n15_r2_cp
mg_pss_tr_3"
ph1=1 3 2 0
ph2=1
ph3=0
ph4=3 1 2 0
ph31=0 2 3 1 2 0 1 3

ph14=0 ; 0 2 2
ph15= 0 0 0 0 2 2 2 2
ph6= 1

ph21=0
ph22=1
ph25=2
```