The University Of Sheffield.

Structural Studies of NADPH:Protochlorophyllide Oxidoreductase

By Andrew William Proudfoot

Department of Molecular Biology and Biotechnology

A thesis submitted for the degree of Doctor of Philosophy

(July 2011)

Abstract

The light driven enzyme Protochlorophyllide Oxidoreductase (POR) is responsible for catalysing the reduction of the $C_{17} - C_{18}$ double bond of the D ring of protochlorophyllide (Pchlide), in the presence of NADPH, forming chlorophyllide (Chlide). The reduction of Pchlide involves a light-induced hydride transfer reaction from the *pro*-S face of nicotinamide adenine dinucleotide phosphate (NADPH) to the C_{17} position, coupled to the addition of a proton to the C_{18} position forming Chlide. The reaction catalysed by POR is a key step in chlorophyll biosynthesis and is essential in the development of chloroplasts.

To achieve the maximum yield of *Thermosynechococcus* POR, *E. coli* were grown at 37 °C to an optical density of 1.7 before being induced with 0.1 mM IPTG and grown for a further two hours at 25 °C. Purification of the over-expressed protein was done using a nickel affinity chromatography column followed by a cation exchange chromatography column. Purified protein could then be concentrated in an Amicon stirred cell concentrator to a maximum concentration of 0.3 mM.

The NMR conditions were optimised, and experiments conducted at 45 °C with a sample in sodium phosphate buffer pH 5.5, containing 100 mM NaCl, 1 mM DTT and Roche complete protease provided the best quality HSQC spectrum, along with stabilising the protein for an indefinite period of time. Full deuteration of the protein did not yield the spectral improvements expected, and as a result the assignment spectra were acquired at 600 MHz using a ^{13}C ^{15}N labelled sample.

A partial assignment of POR identified that, compared to the rest of the protein, the 33-residue loop insertion unique to POR is significantly more mobile and moves on a timescale similar to that associated with substrate binding and product release. Mutagenesis of the loop, whilst not affecting substrate binding, has profound effects on the enzyme activity, implicating the loop to be involved in some way with enzyme activity.

Incubation of the enzyme with Pchlide and $NADP^+$ results in the aggregation of the enzyme and the formation of a complex, which has been shown by EM and AUC to consist of approximately 10 units. This is analogous to the light harvesting Pchlide POR complex identified to exist in plants and proves that *in vitro* such complexes exist in other organisms other than monocotyledonous plants.

Acknowledgements

I must first start by thanking my supervisor Mike, who has supported me throughout the last four years, from near and far. You realised very early on that the vacant expression on my face meant I did not understand and I appreciate all the explanations of talks and papers, which quite frankly had previously gone over my head! You were willing for me to follow up my ideas and respected my input into the project, which has developed me into the scientist I am today, and for which I am grateful. I must also thank Neil Hunter for providing me with a space to work in his lab, along with all the support which you gave me during the second year of my studies, when Mike was away on sabbatical leave, along with all the subsequent years when I just wouldn't leave you alone; your help has been invaluable. I must also thank Jeremy for his continued help with the NMR, along with words of 'support' and 'encouragement' which you have provided over the years; your help has also been very much appreciated and I am not too sure how far I would have got had it not been for your input! Hopefully now I am finished you can have some peace and quiet to do your own work

I also want to thank Mo, Lizzy, Dave M, Pete and all the members of E12, along with Caz, Rob and all the members of the NMR group who have helped, and put up with me along the way. A special thanks must also go to Jo who has been a very good friend to me over the years, both on a personal and professional basis; I am sure you would have got your PhD completed significantly faster had it not been for me rocking up by your side, distracting you whenever I got a little bored or had some inane story to tell!

I also want to thank all my friends including Kate, Adam, Greg, Tom, Moley and Chris along with all you shooting and squash guys; especially Dave, Sam, Sarah and Juliet who were always available for a chat and a drink whenever it was needed. You all helped me relax and switch off after many a stressful day at work and without you all, I certainly would not have got anywhere near to completing my PhD.

Last but by no means least I want to thank my family for always being there. You have always helped me and kept me going along the correct path. Your assistance has been significantly more apparent over the last few months, but it has always been there and where I am today, as well as what I have achieved along the way, is very much a reflection of the help and support which you have all provided me with over the years. THANK YOU VERY MUCH!

Table of Contents

Table of Figures

Table of Tables

Table of Equations

List of Abbreviations

- **AUC** – analytical ultracentrifugation
- **BOG** – n-*octyl*-β-D-glucopyranoside
- **Chlide** – chlorophyllide
- **CMC** – critical micelle concentration
- **CSA** – chemical shift anisotropy
- **DD** – dipole-dipole
- **DPOR** – light-independent protochlorophyllide oxidoreductase
- **DTT** – dithiothreitol
- **EDTA** - ethylenediaminetetraacetic acid
- **HSQC** – homonuclear single quantum coherence spectrum
- **IMAC** - immobilised metal affinity chromatography
- **IPTG** – isopropyl β-D-1-thiogalactopyranoside
- K_d – dissociation constant
- **LB** – liquid Broth
- **LHPP** – light harvesting Pchlide-POR complex
- **NADPH** - nicotinamide adenine dinucleotide phosphate
- **Pchlide** – protochlorophyllide
- pK_a - the pH at which a titrating group is 50 % ionised
- **POR** – NADPH:Protochlorophyllide Oxidoreductase
- R_1 – rate of longitudinal relaxation
- R_2 – rate of transverse relaxation
- **RED** – reductases, epimerases, dehydrogenases
- **SAM** – *S*-adenosyl-L-methionine
- **SDS PAGE** – Sodium dodecyl sulfate polyacrylamide gel electrophoresis
- T_1 – Longitudinal relaxation time
- T_2 – Transverse relaxation time
- **TROSY** – Transverse relaxation optimised spectroscopy
- τ_c – Correlation time

Amino acids:

A	Alanine	**M**	Methionine
C	Cysteine	**N**	Asparagine
D	Aspartate	**P**	Proline
E	Glutamate	**Q**	Glutamine
F	Phenylalanine	**R**	Arginine
G	Glycine	**S**	Serine
H	Histidine	**T**	Threonine
I	Isoleucine	**V**	Valine
K	Lysine	**W**	Tryptophan
L	Leucine	**Y**	Tyrosine

CHAPTER 1

Introduction

1.1 Oxygenic Photosynthesis

Photosynthesis is the process by which sunlight energy is captured and used to convert carbon dioxide into organic compounds, as well as produce atmospheric oxygen. Photosynthesis is the driving force in the carbon cycle and is responsible for removing carbon dioxide from the atmosphere and converting it into the basic building blocks of life. As a result, it is a process which all life is dependant upon and affects climate conditions around the world. The only group of organisms which do not rely in some way on photosynthesis are the chemolithoautotrophic bacteria which grow in complete darkness and use carbon dioxide as their sole carbon source (Falkowski *et al.* 2000).

Oxygenic photosynthesis can be summarised by the general equation:

$$CO_2 + 2H_2O \xrightarrow{Light} [CH_2O]_n + H_2O + O_2$$

Equation 1.1: The general equation for oxygenic photosynthesis in photoautotrophs (Van Niel 1962).

This equation is significantly simplified, where water, acting as the sole reductant, provides the electrons required to reduce the carbon dioxide resulting in the formation of carbohydrate and the release of oxygen. Equation 1.1 corresponds to two chemically different biological processes; a photochemical step and a non-photochemical step. The photochemical step provides the energy and the transferable hydrogen which is essential for the synthesis of ATP and NAD(P)H, whilst the non-photochemical reaction utilises the produced ATP and NAD(P)H to catalyse a series of carbon fixation reactions. It is these non-photochemical reactions which are responsible for the assimilation of carbon dioxide from and release of oxygen into the atmosphere.

The compound responsible for harvesting sunlight and converting it into photochemical energy for the cell is chlorophyll. Chlorophyll 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, with one third of this biomass being found on land and the other two thirds being found in the ocean. The metabolism of chlorophyll is also the only biological process which can be observed from outer space (Rüdiger 1997) (Figure 1.1).

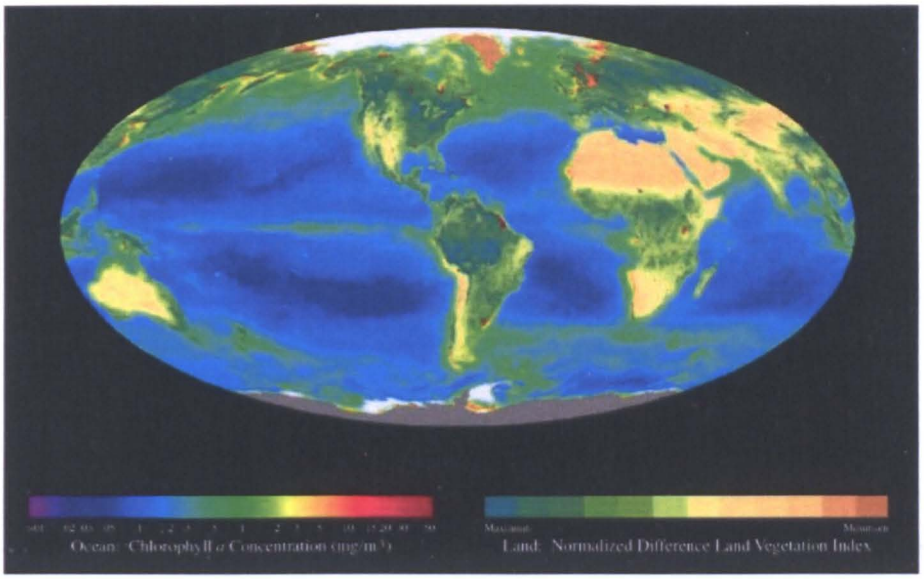


Figure 1.1: View of Chlorophyll biosynthesis from space, showing the hot spots for chlorophyll production around the world (Feldman). On land, regions appearing in green indicate areas of high vegetation whilst regions appearing in brown indicate areas of low vegetation. In the ocean, regions appearing in purple indicate a low concentration of Chlorophyll α, whilst regions appearing in red indicate a high concentration of chlorophyll α.

1.2 Photosynthetic Bacteria

Due to their relatively simple photosystems; simple, rapid and cheap culture; and ease of genetic manipulation; photosynthetic bacteria have been widely used as model systems to study all aspects of photosynthesis. Photosynthetic bacteria are generally found to inhabit marine and freshwater environments and can be classified into four major classes:

Green bacteria: This can be further sub-divided into green sulphur bacteria (*Chlorobi*), which are strict anaerobes that use sulphide or thiosulphide as an electron source, and green non-sulphur bacteria (*chloroflexi*), which are filamentous anoxygenic phototrophs and use reduced carbon compounds as electron donors (Bryant *et al.* 2006).

Purple sulphur bacteria: These are capable of using sulphur or H_2 as electron donors, however very little information is known about these organisms (Bryant *et al.* 2006).

Purple non-sulphur bacteria: These are incapable of using sulphur or H_2 as an electron donor and must use a simple organic compound instead. Several species of purple non-sulphur bacteria have had their genome sequenced, including *Rhodobacter sphaeroides* (Bryant *et al.* 2006).

Cyanobacteria: These are oxygenic phototrophs and contain the well characterised species *Synechocystis*, *Synechococcus* and *Prochlorococcus* (Bryant *et al.* 2006).

The chlorophyll and photosynthetic apparatus found in bacteria is chemically different from the chlorophyll a and b found in plants and has evolved as a result of the organisms demand to harvest light at different optimum wavelengths in different conditions and at different depths. As a result chlorophototrophs can utilise the majority of solar radiation which reaches Earth ranging from near UV to near infrared wavelengths (Chew *et al.* 2007).

1.3 Tetrapyrroles

Cyclic tetrapyrroles consist of four pyrrole rings joined by a single carbon unit, which links position 2 of one pyrrole ring to position 5 of the next. Cyclic tetrapyrroles include compounds such as haem, chlorophyll, bacteriochlorophyll and porphynoids. Porphynoids are generally more reduced and can be further sub-divided into corrinoids (vitamin B_{12}), sirohaem, haem d_1 and coenzyme F_{430}. Whilst the general structure of all cyclic tetrapyrroles is similar, the major difference between them is the central co-ordinating divalent metal ion, which is chelated into the porphyrin ring at various branch points (Figure 1.2).

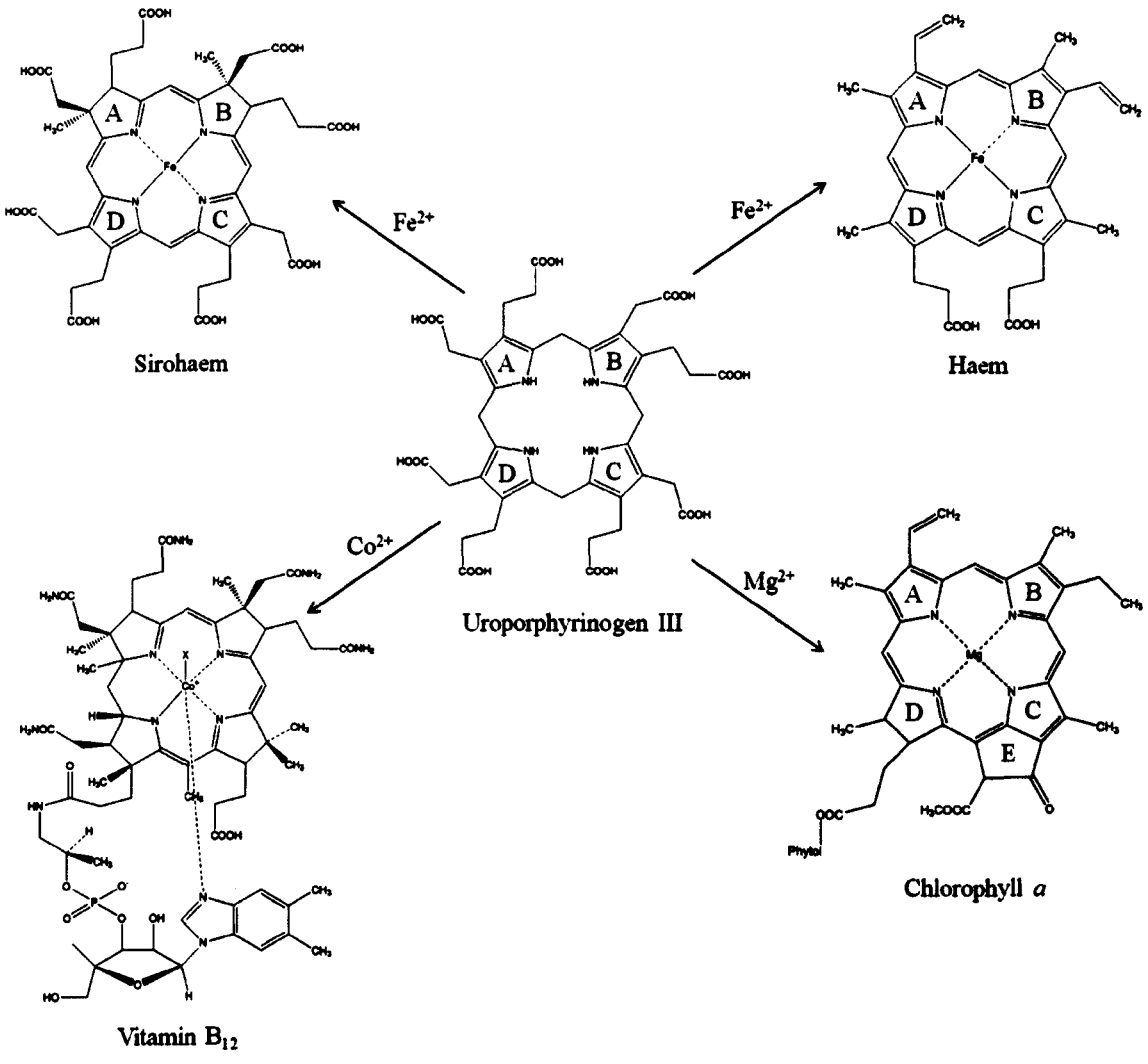


Figure 1.2: Diagram showing the four major tetrapyrroles synthesised in nature from the precursor uroporphyrinogen III.

Haem is a cyclic tetrapyrrole which is co-ordinated by a central ferrous ion (Figure 1.2). In living organisms haem fulfils roles as a prosthetic group in a range of compounds including haemoglobin, myoglobin and sensor proteins for gasses such as O_2 and NO. Haem is also involved with the production and decomposition of hydrogen peroxide as a cofactor of peroxidase and catalase respectively (Layer *et al.* 2010). Sirohaem is also co-ordinated by a central ferrous ion but differs from haem by the chemical groups which are attached to the pyrrole rings (Figure 1.2). Higher plants contain sulphite and nitrite reductases which contain sirohaem as a prosthetic group. In the absence of sirohaem the six electron reduction of sulphite and ntrite would fail to occur and as a consequence there would be no reduced sulphur present in the organism for amino acid or iron sulphur biogenesis (Tripathy *et al.* 2010).

Coenzyme B_{12} is a significantly more reduced tetrapyrrole, which consists of a corin ring that is co-ordinated by a central cobalt ion. The central cobalt ion is also ligated by two other ligands, one is a nitrogen group of a modified base called dimethylbenzimidazole, which is also attached to the corin ring via an aminopropanol linker (Raux *et al.* 2000). Coenzyme B_{12} has roles in the methylations involved with methionine synthesis, along with the reduction of ribonucleotides to deoxyribonucleotides.

(Bacterio)chlorophyll is a tetrapyrrole macrocycle which is co-ordinated by a central magnesium metal ion. Unlike other tetrapyrroles, chlorophyll also contains a characteristic fifth isocyclic ring along with a phytol chain attached to C_{17} (Tanaka *et al.* 2007). Whilst higher plants synthesise only two chlorophyll species, chlorophyll *a* and chlorophyll *b*, which have a methyl and formyl group present at the C_7 position respectively, chlorophotoautotrophs synthesise both bacteriochlorophyll and chlorophyll. To date 12 types of chlorophylls and bacteriochlorophylls have been identified, with the diversity in their structures resulting from enzymes that act late on in the chlorophyll biosynthesis pathway (Chew *et al.* 2007). All photosynthetic organisms produce (bacterio)chlorophyll *a* by the same method, which is then further modified by a series of reactions, resulting in the production of the different forms of (bacterio)chlorophyll (Figure 1.3)

1.4 Common Steps in the Tetrapyrrole Biosynthic Pathway

Tetrapyrrole biosynthesis in all photosynthetic organisms follows a common pathway until it reaches the intermediate protoporphyrin IX. At this stage the biosynthetic pathway branches, with either a Mg^{2+} or Fe^{2+} ion being chelated into the porphyrin ring, resulting in the formation of chlorophyll or haem respectively.

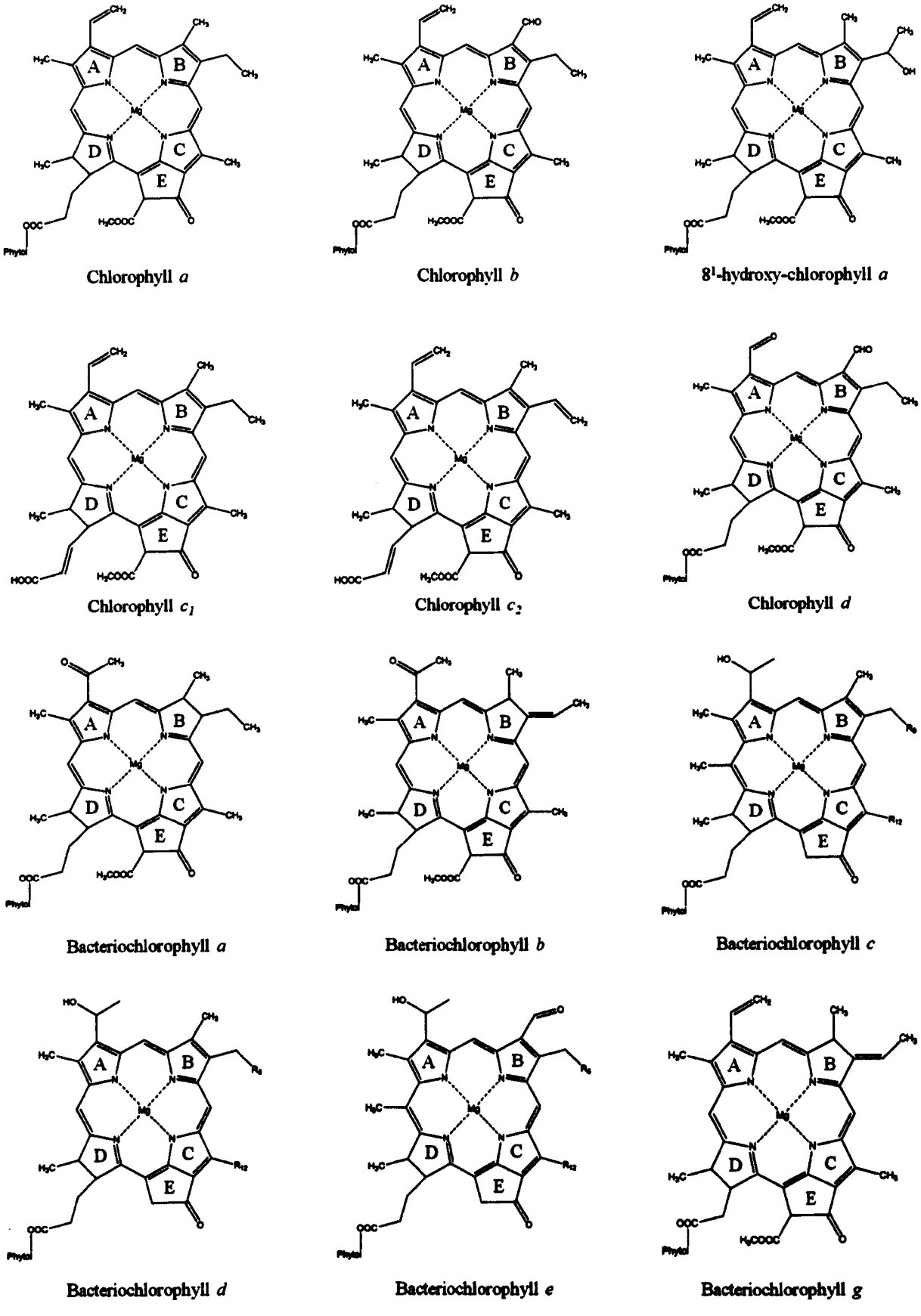


Figure 1.3: Diagram detailing the 12 different forms of chlorophyll and bacteriochlorophyll produced in plants and photosynthetic bacteria.

1.4.1 Formation of δ-aminolevulinic acid

δ-aminolevulinic acid is the common precursor for tetrapyrrole biosynthesis in all organisms and its synthesis is the first committed step in tetrapyrrole biosynthesis. δ-aminolevulinic acid is produced by two distinct pathways in nature: vertebrates, non-photosynthetic eukaryotes and α-proteobacteria produce δ-aminolevulinic acid via the Shemin pathway (Figure 1.4); whilst the C_5 pathway (Figure 1.5) is used by all photosynthetic eukaryotes, archaea and bacteria with the exception of the α-proteobacteria (Gibson *et al.* 1958; Kikuchi *et al.* 1958; Kannangara *et al.* 1988).

In the Shemin pathway, δ-aminolevulinic acid is synthesised by the condensation of succinyl-coenzyme A and glycine which is catalysed by the enzyme δ-aminolevulinate synthase (Kikuchi *et al.* 1958) (Figure 1.4).

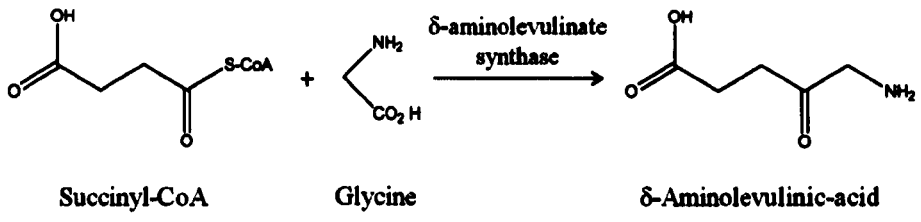


Figure 1.4: The production of δ-aminolevulinic acid from succinyl-CoA and glycine via the Shemin pathway.

Production of δ-aminolevulinic acid by the C_5 pathway is a three step process starting with the precursor glutamate. The initial reactions involve glutamyl tRNA synthase activating glutamate by ligation, forming glutamyl-tRNA, followed by the reduction of the activated carboxyl to a formyl group, by glutamyl tRNA reductase, forming glutamate-1-semialdehyde. The final reaction involves a series of intermolecular amino-exchange reactions, resulting in the trans-amination of glutamate-1-semialdehyde to δ-aminolevulinic acid using the enzyme glutamate-1-semialdehyde aminotransferase (Kannangara *et al.* 1988) (Figure 1.5).

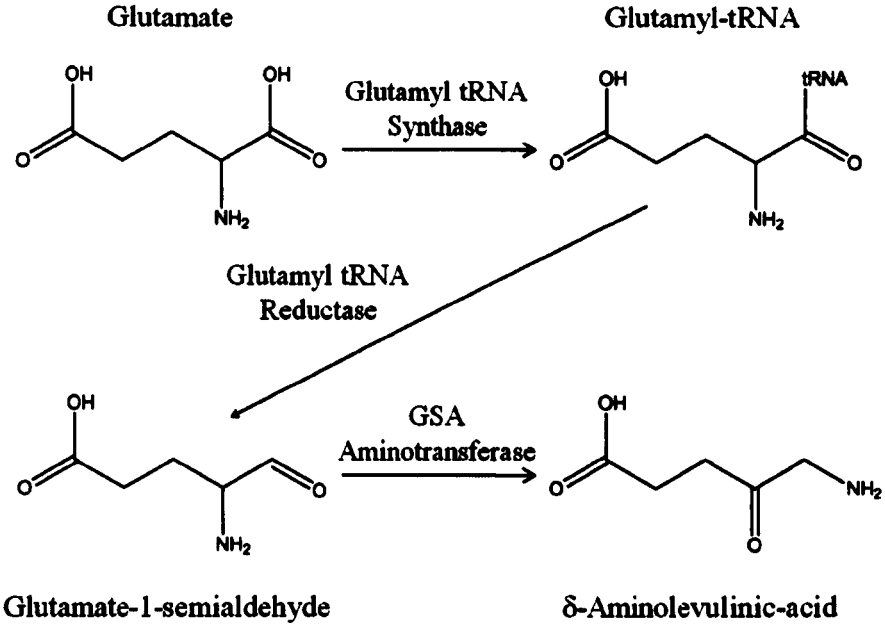


Figure 1.5: The production of δ-aminolevulinic acid from glutamate via the C_5 pathway.

The formation of δ-aminolevulinic acid is the first regulatory point in the chlorophyll biosynthetic pathway which provides a significant advantage, since accumulation of intermediates later on in the pathway will result in the formation of singlet oxygen upon exposure to light. In the C_5 pathway, the enzyme which is regulated is glutamyl tRNA reductase. Failure to repress the activity of glutamyl tRNA reductase in dark grown *Arabidopsis thaliana* resulted in the accumulation of the intermediate protochlorophyllide (Goslings *et al.* 2004). Glutamyl tRNA reductase is encoded by the *HemA* gene family, of which one member that has been found in all higher plants examined to date, is induced by light (Ilag *et al.* 1994; Tanaka *et al.* 1996; Tanaka *et al.* 1997). Regulation of the *HemA* gene in non-photosynthetic organisms has been suggested to occur through end product feedback regulation, rather than the transcriptional regulation observed with *HemA* in photosynthetic organisms (Kumar *et al.* 1996; Tanaka *et al.* 1996; Tanaka *et al.* 1997).

Another class of *A. thaliana* mutants which accumulated protochlorophyllide when grown in the dark were identified to have a mutation in their *flu* gene, a gene which encodes a 27 kDa protein which directly interacts directly with glutamyl tRNA reductase. The presence of an excess of protochlorophyllide or other intermediates in the chlorophyll biosynthetic pathway resulted in the activation of the *flu* gene and the suppression of glutamyl tRNA reductase by FLU.

1.4.2 δ-aminolevulinic acid to Porphobilinogen

The first reaction of the common tetrapyrrole biosynthetic pathway is catalysed by the enzyme δ-aminolevulinic acid dehydratase and involves the asymmetric condensation of two molecules of δ-aminolevulinic acid forming porphobilinogen, with the release of two molecules of water (Granick 1954; Gibson *et al.* 1955; Schmid *et al.* 1955) (Figure 1.6). The protein sequences of δ-aminolevulinic acid dehydratase have been identified from a number of organisms and show a significant level of homology (Jaffe 1995a) indicating that the mechanism of catalysis is probably the same for all organisms (Willows *et al.* 2008).

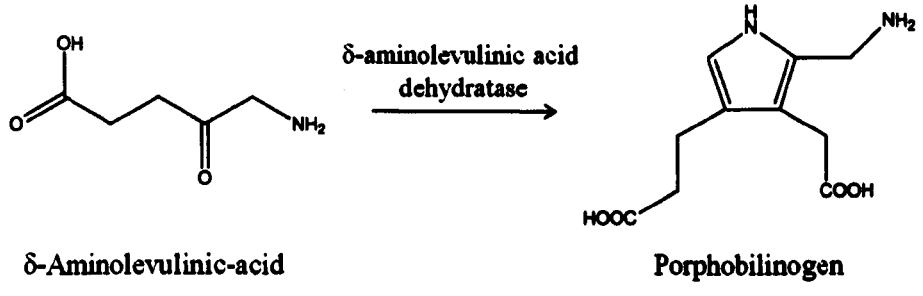


Figure 1.6: Diagram detailing the production of porphobilinogen from δ-aminolevulinic acid by the enzyme δ-aminolevulinic acid dehydratase.

The crystal structure of δ-aminolevulinic acid dehydratase from *E. coli* has been solved to 2.0 Å resolution and has been shown to consist of eight monomers arranged as four dimer pairs forming an octamer (Erskine *et al.* 1999b). Further crystal structures of δ-aminolevulinic acid dehydratase complexed with inhibitors, indicated that both δ-aminolevulinic acid molecules form a Schiff-base between their keto group and the ε-amino group of two conserved lysine residues, which are found in the active site of each monomer (Erskine *et al.* 1999a; Erskine *et al.* 2001; Erskine *et al.* 2005). The two δ-aminolevulinic acid molecules bind to two distinct binding sites of δ-aminolevulinic acid dehydratase, identified as the P and A sites. The molecule bound to the A-site is thought to form the acetic acid side chain and the molecule bound to the P-site is thought to form the propanoic side chain of the porphobilinogen molecule (Jordan *et al.* 1985).

All δ-aminolevulinic acid dehydratase enzymes have been identified to be metalloenzymes with enzymes from different species binding either Mg^{2+} or Zn^{2+}. Whilst the

exact function of the bound metal ions has not been identified, the presence of a tightly bound metal ion has been shown to be essential for catalysis (Jaffe *et al.* 1992). In total, three metal ion binding sites are present and have been proposed to be involved with δ-aminolevulinic acid binding and activity, allosteric activation of the enzyme and proton removal during product formation (Jaffe *et al.* 1994; Jaffe 1995b).

1.4.3 Porphobilinogen to Hydroxymethylbilane

The enzyme hydroxymethylbilane synthase catalyses the formation of a highly unstable linear tetrapyrrole intermediate, hydroxymethylbilane, via the condensation of four molecules of porphobilinogen (Burton *et al.* 1979; Jordan *et al.* 1979a) (Figure 1.7). Hydroxymethylbilane synthase isolated from *Rhodobacter sphaeroides* and *E. gracilis* was used to establish that hydroxymethylbilane is assembled in an ordered way, starting with ring A (the top left pyrrole ring of the cyclic tetrapyrrole) and then continuing in a clockwise manner forming rings B, C and D respectively (Battersby *et al.* 1979b; Jordan *et al.* 1979b).

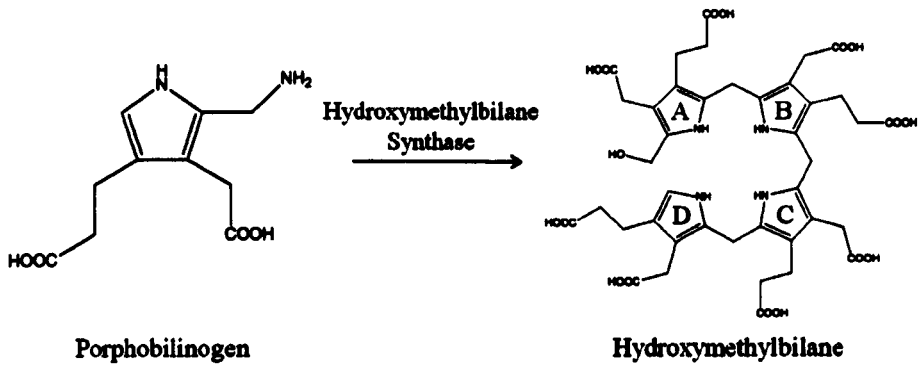


Figure 1.7: Diagram detailing the production of hydroxymethylbilane from porphobilinogen by the enzyme hydroxymethylbilane synthase.

Monopyrrole through to tetrapyrrole enzyme-substrate complexes have been identified to exist and have been documented to be linked through a dipyrromethane cofactor (Jordan *et al.* 1981; Battersby *et al.* 1983b; Jordan *et al.* 1987; Warren *et al.* 1988). In *E. coli,* it has been shown that the dipyrromethane cofactor is post-translationally attached to a cysteine residue of the enzyme, via a thioether linkage. Following formation of hydroxymethylbilane, the binding of another porphobilinogen molecule by

hydroxymethylbilane synthase, results in the release of the formed hydroxymethylbilane and prepares the enzyme to catalyse the addition of porphobilinogen, to the free end of the dipyrromethane (Battersby *et al.* 1983a; Shoolingin-Jordan *et al.* 1996).

Sequence comparison of the *hemC* gene, identified to encode hydroxymethylbilane synthase, shows several conserved regions of amino acids (Avissar *et al.* 1995). The crystal structure of hydroxymethylbilane synthase has been solved at 1.76 Å resolution and has been shown to consist of three distinct domains attached via linkers, which provides the enzyme with a certain level of flexibility as the pyrrole chain elongates (Jordan *et al.* 1987). The structure has also elucidated the position of the active site cleft, which is large enough to contain the growing tetrapyrrole (Louie *et al.* 1992; Louie *et al.* 1996).

1.4.4 Hydroxymethylbilane to Uroporphyrinogen III

Hydroxymethylbilane is cyclised by the enzyme uroporphyrinogen III synthase, forming the first macrocyclic tetrapyrrole, uroporphyrinogen III (Figure 1.8). The catalysed reaction involves an inversion of the D ring and the interchange of C_{α} atoms, via a spiro-pyrrolenine intermediate, resulting in the formation of uroporphyrinogen III (Crockett *et al.* 1991; Stark *et al.* 1993). In the absence of uroporphyrinogen III synthase, the non-enzyme catalysed formation of the biologically inactive isomer uroporphyrinogen I occurs on a rapid time scale (Battersby *et al.* 1979a). This indicates that uroporphyrinogen III synthase must be present immediately after the formation and release of hydroxymethylbilane from hydroxymethylbilane synthase (Burton *et al.* 1979; Jordan *et al.* 1979a). As a result a proposal has been made that hydroxymethylbilane synthase forms a complex with uroporphyrinogen III synthase thus facilitating the transfer of the hydroxymethylbilane and the formation of the correct isomer (Beale 1999). This is supported by previous reports which documented that the presence of uroporphyrinogen III synthase from *R. sphaeroides* facilitates the release of hydroxymethylbilane from hydroxymethylbilane synthase *in vitro* (Rosé *et al.* 1988), however studies using yeast two-hybrid systems have not identified any such interactions *in vivo* (Beale 1999).

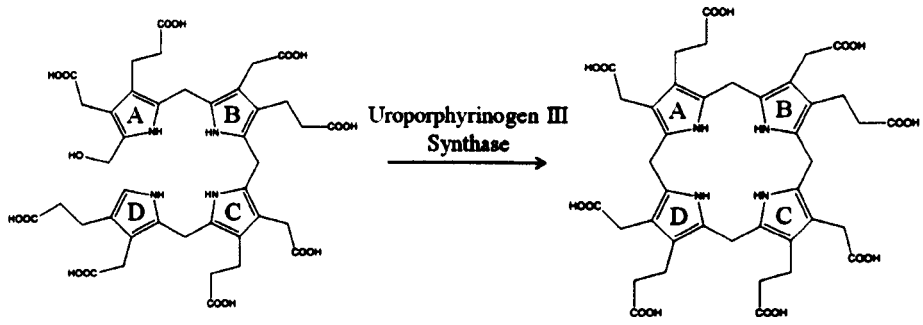


Figure 1.8: Diagram detailing the production of uroporphyrinogen III from hydroxymethylbilane by the enzyme uroporphyrinogen III synthase.

A crystal structure of uroporphyrinogen III synthase, solved at 1.85 Å resolution, has been published and details that the protein folds into a two domain structure connected by a two strand anti-parallel β-ladder, which contains the proposed enzyme active site (Mathews *et al.* 2001).

1.4.5 Uroporphyrinogen III to Coproporphyrinogen III

The conversion of uroporphyrinogen III to coproporphyrinogen III is catalysed by the enzyme uroporphyrinogen III decarboxylase. Starting with the acetyl group present on the D-ring of uroporphyrinogen III and continuing in a clockwise manner, the reaction involves the stepwise decarboxylation of the four acetate residues forming coproporphyrinogen III (Luo *et al.* 1993) (Figure 1.9). It has been shown that at physiological concentrations, the decarboxylation of uroporphyrinogen III occurs in an ordered fashion, as detailed above. However at higher concentrations than that found *in vivo,* the decarboxylations have been identified to occur in a random order (Jones *et al.* 1993). Support for the clockwise decarboxylation theory came from Lash (1991(, who identified that of the four possible type III pentacarboxylate isomers, only one formed catalytically active coproporphyrinogen III. He therefore concluded that uroporphyrinogen III could not be decarboxylated in a random way (Lash 1991).

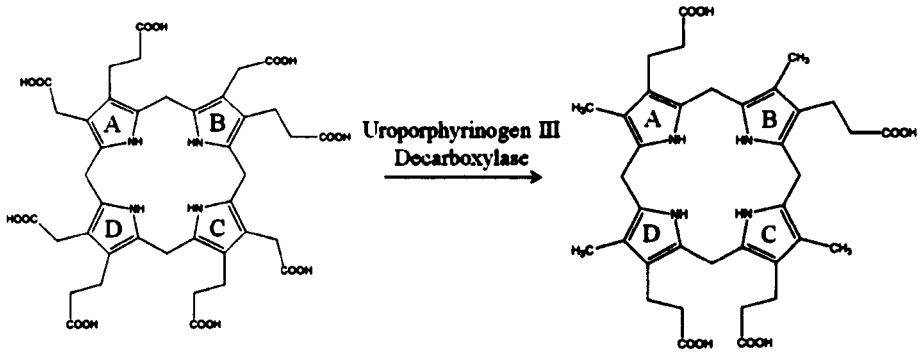


Figure 1.9: Diagram detailing the production of coproporphyrinogen III from uroporphyrinogen III by the enzyme uroporphyrinogen III decarboxylase.

The structures of human and *Nicotiana tabacum* uroporphyrinogen III synthase have been solved at 1.6 Å and 2.3 Å resolution respectively (Whitby *et al.* 1998; Martins *et al.* 2001). The human uroporphyrinogen III synthase has a distorted $(\alpha/\beta)_8$-barrel fold, which contains a deep cleft that appears to form the enzyme active site (Whitby *et al.* 1998). Several highly conserved residues that have been proposed to be involved with substrate binding and catalysis have been located within the deep cleft (Wyckoff *et al.* 1996). Homologs of uroporphyrinogen III synthase have been identified to have a minimum of 33 % similarity and as a result, the two crystallised enzymes appear to have a similar active site, suggesting that the two enzymes catalyse their reactions in a similar way. Since only one active site is present the four decarboxylation events have been suggested to occur at the same catalytic site, with the reaction intermediates being stabilised by a conserved aspartate residue, which coordinates the central pyrrole NH groups (Phillips *et al.* 2003).

1.4.5.1 The Sirohaem Branch

The formation of sirohaem from uroporphyrinogen III represents the first branch point in the tetrapyrrole biosynthesis pathway. In higher plants, methylation of the A and B rings of uroporphyrinogen III by the enzyme uroporphyrinogen III methyltransferase results in the formation of the intermediate dihydrosirohydrochlorin (Tripathy *et al.* 2010). Activity of uroporphyrinogen III methyltransferase is dependent upon an *S*-adenosyl-L-methionine cofactor to act as a methyl donor during the reaction. Following its methylation, dihydrosirohydrochlorin is dehydrogenated to form sirohydrochlorin by the enzyme

dihydrosirohydrochlorin dehydrogenase. The enzyme sirohydrochlorin ferrochelatase is then responsible for inserting a Fe^{2+} into the sirohydrochlorin molecule, forming sirohaem (Sakakibara *et al.* 1996; Leustek *et al.* 1997; Raux-Deery *et al.* 2005). This final reaction is very similar to that of the reaction catalysed by the enzyme protoporphyrin IX ferrochelatase in the haem biosynthesis pathway, and structural studies have shown that the two enzymes have a similar structure (Figure 1.10).

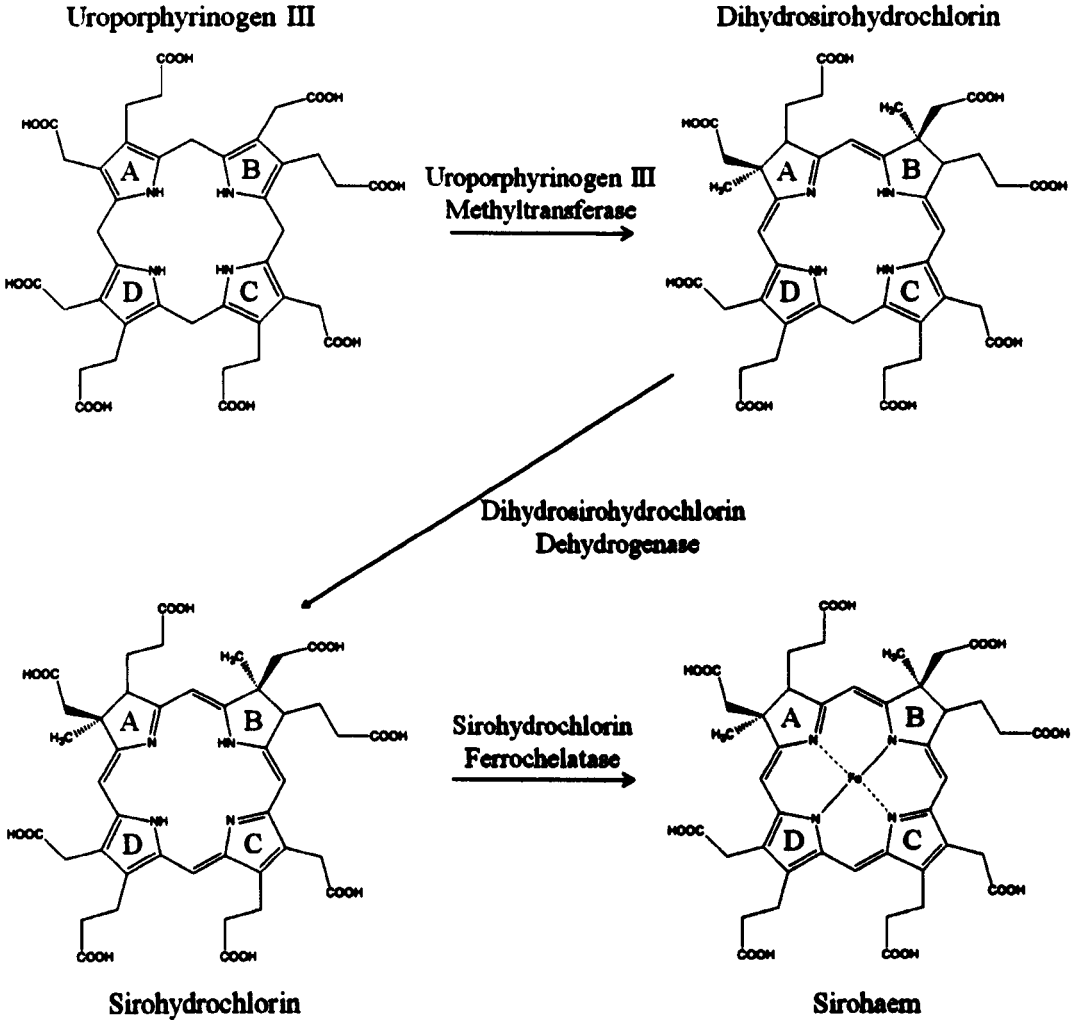


Figure 1.10: Diagram detailing the formation of sirohaem from the precursor uroporphyrinogen III, via the intermediates dihydrosirohydrochlorin and sirohydrochlorin.

Whilst it is generally accepted that the formation of the intermediates described above occurs in all sirohaem biosynthesis pathways, there appears to be little or no conservation of the enzymes involved across different species. In yeast, the methylation of rings A and B is

catalysed by the enzyme Met1p, which exhibits very little homology to the plant uroporphyrinogen III methyltransferase. Likewise, whilst the final two steps of the sirohaem pathway are catalysed by two independent enzymes in higher plants, yeast uses a bi-functional enzyme with a single active site called Met8p, to catalyse the dehydrogenation and chelation reactions (Raux *et al.* 1999). Unlike all other sirohaem producing organisms, *E. coli* and *S. enterica serovar typhimurium* appear to only have a single homodimeric enzyme, termed CysG, which is responsible for catalysing the whole reaction pathway (Warren *et al.* 1990a; Warren *et al.* 1990b; Warren *et al.* 1994).

1.4.6 Coproporphyrinogen III to Protoporphyrinogen IX

Protoporphyrinogen IX is formed from the oxidative decarboxylation of the propionate groups present on rings A and B of coproporphyrinogen III, by the enzyme coproporphyrinogen III oxidase (Figure 1.11). Two forms of the enzyme have been identified; an oxygenic form, coproporphyrinogen III oxidase, which uses oxygen as a terminal electron acceptor and an enzyme which is active under anaerobic conditions, coproporphyrinogen III dehydrogenase, which uses a terminal electron acceptor other than oxygen (Tait 1969; Tait 1972; Breckau *et al.* 2003). Although the conversion of coproporphyrinogen III to protoporphyrinogen IX is catalysed by two different enzymes in different organisms, the stereochemistry has been shown to be the same in both reactions (Battersby *et al.* 1975; Rand *et al.* 2010).

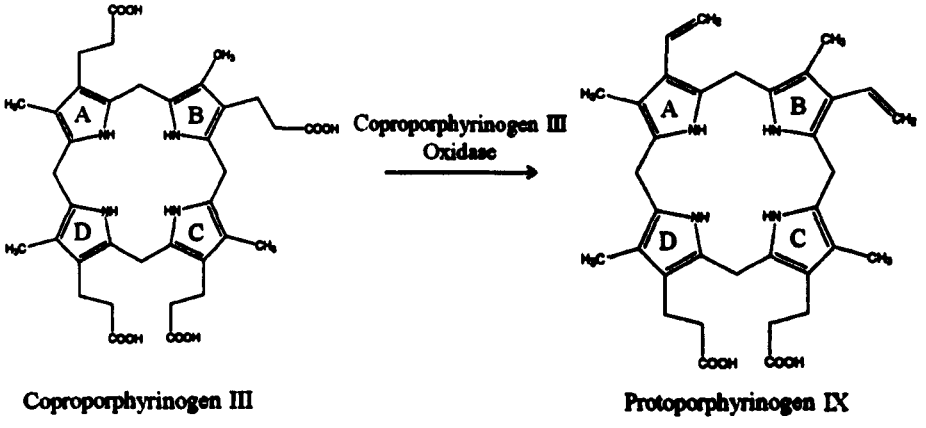


Figure 1.11: Diagram detailing the formation of protoporphyrinogen IX from the precursor coproporphyrinogen III by the enzyme coproporphyrinogen III oxidase.

Coproporphyrinogen III oxidase from *E. gracilis* has been shown to have a substrate specificity for 3-vinyl porphyrinogen (ring A) over 4-vinyl porphyrinogen (ring B), suggesting that decarboxylation of coproporphyrinogen III occurs in an ordered way via the intermediate 2-vinylcoproporphyrin III (Cavaleiro *et al.* 1974; Elder *et al.* 1978). Experiments conducted using deuterium and tritium labelled substrates have indicated that the removal of protons as hydride ions, from the β-carbons of the propionate group, occurs during the reaction. This results in the conversion of the propionate side chain into a vinyl side chain by anti-periplanar elimination, whilst the α-protons of the propionate groups do not appear to be involved (Battersby *et al.* 1975; Zaman *et al.* 1976; Seehra *et al.* 1983).

The crystal structures of coproporphyrinogen III oxidase from *Saccharomyces cerevisiae* and humans have been solved to a resolution of 2.0 and 1.58 Å respectively (Phillips *et al.* 2004; Lee *et al.* 2005). Both crystal structures were similar in their quaternary structure and confirmed that coproporphyrinogen III oxidase was present as a dimer, with each monomer consisting of a seven-stranded anti-parallel β-sheet surrounded on both sides by α-helices. Whilst the detailed catalytic mechanism of the reaction is unknown, two highly conserved arginine residues have been suggested to interact with the propionate side chains and a conserved aspartate residue has been proposed to co-ordinate the pyrrole NH groups (Phillips *et al.* 2004; Lee *et al.* 2005).

Coproporphyrinogen III oxidase generally manifests in eukaryotes with only a few documented cases where it is present in bacteria (Cavallaro *et al.* 2008); the remaining bacteria utilise coproporphyrinogen III dehydrogenase to carry out the reaction of coproporphyrinogen III to protoporphyrinogen IX. The crystal structure of *E. coli* coproporphyrinogen III dehydrogenase was solved to a resolution of 2.1 Å and was shown to be a monomeric two domain protein, which contained an oxygen-labile [4Fe-4S] cluster (Layer *et al.* 2003; Layer *et al.* 2004). The N-terminal domain is made up of a 12-stranded parallel β-sheet surrounded by α-helices, whilst the C-terminal domain is made up of a three stranded anti-parallel β-sheet surrounded by four α-helices and has been identified to contain the [4Fe-4S] cluster. Three of the iron atoms are co-ordinated by three highly conserved cysteine residues, whilst the fourth is ligated by an *S*-adenosyl-L-methionine (SAM) molecule, which also acts as a cofactor by initiating radical based catalysis (Layer *et al.* 2003).

1.4.7 Protoporphyrinogen IX to protoporphyrin IX

The enzyme protoporphyrinogen oxidase is responsible for catalysing the six electron oxidation of protoporphyrinogen IX to protoporphyrin IX (Figure 1.12). Similar to coproporphyrinogen III oxidase, two distinct unrelated forms of protoporphyrinogen oxidase have been identified; an oxygen dependant form, which uses oxygen as the terminal electron acceptor and an oxygen-independent form, which uses nitrate or fumarate as the terminal electron acceptor. Studies have shown however that oxidation can occur by itself in the absence of an enzyme (Porra *et al.* 1961; Sano *et al.* 1961; Porra *et al.* 1964; Jacobs *et al.* 1975; Jacobs *et al.* 1976).

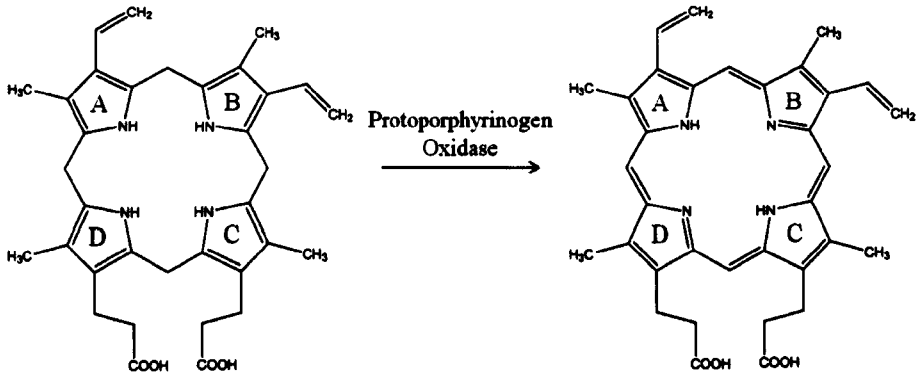


Figure 1.12: Diagram detailing the formation of protoporphyrin IX from the precursor protoporphyrinogen IX by the enzyme protoporphyrinogen oxidase.

Oxygen dependant protoporphyrinogen oxidases have been reported to be homodimeric membrane associated proteins, with the exception of *B. subtilis* and *A. aeolicus* where protoporphyrinogen oxidase is present as a monomer (Wang *et al.* 2001; Qin *et al.* 2010). Whilst all other enzymes, to this point, in the chlorophyll biosynthetic pathway appear only in the plastids, identical protoporphyrinogen oxidases have been reported to be present in both the plastid and mitochondrial fractions of etiolated barley leaves (Jacobs *et al.* 1987). Oxygen dependent protoporphyrinogen oxidases contain a non-covalently bound FAD as a cofactor, which is responsible for the transfer of electrons from the substrate to the oxygen electron acceptor (Siepker *et al.* 1987; Camadro *et al.* 1994). Three molecules of O_2 are reduced throughout the reaction, resulting in the formation of three molecules of H_2O_2 (Ferreira *et al.* 1988). Experiments conducted with tritium-labelled substrates indicated that

three meso protons were removed successively from the α, γ and δ positions of the same face of the molecule, whilst the final tautomerisation is done so by the removal of a hydrogen atom, from the β-meso position of the opposite face of the tetrapyrrole and results in the formation of protoporphyrin IX (Akhtar 2003).

Crystal structures of mitochondrial tobacco and *M. xanthus* protoporphyrinogen oxidase have been solved as dimers to 2.9 and 2.7 Å resolution respectively (Koch *et al.* 2004; Corradi *et al.* 2006), whilst the structure of monomeric protoporphyrinogen oxidase from *B. subtilis* has been solved to 2.9 Å resolution (Qin *et al.* 2010). Although the overall amino acid conservation is low between protoporphyrinogen oxidases from different species, the overall fold of the protein is highly conserved. Each protoporphyrinogen oxidase monomer consists of three domains; an FAD-binding domain and a substrate binding domain, which both exhibit a *p*-hydroxybenzoate-hydroxylase fold-like topology and a third domain involved in dimerisation, which is exclusively α-helical and responsible for membrane-binding for the homo-dimeric enzymes.

Oxygen-independent protoporphyrinogen oxidase has been identified to belong to a family of long chain flavodoxins and contains a FMN cofactor, but whilst the structure of the oxygen-independent protoporphyrinogen oxidase has been solved to a resolution of 2.7 Å and 2.3 Å in the absence and presence of the plant herbicide acifluorfen respectively, significantly less is known about the mechanism of action or the reaction. Jacobs and Jacobs showed that the activity of oxygen independent protoporphyrinogen oxidase isolated from *E. coli* was associated with the membrane fraction and was coupled to the respiratory chain (Jacobs *et al.* 1975; Jacobs *et al.* 1976; Siepker *et al.* 1987). Recent studies have shown that oxygen independent protoporphyrinogen oxidase is active in the presence of menadione, which has been proposed to substitute for the physiological membrane-bound electron acceptor, menaquinone.

1.5 The Branch Point between Chlorophyll and Haem Biosynthesis

The branch point between haem and chlorophyll biosynthesis is the first major branch point in the tetrapyrrole biosynthetic pathway. Insertion of Fe^{2+} into protoporphyrin IX by the enzyme ferrochelatase leads to the formation of haem, whilst insertion of Mg^{2+} by the enzyme Mg-chelatase results in the formation of chlorophyll. In plants the following

reactions are detailed to occur in the chloroplast, whilst in bacteria it is currently unknown whether there is any preferred co-location of the enzymes (Tanaka *et al.* 2007).

1.5.1 The Haem Branch

The final step of haem biosynthesis, the chelation of a ferrous ion into protoporphyrin IX forming protohaem, is catalysed by the enzyme ferrochelatase (Figure 1.13) (Porra *et al.* 1963). The insertion of Fe into protoporphyrin is energetically favourable but will not occur spontaneously and as a result there is no requirement for ATP and studies have shown that ATP actually inhibits the reaction catalysed by ferrochelatase (Cornah *et al.* 2002).

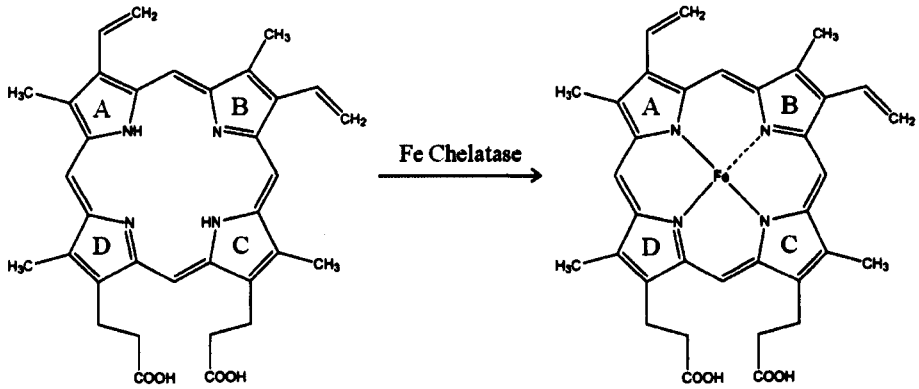


Figure 1.13: Diagram detailing the formation of protohaem from the precursor protoporphyrin IX by the enzyme ferrochelatase.

Ferrochelatase has been crystallised from *B. subtilis* and humans and has been shown to be a membrane-associated homodimer in eukaryotes, but monomeric in bacteria (Wang *et al.* 2001; Wu *et al.* 2001; Grzybowska *et al.* 2002). It has also been reported that in Gram positive bacteria ferrochelatase is a soluble protein located in the cytoplasm (Hansson *et al.* 1994). The crystal structures also identify that each subunit of eukaryotic ferrochelatase contains one [2Fe-2S] cluster whilst there are no iron-sulphur clusters present in bacterial or plant ferrochelatases (Dailey *et al.* 1994; Day *et al.* 1998; Dailey *et al.* 2002; Shepherd *et al.* 2006). The role of the iron-sulphur clusters present in eukaryotic ferrochelatase enzymes is currently unknown; however it has been proposed that the iron-sulphur cluster is indirectly

involved with the interactions between monomers and may therefore explain why ferrochelatase is monomeric in bacteria (Layer *et al.* 2010).

Comparison of amino acid sequences from homologs of ferrochelatase has shown that the level of amino acid conservation is low, yet a high degree of structural conservation is observed in the crystal structures. Each ferrochelatase monomer consists of a four-stranded parallel β-sheet surrounded by α-helices. The N-terminal region is believed to form part of the active site pocket, which binds both the porphyrin ring and metal-binding sites, whilst the C-terminal region is involved with the co-ordination of the iron-sulphur cluster and stabilisation of the homodimer (Al-Karadaghi *et al.* 1997; Medlock *et al.* 2007).

The crystal structure of human ferrochelatase in the presence of protoporphyrin IX has shown that upon binding of the porphyrin ring, the active site closes, resulting in the substrate being completely engulfed within the enzyme. Once bound, the propionate carboxylate groups present on rings C and D of the protoporphyrin molecule, are involved in hydrogen bonding with serine and tyrosine residues in the active site. This causes the porphyrin ring to become distorted by 11.5 ° forming a "modest saddle conformation" (Sigfridsson *et al.* 2003). Following the insertion of the metal ion, conformational changes are observed within the active site, resulting in the propionate hydrogen bonds being broken, allowing the release of protohaem.

1.6 Protoporphyrin IX to Mg Protoporphyrin IX

The insertion of the central magnesium ion into protoporphyrin IX is catalysed by the enzyme Mg-chelatase and represents the first committed step of (bacterio)chlorophyll biosynthesis (Walker *et al.* 1997; Reid *et al.* 2002; Willows 2003) (Figure 1.14). Whilst the formation of Mg-protoporphyrin IX is energetically unfavourable, due to the requirement of removing the shell of co-ordinating water molecules surrounding the Mg^{2+}, the reaction will proceed upon the addition of 15 molecules of ATP (Walker *et al.* 1997). As a result Mg-chelatase has been classified as a type 1, ATP-dependent heterotrimeric chelatase (Brindley *et al.* 2003; Al-Karadaghi *et al.* 2006).

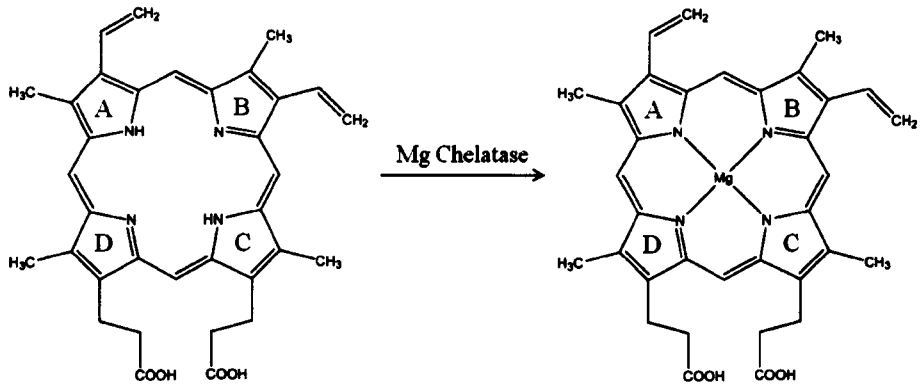


Figure 1.14: Diagram detailing the formation of Mg-protoporphyrin IX from the precursor protoporphyrin IX by the enzyme magnesium chelatase.

The insertion of Mg into protoporphyrin IX forming Mg-protoporphyrin IX, is analogous to the insertion of iron into haem and cobalt into vitamin B_{12}, however the structures of the three enzymes exhibit no similarity. Mg-chelatase consists of three subunits designated I, D and H (Walker *et al.* 1997), which are encoded by the genes *chl/bch I, chl/bch D* and *chl/bch H* respectively, all of which need to be present in order to confer activity (Gibson *et al.* 1995; Jensen *et al.* 1996).

Subsequent studies have shown that subunits D and I are the smaller of the three subunits, with molecular weights of 70 and 40 kDa respectively, and contain several ATP binding sites per subunit (Jensen *et al.* 1999). As a result, subunits D and I have been placed in the ATPase associated with a variety of cellular activities (AAA) family, which are known to form nucleotide dependent ring structures (Willows *et al.* 2008). When analysed by single particle electron microscopy, BchI subunits were identified to form Mg-ATP dependent hexameric rings (Willows *et al.* 2004). This is also the same for the D subunit, however the formation of hexamers of D is dependent upon the presence of subunit I to maintain stability (Lake *et al.* 2004; Axelsson *et al.* 2006). The formation of an I-D complex inhibits the ATPase activity of subunit I and induces binding of the 140 kDa, porphyrin binding H subunit (Willows *et al.* 1996; Jensen *et al.* 1998; Willows *et al.* 1998). The binding of the H subunit to the I-D complex stimulates ATPase hydrolysis and results in chelation of the Mg ion (Jensen *et al.* 1998; Fodje *et al.* 2001). Following chelation, the Mg-chelatase complex dissociates into an I-D-MgADP and a H-MgP complex, which can be recharged with ATP and protoporphyrin IX respectively (Jensen *et al.* 1998).

Prior to chelation a lag period is observed; pre-incubation of subunits I and D with Mg-ATP prior to the addition of subunit H significantly reduced this lag period. This led to the proposal of a two-step reaction: an initial enzyme activation step, followed by a Mg insertion step (Walker *et al.* 1994; Jensen *et al.* 1998; Reid *et al.* 2003). Pre-incubation of the H subunit and the I-D complex with the thiol-modifying reagent N-ethylmaleimide results in enzyme inactivation. This has been documented to be a result of the inhibitor binding to the I subunit, inhibiting ATPase activity (Fuesler *et al.* 1984; Walker *et al.* 1991a; Walker *et al.* 1991b; Jensen *et al.* 2000).

Whilst Mg-ATP is essential for Mg-chelatase activity, it has been identified *in vitro* that the free ATP:ADP ratio participates in regulating Mg-chelatase activity (Reid *et al.* 2004). Upon moving spinach chloroplasts from the dark to light, the concentration of free Mg was observed to increase from 0.5 mM to 2 mM (Ishijima *et al.* 2003). This is lower than the concentration of 6 mM, which had previously been detailed to be needed for Mg-chelatase activity however, the presence of the genome uncoupled protein GUN4 has been documented to lower the concentration of Mg required for activation, down to 2 mM. GUN4 has been proposed to be involved in intracellular signalling and has been identified to bind to both protoporphyrin IX and Mg-protoporphyrin IX, thus stimulating the activity of Mg-chelatase (Davison *et al.* 2005; Verdecia *et al.* 2005).

1.7 Mg Protoporphyrin IX to Mg Protoporphyrinogen IX Monomethyl ester

The enzyme *S*-adenosyl-*L*-methionine:Mg-protoporphyrin IX methyltransferase catalyses the transfer of a methyl group, from the cofactor *S*-adenosyl-*L*-methionine (SAM), to the carboxyl group of the propionate side chain found on ring C of Mg-protoporphyrin IX, forming Mg-protoporphyrin IX monomethyl ester (Gorchein *et al.* 1993) (Figure 1.15). Mg-protoporphyrin IX methyltransferase belongs to the broad family of SAM dependent methyltransferases and the presence of an *S*-adenosyl-*L*-methionine derived methyl group, on the propionate side chain, was confirmed by the use of ^{14}C-methyl labelled SAM (Bollivar *et al.* 1994a; Kagan *et al.* 1994). The addition of a methyl group to the propionate side chain is thought to be essential, to prevent the spontaneous decarboxylation of the propionate group during the formation of the isocyclic ring E (Beale 1999).

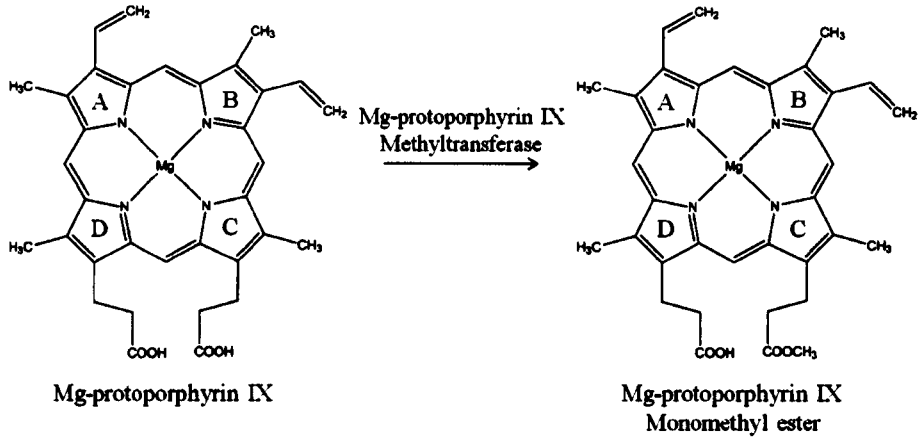


Figure 1.15: Diagram detailing the formation of Mg-protoporphyrin IX monomethyl ester from the precursor Mg-protoporphyrin IX by the enzyme magnesium protoporphyrin IX methyltransferase.

The product of the *bchM* gene from *Rhodobacter sphaeroides* has a molecular weight of 27.5 kDa and has been identified as Mg-protoporphyrin IX methyltransferase, however to date no structure has been published. The deduced amino acid sequences of the *BchM* and *ChlM* genes have a relatively low sequence similarity, however both sequences do contain putative SAM-binding motifs (Bollivar *et al.* 1994a; Gibson *et al.* 1994; Smith *et al.* 1996). Within the broad family of SAM dependent methyltransferases, Shepherd *et al.* placed Mg-protoporphyrin IX methyltransferase in the small molecule class of methyltransferases (Shepherd *et al.* 2003). All small-molecule SAM dependent methyltransferases contain a mixed seven-stranded β-sheet catalytic domain, along with a SAM binding site and a binding site for the methyl acceptor. In other catalytically defined methyltransferases, the SAM methyl group is bound to a charged sulphur atom which converts the inert methyl group into a highly reactive moiety for nucleophilic and carbanionic attack. The methyl group replaces a proton on the methyl acceptor and the methyltransferase has been proposed to stabilise the transition state, by bridging the exchanged proton to an amino acid functional group (Shepherd *et al.* 2003).

Whilst the methyl donor was detected to be SAM, a range of porphyrin molecules including Mg-protoporphyrin IX, Zn-protoporphyrin and Ca-protoporphyrin have been identified to act as methyl acceptors for the mono-methyltransferase (Gibson *et al.* 1963). Reports have also been made of the enhancement of Mg-protoporphyrin IX methyltransferase activity upon the addition of Mg-chelatase subunit, BchH (Hinchigeri *et al.* 1997). This

observation led to the proposal that Mg-protoporphyrin IX may be channelled directly from Mg-chelatase, to Mg-protoporphyrin IX methyltransferase (Masuda 2008).

1.8 Formation of the Isocyclic Ring

The enzyme Mg-protoporphyrin IX monomethyl ester cyclase is responsible for the incorporation of molecular oxygen into the intermediate Mg-protoporphyrin monomethyl ester, forming divinyl protochlorophyllide *a* (Suzuki *et al.* 1997) (Figure 1.16). The incorporation of oxygen oxidises the C_{13} propionic acid side chain of Mg-protoporphyrin IX monomethyl ester, resulting not only in the formation of an isocyclic fifth ring (ring E) but also a change in colour from red to green (Beale 1999; Willows 2003). Two types of the cyclase have been identified, an aerobic cyclase which incorporates molecular oxygen (Walker *et al.* 1989) and an anaerobic cyclase which incorporates atomic oxygen from water (Bollivar *et al.* 1994b).

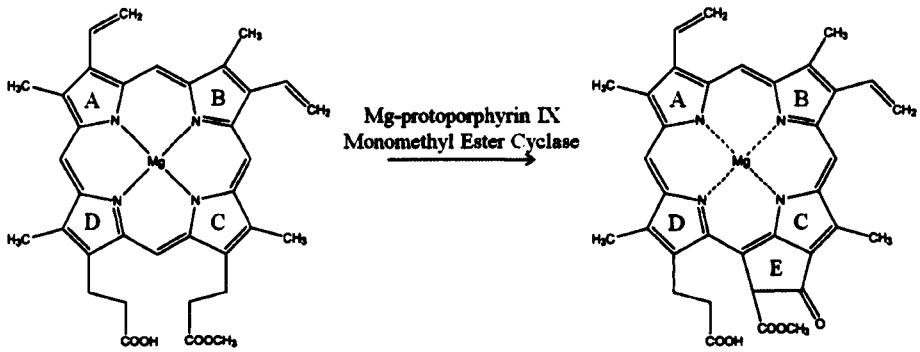


Figure 1.16: Diagram detailing the formation of divinyl protochlorophyllide a from the precursor Mg-protoporphyrin IX monomethyl ester by the enzyme Mg-protoporphyrin IX monomethyl ester cyclase.

The origin of the oxygen used in the oxidation of the C_{13} propionic acid side chain has been studied in purple bacteria using $^{18}O_2$ and $H_2{}^{18}O$. Whilst in *Rhodobacter sphaeroides* the oxo group present on ring E is derived only from water, in *Roseobacter denitrifi* and *Rhodovulum sulphidophilum* which can grow both aerobically and anaerobically, ^{18}O appears to be incorporated from both oxygen and water (Porra *et al.* 1995; Porra *et al.* 1996; Porra *et*

al. 1998). When grown under aerobic conditions, atmospheric oxygen is preferentially incorporated into ring E, whilst when grown anaerobically, oxygen from water is preferentially incorporated. This led to the proposal that two different enzymes may exist for the oxidative cyclization (Porra *et al.* 1996; Porra *et al.* 1998).

Two genes have been discovered, which when disrupted result in the accumulation of the intermediate, Mg-protoporphyrin IX monomethyl ester. These genes have been named *bchE* and *acsF* and are thought to encode the anaerobic and aerobic cyclase respectively (Bollivar *et al.* 1994b; Pinta *et al.* 2002). Whilst structures of either of the proteins encoded by the cyclase genes are missing, analysis of the *bchE* gene has led to a number of structural predictions being made. *bchE* contains a conserved CXXXCXXC motif which forms a 4Fe-4S cluster and is found in all free radical SAM enzymes (Sofia *et al.* 2001). A putative cobalamin binding site has also been identified by sequence analysis of the gene and together with the previous observation, has been used to predict the mechanism by which the formation of the isocyclic ring occurs (Gough *et al.* 2000).

The formation of the isocyclic ring in anaerobically grown bacteria has been proposed to occur in a similar way to that of the β-oxidation of fatty acids, via acrylate, β-hydroxy and β-keto intermediates (Beale 1999). The proposed mechanism involves the use of a dehydrogenase to remove two hydrogen atoms producing the acrylate intermediate, which is then subsequently hydrated forming the β-hydroxy intermediate (Beale 1999). It has also been proposed that the 4Fe-4S cluster acts as the electron donor, to generate a 5'-deoxyadenosyl radical from adenosyl cobalamin. This can then act as proton radical acceptor, thus generating a radical on the propionate side chain (Gough *et al.* 2000; Sofia *et al.* 2001).

Organisms which grow aerobically have been proposed to follow a different route, which involves the use of a hydratase and does not include the formation of the acrylate intermediate (Porra *et al.* 1996). Analysis of the *AcsF* gene has identified that it contains a conserved putative bi-nuclear-iron-cluster motif (Pinta *et al.* 2002). An adenosyl radical is proposed to be formed from adenosylcobalamin, which leads to the formation of a 13^1-radical of Mg-protoporphyrin IX monomethyl ester and the subsequent withdrawal of an electron forming a 13^1-cation. This is then attacked by a hydroxyl ion which gives rise to a 13^1-hydroxy intermediate, which then undergoes the withdrawal of three hydrogen atoms resulting in the cyclisation, forming divinyl protochlorophyllide *a* (Pinta *et al.* 2002).

1.9 Reduction of the 8-vinyl Group

The enzyme 8-vinyl reductase uses NADPH as a reductant and catalyses the reduction of the 8-vinyl group of ring B, to an ethyl group (Parham *et al.* 1995). The reduction of the 8-vinyl group is an essential step in the chlorophyll biosynthetic pathway, however there is still ambiguity as to where in the pathway the reaction is catalysed. Multiple monovinyl and divinyl chlorophyll precursors have been identified in plants and algae, which can not be explained by a stepwise linear pathway. Experiments conducted *in vivo* and *in vitro* have shown that 8-vinyl reductase can reduce both 3,8-divinyl protochlorophyllide and 3,8,-divinyl chlorophyllide, however the efficiency of the latter substrate is significantly higher than that of the former (Tanaka *et al.* 2007; Masuda 2008). Likewise, studies conducted with light dependent POR and highly purified forms of monovinyl and divinyl protochlorophyllide, indicated that the enzyme has no preference for either substrate (Heyes *et al.* 2006b). The implication of these observations is that the 8-vinyl reductase can act before or after the double bond reduction catalysed by protochlorophyllide oxidoreductase (Rüdiger 1997; Beale 1999) (Figure 1.17).

The lack of structural and genetic information has hindered the understanding of 8-vinyl reductase. *R. capsulatus* which has mutations in the *bchJ* gene, are still capable of synthesising bacteriochlorophyll, however they do accumulate divinyl protochlorophyllide, whilst strains with a *bchJ/bchL* double mutation, have significantly altered ratios of divinyl to monovinyl protochlorophyllide present in the organism. This led to the proposal that *bchJ* encoded the 8-vinyl reductase (Suzuki *et al.* 1995). Conversely the unicellular red alga *Cyanidioseyzon merolae* synthesises monovinyl chlorophyll *a*, but no homolog of the *bchJ* gene has been identified in the organism's genome. Whilst this implies that another type of 8-vinyl reductase exists, to date no progress has been made into elucidating such a homolog (Nagata *et al.* 2005).

The genus *Prochlorococcus* are unique, as they are the only chlorophyll producing organism which lacks 8-vinyl reductase. This results in the organism only producing divinyl chlorophyll *a* and *b*. The consequence of this is that due to the presence of the two vinyl groups, the bacterium has enhanced light absorption at the blue end of the spectrum (Chisholm *et al.* 1992; Partensky *et al.* 1999; Coleman *et al.* 2007).

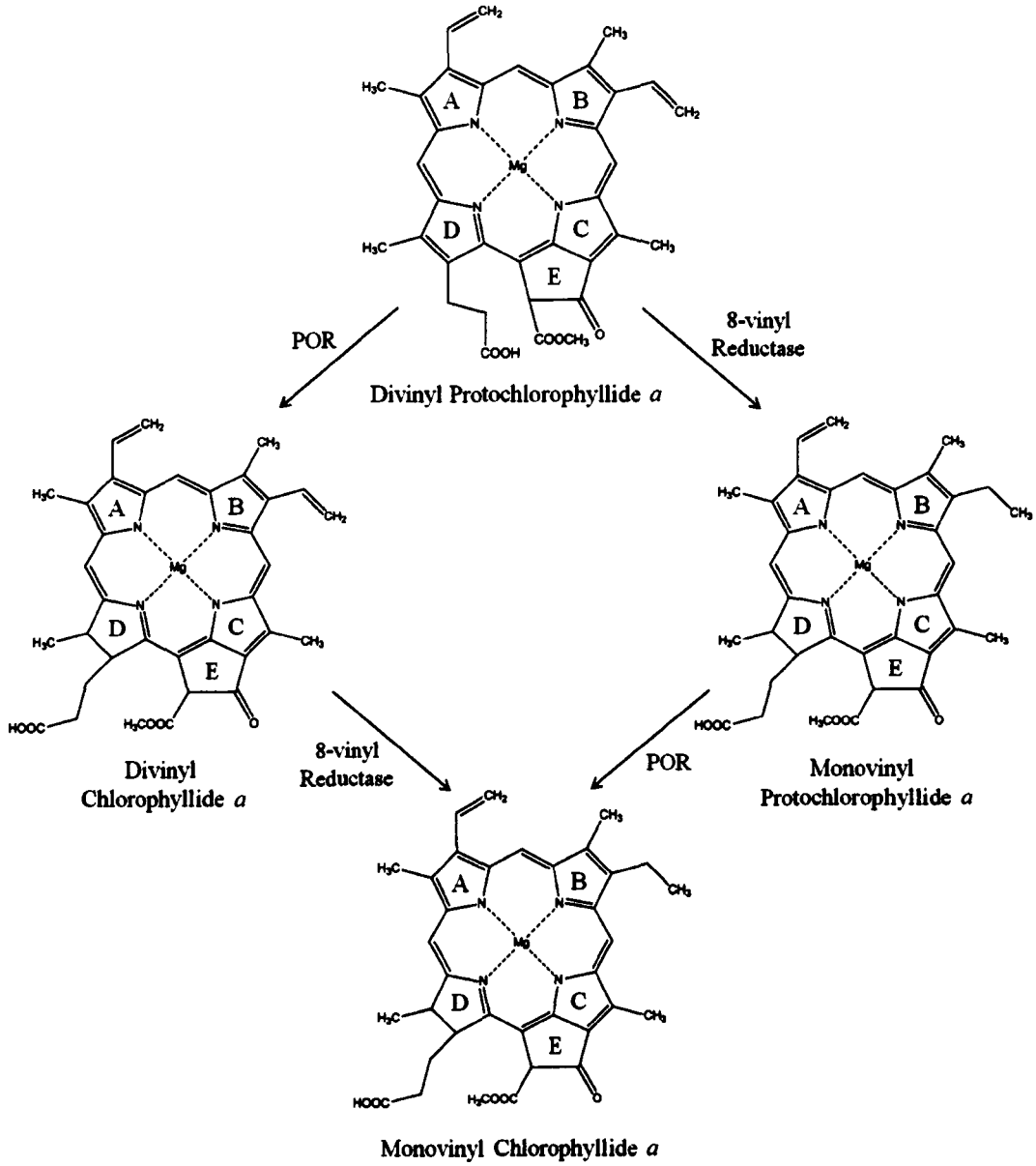


Figure 1.17: Diagram showing the two pathways involved in the production of monovinyl chlorophyllide *a* from the precursor divinyl protochlorophyllide *a* via the intermediates divinyl chlorophyllide *a* and monovinyl protochlorophyllide *a*.

1.10 Reduction of Protochlorophyllide to Chlorophyllide

Protochlorophyllide Oxidoreductase catalyses the reduction of the $C_{17} - C_{18}$ double bond of the D ring of Protochlorophyllide (Pchlide) to form Chlorophyllide (Chlide) (Lebedev *et al.* 1998). The observation that angiosperms germinated in the dark exhibit etiolated yellow leaves, which green upon exposure to light, led to the identification of the light-dependent enzyme NADPH:Protochlorophyllide Oxidoreductase (POR) (Griffiths 1978). However anoxygenic photosynthetic bacteria produce bacteriochlorophyll exclusively

by a light-independent method; this led to the identification of the light-independent protochlorophyllide oxidoreductase (DPOR) (Bauer *et al.* 1993). POR and DPOR can be found in all chlorophyll-containing organisms, with the exception of angiosperms which contain only POR, whilst bacteriochlorophyll-containing organisms contain solely DPOR (Lebedev *et al.* 1998) . Although DPOR and POR are functionally similar, the two enzymes differ in their catalytic mechanism, molecular structure, subunit composition and genomic coding (Masuda *et al.* 2004). All these factors indicate that photosynthetic organisms have independently evolved the two mechanisms of Pchlide reduction (Muraki *et al.* 2010).

1.10.1 Light-independent Protochlorophyllide Oxidoreductase

All anoxygenic photosynthetic bacteria and cyanobacteria contain DPOR (Yamazaki *et al.* 2006) and are able to synthesise (B)Chl in the dark. *In vivo* and *in vitro* analysis indicated that the activity of DPOR is dependent upon the presence of three subunits, BchL/ChlL, BchB/ChlB and BchN/ChlN and the reductant dithionite (Fujita *et al.* 2000; Nomata *et al.* 2005). The three subunits of DPOR show significant homology to the three subunits of nitrogenase NifH, NifD and NifK (Dean *et al.* 1993), indicating that the genes encoding DPOR possibly arose by gene duplication and divergence of the nitrogenase genes (Lockhart *et al.* 1996). DPOR is oxygen sensitive due to the presence of a 4Fe–4S metallocentre in the BchL subunit, which acts as the reductant in the POR reaction. EPR measurements conducted indicated that the BchL subunit is present in dimers in the organism *R. capsulatus,* as found with the nitrogenase subunit NifH (Nomata *et al.* 2006). Since DPOR is present in a number of oxygenic phototrophs, this suggested that DPOR has either developed a mechanism of protection from the produced oxygen, or is tolerant to low levels of oxygen (Masuda 2008).

Three genes encoding DPOR have been identified, which upon mutation result in the accumulation of Pchlide; these have been termed *bch/chl B, bch/chl L* and *bch/chl N* (Zsebo *et al.* 1984; Coomber *et al.* 1990; Yang *et al.* 1990; Burke *et al.* 1993; Bollivar *et al.* 1994b). Colonies of *R. capsulatus* which have mutations in any of the three genes responsible for the three subunits of DPOR accumulate Pchlide when grown in the dark. Upon transfer to the light, the presence of a recombinant pea POR encoded on a separate plasmid was sufficient to complement the mutations in all three genes, thus restoring the ability to form bacteriochlorophyll and the subsequent photosynthetic apparatus. This confirmed that the

three genes identified encoded the three subunits of DPOR, all of which are essential for activity (Wilks *et al.* 1995).

The reasons why some organisms contain both POR and DPOR still remain unclear, however it has been postulated that the two enzymes may be differentially regulated in response to environmental signals, or in response to organism developmental stages (Heyes *et al.* 2009a). This is supported by the cyanobacterium *Plectonema boryanum*, where both POR and DPOR have been observed to contribute to the reduction of Pchlide in the light, however the extent of the contributions from POR increases with increasing light intensities (Fujita 1996).

1.10.1.1 Structure and Mechanism of DPOR

The crystal structure of the catalytic component of DPOR has recently been solved by Muraki *et al* (2010) (Figure 1.18). The authors crystalised the NB protein complex from *R. capsulatus* in the presence and absence of Pchlide. Following crystallisation, they determined that the NB protein exists as a $BchN_2 - BchB_2$ heterotetramer, with a single iron-sulphur cluster at each BchN – BchB interface; this is analogous to the nitrogenase MoFe protein, which is also a $NifD_2 - NifK_2$ heterotetramer. The NB-cluster is co-ordinated by three cysteine residues from BchN and a single highly conserved asparagine residue from BchB. Mutagenesis of the three cysteine residues resulted in a complete loss of activity of the enzyme due to an inability to form a complex with BchB, whilst mutagenesis of the asparagine resulted in a partially active enzyme, due to the mutants still being able to co-ordinate the 4Fe – 4S cluster.

The authors also proposed a mechanism for the *trans*-specific reduction of the $C_{17} - C_{18}$ double bond of Pchlide. Upon binding of Pchlide to DPOR, the C_{17} propinonate of Pchlide is distorted so that it is almost perpendicular to the porphyrin plane. Asp-274 of the BchB subunit is then positioned ideally below the D ring of Pchlide and was therefore proposed to be the candidate for the proton donor for C_{17}. Replacement of the asparagine with an alanine rendered the enzyme inactive, confirming the role of Asp-274 in the catalytic mechanism. As there is no polar amino acid suitably located to donate a proton to the C_{18}, the authors proposed that the C_{17} propionate side chain could act as the proton donor for C_{18}. Substitution of Pchlide with chlorophyll *c*, which has an acrylate instead of a propionate at

the C_{17} position, abolished enzyme activity with the C_{17}–C_{18} double bond of chlorophyll *c* failing to be reduced, thus supporting the proposed theory (Muraki *et al.* 2010).

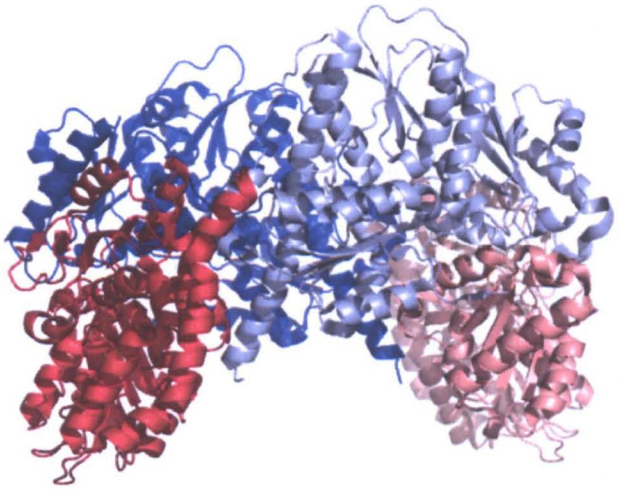


Figure 1.18: Crystal structure of the NB protein of DPOR from *Rhodobacter capsulatus*. The BchN and BchB subunits in one dimer are coloured red and blue respectively, with the identical subunits, which are related by non-crystallographic two-fold symmetry, being coloured light red and light blue respectively (Muraki *et al.* 2010). This model and all other models created have been done using the software Pymol (Schrödinger 2010).

1.10.2 Light-dependent NADPH: Protochlorophyllide Oxidoreductase

NADPH:Protochlorophyllide Oxidoreductase (EC 1.3.1.33), the subject of this thesis, catalyses the light dependent reduction of the C_{17}–C_{18} double bond of Pchlide, forming Chlide (Griffiths 1978). POR is one of only two known enzymes which requires light for catalysis; the other enzyme is DNA photolyase, which uses the energy of near-UV light for pyrimidine dimer repair (Aubert *et al.* 2000). The POR catalysed reaction is an important regulatory step in the biosynthesis of chlorophyll and the subsequent assembly of the photosynthetic apparatus (Lebedev *et al.* 1998).

Dehydrogenase enzymes have been have been reported to enhance the rate of proton and hydride transfer reactions by a factor of up to 10^{17}, when compared to the equivalent reaction in solution (Benkovic *et al.* 2003). With proton and hydride transfers being involved in nearly all reductase reactions (Heyes *et al.* 2003a), understanding of the reaction dynamics is crucial. POR provides an opportunity to fully investigate the reaction dynamics of such

reactions, as POR-Pchlide complex can be pre-formed in the dark and catalysis of the 'caged complex' can be initiated by a single pulse of light. Studies using such a complex removes the diffusion associated substrate-binding events, which occur prior to catalysis, thus allowing the hydride transfer reaction to be studied in greater detail (Heyes *et al.* 2005).

POR is present in all chlorophyll producing organisms and is encoded by a single gene. It is generally accepted that POR evolved from cyanobacteria and its evolution is closely related to the evolution of oxygenic photosynthesis (Fujita 1996; Reinbothe *et al.* 1996; Armstrong 1998). Sequence comparisons of POR with other sequences in protein databases led to the identification of two highly conserved sequences in POR; these were a glycine rich GxGxxG motif, which is characteristic of the Rossmann fold, and the catalytic diad YxxxK, which is characteristic of a tyrosine dependant oxidoreductase (Baker 1994; Wilks *et al.* 1995). The presence of the Rossmann fold is indicative of NAD(P)H binding enzymes however the glycine rich form of the motif is generally found in NADH-dependant oxidoreductases (Perham *et al.* 1991). Whilst this may be the case, experiments conducted with NADH as the co-factor showed that the enzyme was inactive and only active in the presence of NADPH (Griffiths 1978). Identification of these two motifs placed POR in the short-chain alcohol dehydrogenase family, which is a member of the 'RED' super-family of enzymes (Reductases, Epimerases, Dehydrogenases) (Baker 1994; Wilks *et al.* 1995). Members of the short chain alcohol dehydrogenase family are reported to exist generally as dimers or tetramers (Oppermann *et al.* 2003) however initial characterisation of the POR polypeptide using radioactive thiol-specific reagents led to the discovery that POR is a monomer with a molecular weight corresponding to approximately 36,000 Daltons (Oliver *et al.* 1980). Reports of dimers and higher order aggregates termed the light harvesting Pchlide protein complex (LHPP) have been reported in dark grown angiosperms, however reports of the production of such high molecular weight complexes have been limited to dark grown angiosperms (Reinbothe *et al.* 1999).

When the primary sequence of POR is compared with those of other short chain dehydrogenase enzymes, a significant level of sequence homology is observed throughout. There is however a stretch of 33 amino acids in the POR sequence which is not present in any of the other members of the family and have been proposed to form an extra loop (Figure 1.19). Although the function of this loop remains unknown, roles in substrate binding, membrane association and protein–protein interactions have been postulated (Wilks *et al.* 1995).

```
T. elongatus  --MSDQPRP- --TVIITGAS SGVGLYATKA LANRGWHVIM ACRNLEKAEQ
Synechocystis --MTDQQK-- --TAIITGAS SGVGLYGAKA LADKGWHVVM ACRNLEKTER
E. coli       MFNSDNLKLD GKCAIITGAG AGIGKEIAIT FATAGASVVV SDINADAANH

              AAKNLQIPPE AYTILHLDLS SLASVRGFVE SFRALNRPLR ALVCNAAVYY
              VAKEVGIPEA SRTIMHLDLA DFDSVRKFVA DFRATGKTLN SLVCNAAVYL
              VVDEIQQLGG QAFACRCDIT SEQELSALAD FAISKLGKVD ILVNNAGGGG

              PLLKEPIYSV DGYEITVATN HLGHFLLINL LLEDLKNSPE SDKRLVILGT
              PLAKEPQRNK DGYELCVATN HLGHFLLCNL MLEDLKNSPA ADKRLVILGT
              PKPFD--MPM ADFRRAYELN VFSFFHLSQL VAPEM----E KNGGGVILTI

              VTANRKELGG KIPIPAPPDL GNLEGFEKGF KKPIAMINGK PFKSGKAYKD
              VTANPKEVGG KIPIPAPPDL GDLQGMAAGF KPPVAMIDGK IFKPGKAYKD
              TSMAAENK-- ---------- ---------- ---------- -NINMTSYAS

              SKLCNMLTAR ELHRRFHEST GIVFNSLYPG CVADTPLFRH HFPLFQKLFP
              SKLCNILTMR ELHNRYHKDT GIIFNSFYPG CVAETGLFRN HYGLFRKIFP
              SKAAASHLVR NMAFDL-GEK NIRVNGIAPG AIL-TDALKS VIT-PEIEQK

              LFQKKITGGY VSQELAGERV AMVVADPEFR QSGVHWSWGN RQKEGRKAFV
              WFQKNITGGY VTEEVAGERL AKVVADSGFD VSGVYWSWGN RQQQGREAFM
              MLQHTPIRRL GQPQDIANAA LFLCSPAASW VSGQILTVSG GGVQELN---

              QELSAEASDE QKARRLWELS EKLVGLA-
              QEVSDEALDD NKADVLWDLS AKLVGMPA
              ---------- ---------- --------
```

Figure 1.19: Sequence homology stacks of POR from the organisms *Thermosynechococcus elongatus* and *Synechocystis*, along with the 7α-hydroxysteriod dehydrogenase from *E. coli*. The loop is coloured in red (Townley *et al.* 2001).

1.10.2.1 Plant Isoforms of POR

When angiosperms are grown in the light, only trace levels of POR are found. However when angiosperms are germinated in the dark, large quantities of POR–Pchlide–NADPH complexes accumulate, forming highly ordered aggregates known as prolamellar bodies, which extrude from prothylakoids (Apel *et al.* 1980; Apel 1981; Batschauer *et al.* 1984; Mosinger *et al.* 1985; Benli *et al.* 1991; Forreiter *et al.* 1991). Prolamellar bodies have a fluorescence maximum at 655 nm which is significantly red shifted from that of free Pchlide, which has a fluorescence maximum of 631 nm. The reason for this red shift was suggested to be a result of the pigment aggregating together upon the formation of prolamellar bodies (Lebedev *et al.* 1995; Younis *et al.* 1995; McEwen *et al.* 1996). Following exposure to light, the aggregates disassociate and the stored Pchlide is converted to Chlide. This is accompanied by a spectral blue shift, known as the 'Shibata' shift, which is thought to be a result of the disassociation of the aggregates (Zhong *et al.* 1996).

In higher plants three isoforms of POR have been identified to exist; these are POR A, POR B and POR C (Armstrong *et al.* 1995; Holtorf *et al.* 1995; Oosawa *et al.* 2000). All isoforms of POR are nuclear encoded and translated in the cytoplasm as a 41 kDa precursor protein, before being transported across the plastid envelope of chloroplasts. Once in the chloroplast, the precursor protein is cleaved by a stromal-processing peptidase, which results in the mature 37 kDa mature protein (Teakle *et al.* 1993; Masuda *et al.* 2004). The discovery of this method of transport into the chloroplast was made by fusing the POR A precursor sequence to a reporter protein of cytosolic dihydrofolate reductase from mouse. Upon stimulation by Pchlide, import of the dihydrofolate reductase protein into the chloroplast was observed. This not only identified the precursor protein responsible for import, but also suggested that the POR A precursor interacts with Pchlide in the plastid envelope (Reinbothe *et al.* 1997). Conversely, reports have also appeared indicating that in the absence of Pchlide, the import of POR into chloroplasts is not significantly affected (Dahlin *et al.* 1995; Aronsson *et al.* 2000).

The three different isoforms of POR present in higher plants are all highly similar, however no phylogenetic relationship is observed between them, indicating that each isoform originated from individual gene duplication events (Masuda *et al.* 2004). The three different isoforms of POR show different patterns of light and developmental regulation, implying that each isozyme is required at a different stage of the greening process, to enable plants to adjust their requirement for chlorophyll biosynthesis (Schoefs *et al.* 2003).

1.10.2.1.1 The Light Harvesting Pchlide Binding Protein Complex

It has been proposed that two of the POR isoforms, POR A and POR B can form higher molecular weight light harvesting Pchlide protein complexes (LHPP). These complexes are proposed to consist of 6 units of POR A and POR B in the ratio of 5:1 respectively. In plants a number of different Pchlide analogues exist, including Pchlide *a* and Pchlide *b* which bind to POR B and POR A respectively. It was observed that upon illumination, only the POR B–Pchlide *a* complexes were photoconvertible which led to the suggestion that Pchlide *b* may operate as a light scavenger, thus allowing chlorophyll synthesis to occur in low light intensities (Reinbothe *et al.* 1999; Reinbothe *et al.* 2003a; Reinbothe *et al.* 2003b). However the formation of LHPP has been shown not to be a plant POR specific reaction. Masuda *et al.* (2008) showed that an *Arabidopsis* POR A knockdown

mutant, which contained a cyanobacterial POR could complement the mutation and a functional LHPP was produced. Although it was unclear as to whether the cyanobacterial POR had bound both Pchlide *a* and Pchlide *b*, the formation of such complexes was shown to occur with a cyanobacterial enzyme (Masuda *et al.* 2009). Further to the discovery of the LHPP Reinbothe *et al.* made a series of POR A and B mutants where the 33-residue loop had been mutated out completely (Reinbothe *et al.* 2003b). The authors reported that the deletion of the loop had not affected the import of the protein into the chloroplast, the stability of the mature protein or even the ability of the enzyme to bind its pigments. However the mutants were not able to form LHPP thus implicating the loop in protein–protein interactions. More recent work published has detailed that the formation of such complexes is highly controversial and the work conducted above has been criticised for a number of reasons (Armstrong *et al.* 2000). The main point of concern is that according to the proposed model the predominant Pchlide species present in dark grown plants would be Pchlide *b*. Whilst Pchlide *b* has been detected in green plants, it is not the predominant species (Shedbalkar *et al.* 1991) and when the etioplast membrane is supplemented with Pchlide *b* it is photo-reduced to Chlide *a* (Scheumann *et al.* 1999). Also whilst this model is suitable for organisms which have multiple isoforms of POR, no such complexes have been reported in organisms which contain only a single POR protein. It has therefore been reported that the formation of the LHPP may be specific to monocotyledonous plants (Masuda *et al.* 2004).

1.10.2.1.2Regulation of POR in Plants

POR A and POR B mRNAs are both found in etiolated seedlings; whilst POR B mRNA continues to accumulate in light grown plants, PORA mRNA rapidly depletes following illumination. This suggests that POR A is only active in etiolated seedlings at the beginning of illumination, whilst POR B operates throughout the entire greening process in light-adapted mature plants. The activity and expression of POR C has also been shown to be induced by light. It has also been documented that the level of POR C increases upon illumination (Holtorf *et al.* 1995; Oosawa *et al.* 2000; Su *et al.* 2001; Pattanayak *et al.* 2002; Masuda *et al.* 2003). Experiments were done where mRNA synthesis was inhibited by a transcription inhibitor, cordycepin. The POR A mRNA was found to be unstable in the presence of the inhibitor, whilst the presence of the inhibitor appeared to stabilise the POR B mRNA. This suggested that two independent mechanisms existed for the degradation of

POR A and POR B mRNA (Holtorf *et al.* 1996b). Proposals have been made that POR A mRNA is mediated by phytochrome, whilst POR B mRNA is not (Batschauer *et al.* 1984; Mosinger *et al.* 1985; Holtorf *et al.* 1995)

Reports have also appeared indicating that the expression of the POR gene in plants, is under the control of circadian and diurnal rhythms. The accumulation of Pchlide in the leaves of *Phaseolus vulgaris* followed a circadian oscillation when grown in complete darkness (Argyroudi-Akoyunoglou *et al.* 1996). Conversely the levels of POR B mRNA in greening barley showed a diurnal rhythm, when grown under a diurnal 12 h light/12 h dark cycle. When plants were grown in this way, trace amounts of POR A mRNA were observed to appear at the end of the night period, thus implicating POR A in the synthesis of chlorophyll at the start of each day. The regulation of the POR genes by diurnal and circadian regulation may provide a way for the plant to regulate the POR transcripts present thus optimising the synthesis of chlorophyll during the day and night (Holtorf *et al.* 1995; Holtorf *et al.* 1996a; Holtorf *et al.* 1996b).

1.10.2.2 POR Homology Model

A homology model of POR from *Synechocystis* has been published, based on the known structures of other members of the RED superfamily (Townley *et al.* 2001). Townley *et al.* compared the *Synechocystis* POR sequence to known structures from the SCOP database and identified three tyrosine dependent oxidoreductases, which were the most suitable structural templates for modelling. The three structures identified were the short chain alcohol dehydrogenase from *Mus musculus*, 3α, 20β–hydroxysteroid dehydrogease from *Streptomyces hydrogenans* and the 7α–hydroxysteroid dehydrogenase from *E. coli*. All of the three structures identified showed complete conservation of the glycine rich motif of the Rossmann fold along with the tyrosine and lysine catalytic diad, however none of the structures identified had any equivalent homology to the 33-residue insertion which is present in POR (Figure 1.19). The substrate for 7α–hydroxysteroid dehydrogenase was the most similar in shape to Pchlide and as a result Townley *et al.* selected this as the structural template for POR (Townley *et al.* 2001).

The structure of POR was predicted to contain a central parallel β-sheet comprised of 7 β–strands surrounded by 9 α-helices (Figure 1.20); the β-sheet corresponds to the

Rossmann fold, whilst the α-helices are arranged in a way to provide a cofactor binding pocket. The conserved tyrosine and lysine residues have been modelled into the 6th α-helix in such a way that the side chains are in a position to interact with the Pchlide. Based purely upon secondary structure prediction, the 33–residue insertion was built as a strand–turn helix motif and was placed in the model between the fifth and sixth β-sheets. Following predicting the structure of the loop, a sequence motif of a similar structure was identified in the structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (Townley *et al.* 2001).

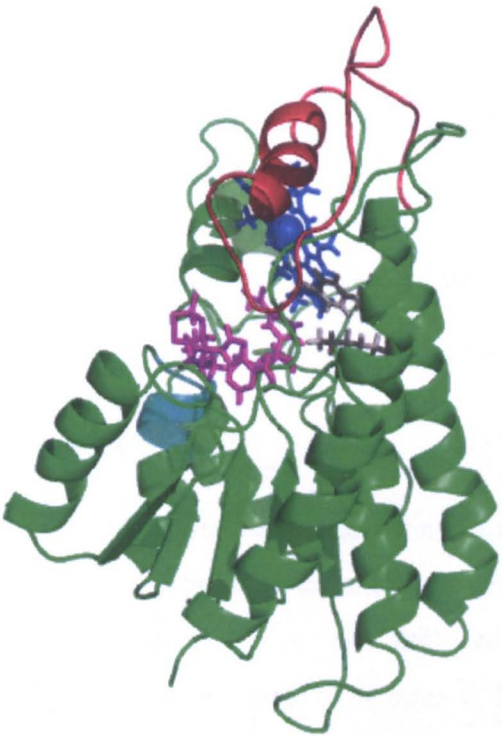


Figure 1.20: Homology model of POR (Townley *et al.* 2001) containing a bound NADPH (magenta) and Pchlide (Dark blue) molecule. The glycine rich NADPH section of the Rossmann fold is indicated in light blue, whilst the highly conserved catalytic diad is shown in grey. The 33-residue insertion is shown in red to form a lid over the top of the protein following the binding of Pchlide.

The authors checked the predicted structure of POR by making site directed mutants. Arg-34 and Cys-33 are located in the cofactor binding pocket and have been proposed to play a role in the enzymes specificity and ability to bind NADPH only. It has been proposed that arg-34 can form a salt bridge with the 2'-phosphate group of the NADPH and mutagenesis of the arginine residue to valine results in more than a 10-fold reduction in enzyme activity,

whilst an arginine, cysteine double mutant abolishes all detectable activity completely. Enzymes which use NADH as a cofactor do not have an arginine residue in position 34 and are also generally preceded by an aspartic acid instead of a cystine. All of the above data supports the proposed hypothesis that the two residues are involved in the enzymes specificity to NADPH (Townley *et al.* 2001).

Prior to the production of a homology model Circular Dichroism (CD) experiments were conducted with POR to ascertain the proportion of α-helix, β-sheet and random coil present in POR. The average secondary structure of POR was calculated using CD to be 33 ± 3 % α-helix, 19 ± 0.2 % β-sheet and 48 ± 4 % turns and random coil. Secondary structure prediction by two independent computer programmes PHD and SOPMA identified the proportion of α-helix, β-sheet and random coil present in POR, within error, to be the same as that calculated by CD (Birve *et al.* 1996). The proportions of α-helix, β-sheet and random coil present in the model produced by Townley *et al.*(2001) was 40, 15 and 35 % respectively which is in agreement with the values calculated by Birve *et al.* (1996).

1.10.2.3 POR Catalytic Cycle

Due to the requirement of light to initiate its catalytic reaction, POR has provided an opportunity to study the hydride and proton transfer reactions, which are characteristic of the RED super-family of enzymes (Wilks *et al.* 1995). The ability to form the enzyme-substrate complex has allowed the reaction to be initiated with light at low temperatures, allowing different intermediates of the reaction to be formed and frozen in their catalytic state (Heyes *et al.* 2002b; Heyes *et al.* 2003b; Heyes *et al.* 2004). As a result, the catalytic cycle of POR has been broken down into five distinct steps; an initial light-driven step followed by four 'dark reactions' (Heyes *et al.* 2004).

1.10.2.3.1 The Light Driven Reaction

Using POR from *Synechocystis*, an initial light dependant reaction has been identified in the POR catalytic cycle, which has been documented to occur at temperatures between 120 and 200 K (Heyes *et al.* 2002b). The majority of studies conducted have been done under light saturating conditions, however it has been reported that catalysis can be triggered

following illumination for 50 fs (Heyes *et al.* 2003a). Following illumination at 180 K, fluorescence emission spectra acquired at 77 K show that a peak, corresponding to POR–Pchlide–NADPH complex at 644 nm disappears and is replaced with a non-fluorescent intermediate. Absorbance experiments conducted under the same conditions show that this non-fluorescent intermediate has an absorbance at 696 nm. The experiment was repeated in the presence of $NADP^+$ and no change was observed in the fluorescence spectrum following illumination at 180 K proving that the fluorescence signal at 644 nm is the only signal corresponding to photoactive POR. Following illumination at 180 K, the sample was warmed to room temperature in the dark, before being cooled back down to 77 K. Fluorescence spectra acquired indicated the presence of a new fluorescent species at 674 nm, which corresponds to free Chlide, along with the re-appearance of the fluorescent species at 644 nm. This was also coupled with a decrease in the fluorescence of free Pchlide at 631 nm (Heyes *et al.* 2002b).

At temperatures of 200 K and below, proteins undergo a transition called the "glass transition" which stops all protein motions and results in a complete loss of activity for the majority of enzymes (Dvorsky *et al.* 2000; Vitkup *et al.* 2000; Teeter *et al.* 2001). This, coupled to the results detailed above, implies that enzyme motions are not involved with the light driven initial stage of the catalytic reaction (Heyes *et al.* 2003b) and that illumination of the sample is all that is required to ensure turnover of the enzyme is initiated (Heyes *et al.* 2002b).

Initial studies of the non-fluorescent intermediate, indicated that it may correspond to an ion-radical complex. Illumination of the enzyme-substrate complex at 180 K gives rise to two distinct EPR signals, supporting this theory (Belyaeva *et al.* 1988; Lebedev *et al.* 1999; Heyes *et al.* 2002b), however subsequent ENDOR experiments have suggested that these distinct species may correspond to Pchlide and Chlide cations and are thus by-products of the charge transfer state. Subsequent Stark spectroscopy conducted on the non-fluorescent intermediate, has indicated that the species exhibits a large Stark effect, which is indicative of charge transfer character and has therefore provided clear evidence that charge-separated states are formed between the Pchlide and NADPH substrates, during the initial illumination at 180 K (Raskin *et al.* 2002; Heyes *et al.* 2006a).

1.10.2.3.2Subsequent Dark Reactions

Using POR from *Synechocystis*, two dark reactions were identified to occur to complete the POR catalytic cycle (Heyes *et al.* 2003b). The first dark reaction results in the conversion of the non-fluorescent intermediate into a species which has a fluorescence maximum at 684 nm and an absorbance maximum at 681 nm. The second dark reaction was detailed to occur at temperatures above 205 K and resulted in the conversion of the 684 nm fluorescent species into a new species, which fluoresces at 674 nm. No distinct change in absorbance is observed upon the formation of this new species; however the peak formed in the first dark reaction at 681 nm is detailed to broaden during the second dark reaction. The products of the two dark reactions have been proposed to correspond to a POR–$NADP^+$–Chlide and a free Chlide respectively. Since the second dark reaction can only occur at temperatures above 200 K, it has been inferred that this part of the reaction is associated with protein motions (Heyes *et al.* 2003b).

The use of a thermophilic POR from the organism *Thermosynechococcus elongatus* allowed the identification of an additional two dark reactions, which had previously been overlooked whilst analysing the mesophilic POR (Heyes *et al.* 2004). Characterisation of a thermophilic POR allows the reaction to be followed over a much wider temperature range, since thermophilic enzymes are fractionally active at room temperature when compared to their relative mesophilic forms, due to reduced conformational flexibility (Závodszky *et al.* 1998; Kohen *et al.* 1999; Fields 2001).

The first dark step identified using *Thermosynechococcus* POR was shown to be the same as that previously observed using *Synechocystis* POR, whilst the second dark step, which had previously shown a broadening in absorbance at 681 nm, exhibited a blue shift in absorbance to 671 nm when analysed using the thermophilic enzyme. The following third and fourth dark steps have been shown to be temperature dependent, with the third and fourth dark step only occurring at temperatures above 260 and 294 K respectively. During the third dark step, the fluorescent species at 674 nm is converted into a new species which fluoresces at 689 nm; this is then further converted into a species that fluoresces at 674 nm during the fourth dark step. Each of the fluorescent species has a distinct absorbance with the product of the third dark reaction absorbing at 685 nm and the product of the fourth dark reaction absorbing at 671 nm (Heyes *et al.* 2004).

Further analysis of the product release and substrate binding states of the enzyme, resulted in each one of the previously identified species being linked to distinct enzyme-product/substrate complex. It was also proposed that product release and substrate binding occurs in an ordered way (Heyes *et al.* 2007). Initially the $NADP^+$ is released from the complex in a biphasic process, leaving a POR–Chlide complex. The initial phase was thought to represent a conformational change in the protein which facilitates the subsequent product release and substrate binding events, with the second phase corresponding to the release of $NADP^+$. Following the release of $NADP^+$ from the complex, NADPH rebinds to the complex, forming a POR–Chlide–NADPH complex. This then facilitates the release of the Chlide and the re-binding of Pchlide reforming the active POR–Pchlide–NADPH complex (Heyes *et al.* 2007). A summary of the POR catalytic cycle was created by Heyes *et al.* which detailed each of the characterised light and dark steps along with their corresponding absorbance and fluorescence values (Figure 1.21) (Heyes *et al.* 2004).

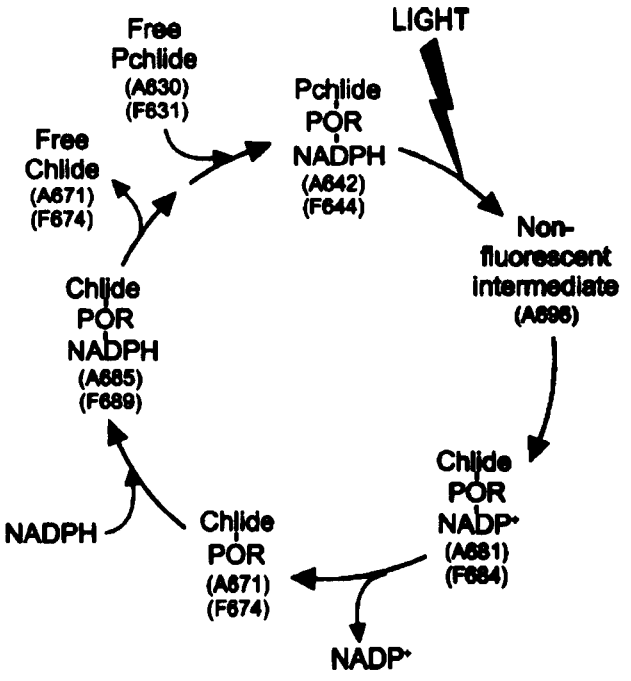


Figure 1.21: The POR catalytic cycle detailing the fluorescence and absorbance values of all enzyme complexes found in the reaction (Heyes *et al.* 2004).

1.10.2.4 Biochemistry of POR

The reaction catalysed by POR is a hydride transfer reaction, with the hydride transferred to the C_{17} position being derived from the pro-*S* face of the cofactor NADPH and the proton transferred to the C_{18} position coming from the conserved tyrosine residue (Valera *et al.* 1987; Begley *et al.* 1989; Wilks *et al.* 1995; Heyes *et al.* 2002a). The absorption of a photon by the Pchlide has been proposed to create a separation of charge across the C_{17}, C_{18} double bond, which induces tensional strain in the molecule, inducing favourable conditions for the initial hydride transfer from NADPH (Wilks *et al.* 1995; Griffiths *et al.* 1996; Heyes *et al.* 2006a). Following the initial reduction step, the negative charge at the C_{18} position is then neutralised with a proton from the conserved tyrosine residue (Griffiths *et al.* 1996). The close proximity of the conserved lysine residue is proposed to be important in lowering the apparent pK_a of the phenolic group of the tyrosine residue, facilitating the de-protonation of the tyrosine residue (Wilks *et al.* 1995; Heyes *et al.* 2002a; Heyes *et al.* 2006a) (Figure 1.22). It was previously thought that POR used FAD as a reducing agent (Walker *et al.* 1988) however this was disproved when NADPH was discovered as the sole reductant (Townley *et al.* 1998).

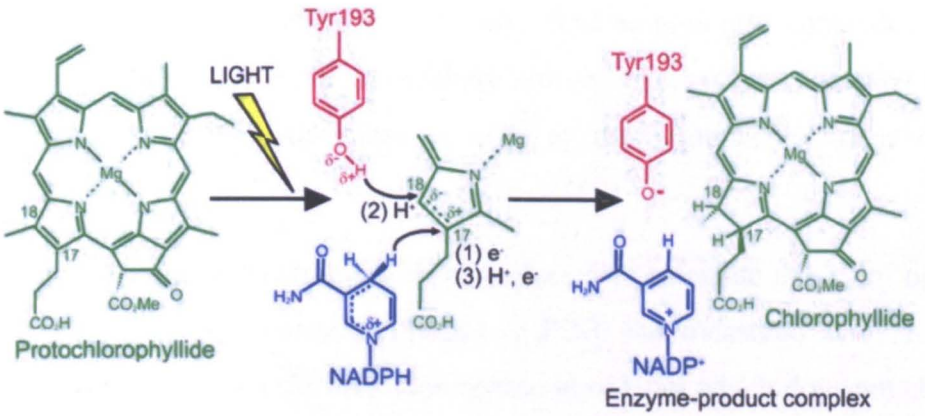


Figure 1.22: The reaction mechanism of protochlorophyllide to chlorophyllide showing the involvement of NADPH and the conserved tyrosine residue (Menon *et al.* 2010).

As previously stated, POR has been assigned to be a member of the SDR family of enzymes. A characteristic of this family is the conservation of a tyrosine and lysine residue, arranged in the YXXXK catalytic motif, which are critical for activity of all enzymes in the SDR family (Ensor *et al.* 1991; Obeid *et al.* 1992; Chen *et al.* 1993; Varughese *et al.* 1994).

Mutagenesis of the two conserved residues in POR also results in an enzyme which is inactive (Wilks *et al.* 1995; Lebedev *et al.* 2001; Heyes *et al.* 2002a). Recent studies have also shown that mutagenesis of the two conserved catalytic residues also affects the enzyme's ability to bind Pchlide (Menon *et al.* 2009). Lysine mutants have more of an effect on Pchlide binding than tyrosine mutants. When coupled to the observations made by Klement *et al* (1999), that Pchlide analogues that have different C_{17} side chains were not accepted as substrates for POR (Klement *et al.* 1999), this supports claims previously made about the conserved lysine residue interacting with the carboxyl group attached to the C_{17} position of Pchlide (Lebedev *et al.* 2001).

Low temperature analysis of the catalytic mechanism, with a tyrosine to phenylalanine mutant, shows that a similar yield of non-fluorescent intermediate is formed with the mutant enzyme to that observed with wild type. Analysis also shows that the rate of hydride transfer in the mutant and wild type enzymes, is similar when assayed at cryogenic temperatures (Menon *et al.* 2009). A difference is observed however in the fluorescent species formed during the first dark reaction, with the mutant exhibiting a lower fluorescence yield which is also blue-shifted by 5 nm, when compared to wild type. The authors infered from this that the rate of hydride transfer in the mutant enzyme was not affected and that the conserved residues are involved in stabilisation of the photo-excited state, allowing the hydride transfer reaction to occur more efficiently. The authors also concluded that whilst the conserved tyrosine residue is the most likely proton donor, other conserved residues or the solvent could be sufficiently close enough to the pigment to transfer a proton (Menon *et al.* 2009).

Fluorescence measurements have been conducted to calculate the K_d of both Pchlide and NADPH binding. The binding of Pchlide to POR was measured at 77 K. At such temperatures, free Pchlide has a distinct fluorescence at 631 nm which does not change upon the addition of POR alone. The inclusion of NADPH into the sample resulted in the fluorescence maximum red shifting to 644 nm. The ratio of the 644:631 could therefore be used to determine the equilibrium binding constant of Pchlide to POR. In the presence of NADPH the K_d for Pchlide was calculated to be 7.7 ± 0.7 μM for POR from *Synechocystis* and 5.6 ± 0.6 μM for POR from *Thermosynechococcus elongatus* (Heyes *et al.* 2002b; Menon *et al.* 2010). When twice the concentration of $NADP^+$ is included in the sample instead of NADPH, a shoulder appears on the red edge of the Pchlide signal at 641 nm, implying that POR can also bind Pchlide in the presence of $NADP^+$ (Heyes *et al.* 2002b).

NADPH binding has been measured by both FRET and NADPH fluorescence. The FRET signal has been shown to arise from a single tryptophan in the active site, whilst an enhancement of NADPH fluorescence is observed to occur at 454 nm following binding to the enzyme (Heyes *et al.* 2000; Townley *et al.* 2001). In the absence of Pchlide, the K_d for NADPH was calculated to be 0.021 ± 0.009 μM for POR from *Thermosynechococcus elongatus,* whilst the K_d for $NADP^+$, in the absence of Pchlide, was calculated to be 40 times weaker at 0.87 ± 0.22 μM, thus implying that POR binds NADPH significantly stronger than it binds $NADP^+$ (McFarlane *et al.* 2005; Menon *et al.* 2010). Recent NADPH binding studies have shown that the kinetics of NADPH binding were very complex and have subsequently been fitted to three distinct phases; an initial phase, thought to involve the anchoring of the ADP moiety of NADPH to the active site, followed by rearrangements of the enzyme, to ensure the correct positioning of the nicotinamide ring. Finally small structural changes are thought to occur within the enzyme to optimally align the cofactor, prior to hydride transfer from the NADPH to Pchlide (Heyes *et al.* 2009b).

The involvement of cysteine residues in the POR reaction has been demonstrated by the use of thiol-modifying reagent N-ethyl maleimide (Oliver *et al.* 1981; Heyes *et al.* 2000). Three cysteine residues have been identified to be absolutely conserved throughout all POR enzymes; these are Cys-37, Cys-199 and Cys-226 (*Thermosynechococcus elongatus* numbering). Mutants which contained the Cys-226 mutation were considerably reduced in their ability to form Chlide and showed a significant decrease in the ability to bind pigment. Mutants which contained the Cys-37 mutation, however showed a significantly higher K_d for NADPH which has been attributed to affecting the interaction of the neighbouring Arg-38 with the 2'-phosphate group of the NADPH molecule (Menon *et al.* 2010).

Analysis of the Cys-226 ternary complex has shown that what little activity occurs, does so below the glass transition temperature, on a faster time scale and by a different catalytic mechanism to that of the previously observed wild type POR. It has been proposed that the Cys-226 mutation alters the POR active site, decreasing the donor-acceptor distance for proton transfer and allowing ring A of the Pchlide molecule to interact with Cys-226. Removal of the interaction has been proposed to alter the position of the Pchlide molecule in the enzyme active site and to increase the distance between the C_{17} of Pchlide and NADPH. It has also been suggested that this event is coupled to a shortening of the distance between the catalytically active tyrosine and the C_{18} of Pchlide. This provides a possible explanation

for how protein motions could be eradicated and how the reaction can occur below the glass transition temperature (Menon *et al.* 2010).

1.11 Esterification of Chlorophyllide to Chlorophyll

The final step in the biosynthesis of chlorophyll *a* is the esterification of a phytol tail onto the propionate side chain of ring D; a reaction catalysed by the enzyme chlorophyll synthase (Beale 1999) (Figure 1.23). The enzyme responsible for the final reaction has been shown in cucumber to use both monovinyl and divinyl chlorophyllide *a* as substrates, whilst the enzyme from *Synechocystis* sp. PCC 680 can use both chlorophyllide *a* and *b* but not bacteriochlorophyllide (Oster *et al.* 1997; Adra *et al.* 1998).

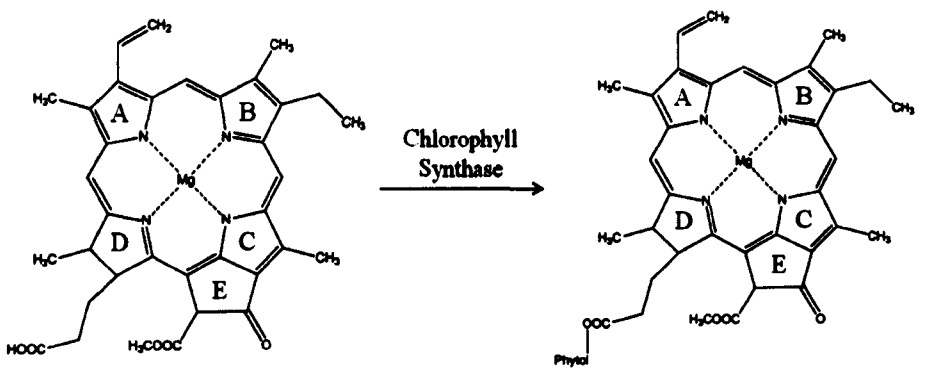


Figure 1.23: Diagram detailing the formation of chlorophyll a from the precursor monovinyl chlorophyllide *a* by the enzyme chlorophyll synthase.

The phytol tail of chlorophyll molecules constitutes approximately 30 % of the molecular weight of chlorophyll and makes the chlorophyll molecule hydrophobic. Whilst the use of other alcohol moieties has been identified, the phytol tail is generally a C_{20} isopropenoid alcohol (Scheer 1991). It has been suggested that its addition to the chlorophyllide molecule requires both NADPH and ATP and occurs in two stages: an initial esterification of chlorophyllide *a* and geranylgeraniol and three successive hydrogenations of the alcohol moiety. Chlorophyllide *a* has been identified to be esterified with geranylgeranyl-PP, an activated form of the precursor to phytol. This is followed by the successive

reductions, of three of the four C-C double bonds present, by a hydrogenase forming the phytol tail (Schoch 1978; Schoch *et al.* 1978; Benz *et al.* 1980).

It is currently unknown whether this reaction is catalysed by a single or multiple enzymes. Two genes have been identified to be crucial for this final reaction, *chl/bch P* and *chl/bch G*, however mutations in the *bchP* gene resulted in bacteriochlorophyll being esterified with geranylgeraniol (Addlesee *et al.* 1996; Addlesee *et al.* 1999). Suggestions have therefore been made that the product of *chl/bch P* is the reductase, whilst the gene product of *chl/bch G* is the synthase (Oster *et al.* 1997; Addlesee *et al.* 2000). A single step esterification process has been observed in spinach chloroplasts, where phytyl-PP is attached to the chlorophyllide. The enzyme geranylgeranyl-PP reductase has been identified in the chloroplast envelope and is responsible for the conversion of free geranylgeranyl-PP to phytyl-PP (Soll *et al.* 1983; Rüdiger 1987).

1.12 Thesis Overview

The work conducted in this thesis has been done solely on the POR isoform from the thermophilic organism *Thermosynechococcus elongatus,* which has been over-expressed and purified in *E. coli.* In comparison to plants, bacteria contain only a single light-dependent POR gene, however sequence alignment has shown that the single bacterial POR sequence is homologous to all POR proteins found in plants. The use of a thermophilic form of POR will allow the NMR experiments to be conducted at a higher temperature, thus increasing the quality of the spectrum acquired, by decreasing the correlation time of the protein.

This thesis attempts to assign the POR HSQC spectrum and determine the secondary structure of the enzyme POR using NMR. POR has a molecular weight of 37 kDa which is large for NMR and means that there is a large gain in sensitivity and lineshape from deuteratation of the protein, however work done before the start of this project resulted in a relatively low yield of protein per litre of LB medium. The thesis therefore starts with a discussion of the work done to optimise the over-expression and purification of the unlabelled enzyme, enabling a triple labelled sample to be produced at much higher yield (chapter 3). As no previous NMR work has been conducted on POR, I then go on to discuss the conditions required to optimise the quality of the NMR spectrum achieved along with the problems encountered whilst working with a triple labelled protein sample (chapter 4). The

results discussed in chapter 4 led to the conclusion that the sample which would provide the best quality data would be a ^{13}C ^{15}N labelled sample; the analysis of the backbone assignment and the structural studies conducted on such a sample are discussed in chapter 5. It was wondered if the poor quality of spectra obtained was a result of restricted mobility of the 33-residue insertion loop present in POR. It was hoped that binding of the substrates to POR would reduce the mobility of the loop and improve the spectral quality. I therefore go on to discuss the protocol developed to deliver the substrates to POR and purify the ternary complex into a detergent- and solvent-free environment, along with the analysis of the complex produced and the reaction kinetics (chapter 6).

CHAPTER 2

Materials and Methods

2.1 Materials

All chemicals were obtained from Sigma and were of the highest analytical grade, unless otherwise stated. Chemicals used for SDS and NATIVE PAGE were obtained from Invitrogen whilst all chromatographic resins were obtained from G. E. Healthcare.

2.2 Standard Buffers, reagents and media

All growth media and buffers were produced as described in (Sambrook *et al.* 1989) and sterilised by autoclaving at 126 °C, unless detailed otherwise. All media and buffers were produced using distilled water further purified by a Milli-Q system. Water based antibiotics were sterilised by passing through a 0.2 μm syringe filter. Media and buffer recipes are detailed in Appendix 1.

2.3 Bacterial strains and plasmids

The bacterial strains and plasmids used are detailed in table 2.1 and 2.2 respectively. All conditions were optimised in Liquid Broth (LB) medium with antibiotics added, when required, at the following concentrations (μg/ml): chloramphenicol, 25, kanamycin, 30, rifampicin, 25.

All labelled growths were grown in M9 minimal media containing ^{13}C labelled carbon and ^{15}N labelled nitrogen with antibiotics added, when required, at the same concentrations as stated above. For triple labelled growths, heavy water (D_2O) was used instead of Milli-Q water and fully deuterated ^{13}C glucose was used as a carbon source.

Strain	Properties	Source/Reference
***E. coli* BL21(DE3) pLysS-T1^R**	*E. coli* B F$^-$ dcm ompT hsdS(r_B^- m_B^-) gal λ (DE3) [pLysS Can^R]	Stratagene
R. capsulatus ZY5	*bchL*::Km, *rif*-10	(Yang *et al.* 1990)

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

Plasmid	Properties	Source/Reference
pET-9His	pET9a derivative containing the XbaI-NdeI (ribosome binding site and six histidine tag) DNA fragment of pET14b (Novagen), Nm^R	Dr. Lucien Gibson, unpublished data. (Heyes *et al.* 2000)

Table 2.2: Details of the plasmids used.

2.4 Production of competent *E. coli* cells

E. coli cells were plated out on LB agar (detailed A.1.1) containing appropriate selection. A single cell was picked and grown up in 5 ml LB media (detailed A.1.2) containing appropriate selection and grown overnight at 37 °C with agitation. 200 µl of *E. coli* cells were grown in 10 ml of LB media (detailed A.1.2) in a 50 ml falcon tube, at 37 °C with agitation until an OD_{600} of 0.6 was obtained. The cells were pelleted, and resuspended in 3.3 ml RF1 media (detailed A.3.1) and incubated on ice for 30 min. The cells were centrifuged and the pellet was re-suspended in 1 ml RF2 media (detailed A.3.2). The cells were incubated on ice for 30 min prior to aliquoting into 200 µl samples and storage at -86 °C.

2.5 Transformation of *E. coli* cells

200 µl competent cells, prepared 2.4, were placed in a 14 ml falcon tube and inoculated with 2 µl pET9His + T. POR plasmid DNA. Following incubation on ice for 30 min, the cells were heat shocked at 42 °C for 90 seconds. The cells were incubated on ice for a further 2 minutes before the addition of 800 µl LB media (detailed A.1.2). The cells were incubated at 37 °C with agitation, for 90 minutes and then plated out onto LB agar plates (detailed A.1.1) containing the appropriate selection. Plates were left to grow overnight at 37 °C, then stored at 4 °C.

2.5.1 Production of an *E. coli* BL21pLysS pET9His T. POR glycerol stock

A single transformed *E. coli* (detailed 2.5) was picked and grown overnight in LB medium (detailed A.1.2) containing the appropriate selection, at 37 °C with agitation. 1 ml of prepared cells were placed into a sterile 1.5 ml eppendorf and centrifuged at 13,000 rpm using a bench-top centrifuge. The LB medium was removed from the eppendorf and the transformed *E. coli* were re-suspended in 1 ml of 50:50 LB medium and glycerol, which had been filtered through a 0.2 µm filter. The re-suspended *E. coli* were stored at -86 °C.

2.6 Over-expression of pET9His T.POR in *E. coli*

2.6.1 Growth of starter cultures

Using a sterile inoculating loop, a single transformed colony (detailed 2.5) was picked from the plate prepared and inoculated into 100 ml of desired medium, containing the appropriate selection markers. The cells were grown up overnight at 37 °C with agitation.

2.6.2 Small Scale Culture Growths

62.5 ml of desired medium was placed into a 250 ml conical flask along with the appropriate antibiotic selection markers. 0.125 ml starter culture, prepared 2.6.1, was used to inoculate the media and the culture was incubated at 37 °C with agitation.

2.6.3 Large Scale Culture Growths

500 ml media was placed into a 2 l conical flask along with the appropriate antibiotic selection. 1 ml starter culture, prepared 2.6.1, was used to inoculate the media, and the culture was incubated at 37 °C with agitation.

2.6.4 Measuring the Growth of *E. coli*

1 ml of *E. coli* growth was removed from the desired flask using a sterile pipette and transferred to a 1.5 ml cuvette. The cuvette was placed into a Cary 50 spectrophotometer, which had been blanked against sterile medium, and the absorbance at 600 nm was measured. Once cell growths were above an optical density of 0.6 serial dilutions of the samples were conducted using sterile medium.

2.6.5 Protein Induction

When the *E. coli* growths (detailed 2.6.2 and 2.6.3) achieved a suitable OD_{600}, the cells were induced with IPTG (Isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 0.1 mM (detailed A.2.4) and incubated at the appropriate temperature for 2 hours with agitation.

2.7 Harvesting *E. coli* cells

Cell growths were separated out into 400 ml plastic Nalgene centrifuge flasks and centrifuged at 10,000 rpm in a Beckmann centrifuge using a JA14 rotor for 15 minutes at 4 °C. The cell pellet was re-suspended in low salt buffer (detailed A.3.7) and transferred to sterile universal tubes. The re-suspended *E. coli* were stored at -20 °C.

2.8 Fractionation of *E. coli* cells

Re-suspended cell pellets (detailed 2.7) were defrosted and Roche complete protease inhibitor was added to the samples as detailed by the manufacturers instructions. The re-suspended *E. coli* were transferred to a glass sonicating vessel and were sonicated on ice at 15,000 amplitude microns for 15 cycles (15 seconds sonication, 30 seconds rest). The resulting crude mixture was placed into 50 ml Beckmannn centrifuge tubes and was heated at 42 °C for 15 minutes before being centrifuged at 25,000 rpm for 30 minutes at 4 °C in a JA20/25.1 rotor.

2.9 Protein purification

2.9.1 Ammonium Sulphate precipitation

Ammonium sulphate was weighed out and ground down to a fine powder using a pestle and mortar. 25 ml of supernatant (detailed 2.8) was placed into a 50 ml Beckmann centrifuge tube and the appropriate mass ammonium sulphate corresponding to an 85 % cut, was dissolved gradually into the protein solution. The supernatant was left for one hour on a spiromix before centrifuging the precipitate at 20,000 rpm for 30 minutes at 4°C in a JA20/25.1 rotor. Following the removal of the supernatant, the resulting pellet was stored at -20 °C.

2.9.2 Purification of His-tagged POR on a Nickel column

Protein pellets (detailed 2.9.1) were defrosted and re-suspended in 15 ml binding buffer (detailed A.3.3), containing Roche complete protease without EDTA. The re-suspended protein solution was filtered through a 0.45 μm filter before being loaded onto a C 10/20 column (GE Healthcare) containing 5 ml nickel His-60 resin, which had been equilibrated with binding buffer (detailed A.3.3), at a rate of 0.5 ml/min, before being attached to an Äkta Prime chromatography machine. The column was washed with 15 ml of binding buffer (detailed A.3.3) before a 180 mM imidazole gradient was applied to the column over 10 ml. The column was further washed with 20 ml of 180 mM imidazole before a 1 M imidazole gradient was applied to the column over 90 ml. The column was then further washed with a following 15 ml of uber elute buffer, containing 1 M imidazole (detailed A.3.6). All buffers were passed through the column at a rate of 1 ml/min and all eluant was collected in 5 ml fractions using a fraction collector. The absorbance of the eluant was measured using a UV spectrometer and elution traces were created using the software Primeview. The column was stored at 4 °C in 20 % ethanol.

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

20 ml of SP sepharose resin (GE Healthcare) was packed into a C 10/20 glass chromatography column (GE Healthcare) at a rate of 1 ml/min. Following packing of the column, the resin was buffer exchanged into low salt buffer (detailed A.3.7). All fractions eluted from the nickel His-60 column (detailed 2.9.2) which were detected to contain POR were loaded onto the SP sepharose column at a rate of 0.5 ml/min before being attached to an Äkta Prime chromatography machine. The column was washed with 20 ml low salt buffer (detailed A.3.7) before a 2M salt gradient was applied to the column over 100 ml. The column was further washed with 20 ml of high salt buffer (detailed below A.3.8). All buffers were passed through the column at a rate of 1 ml/min and all eluant was collected in 5 ml fractions using a fraction collector. The absorbance of the eluant was measured using a UV spectrometer and elution traces were created using the software Primeview. The column was stored at 4 °C in 20 % ethanol with 0.2 M sodium acetate.

2.9.4 Purification of His-tagged POR on an Mono-S column

A pre-packed Mono-S PC 1.6/5 column was purchased from GE Healthcare and was buffer exchanged into low salt buffer (detailed A.3.7). All fractions eluted from the SP sepharose column (detailed 2.9.3) which were detected to contain POR, were concentrated to 2 ml and diluted 5-fold with low salt buffer. The Mono-S column was attached to an Äkta Prime chromatography machine before loading the sample at a rate of 0.5 ml/min. The column was washed with 20 ml low salt buffer (detailed A.3.7) before a 2M salt gradient was applied to the column over 100 ml. The column was further washed with 20 ml of high salt buffer (detailed A.3.8). All buffers were passed through the column at a rate of 1 ml/min and all eluant was collected in 5 ml fractions using a fraction collector. The absorbance of the eluant was measured using a UV spectrometer and elution traces were created using the software Primeview. The column was stored at 4 °C in 20 % ethanol with 0.2 M sodium acetate.

2.9.5 Purification of His-tagged POR on a Gel Filtration column

A pre-packed HiLoad 16/600 Superdex 200 column purchased from GE Healthcare was attached to an Äkta Prime chromatography machine, ensuring the matrix was kept under positive pressure at all times, and buffer exchanged into gel filtration buffer (detailed A.3.9). All fractions eluted from the Mono-S column (detailed 2.9.4) which were detected to contain POR, were concentrated to 2 ml and loaded onto the column at a rate of 0.5 ml/min. The column was washed with 120 ml gel filtration buffer (detailed A.3.9) at a rate of 1 ml/min and all eluant was collected in 1 ml fractions using a fraction collector. The absorbance of the eluant was measured using a UV spectrometer and elution traces were created using the software Primeview. The column was stored at 4 °C in 20 % ethanol.

2.10 Concentration of His_6-POR

All fractions eluted from the SP sepharose column (detailed 2.9.3) which were detected to contain POR, were combined and placed into an Amicon stirred cell concentrator with a 10 kDa PES membrane. Samples were buffer exchanged into the desired buffer and concentrated to the desired concentration, using nitrogen gas at a pressure of 50 psi.

2.11 Estimating the Concentration of His_6-POR

The concentration of POR was estimated using the Bradford assay. 210 µl of BioRad protein assay was placed into 3.3 ml cuvette and made up to 3 ml with milli-Q H_2O. 3 µl of protein sample was added to two of the cuvettes and mixed with the other cuvette being left as a blank. A Cary 50 was used to measure the absorbance of the protein solutions at 595 nm. The two readings obtained were averaged and the concentration of POR was worked out using Equation 2.1.

$$\mu g\ POR = \frac{A_{595} - 0.02844}{0.01908}$$

Equation 2.1: Equation of the standard curve produced following the using the Bradford assay to measure the absorbance at 595 nm of known concentrations of Bovine Serum Albumin.

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

A 10 ml *E. coli* culture was produced as described in 2.6.1. 2 ml of the *E. coli* culture produced were centrifuged for 5 minutes at 13,000 rpm using a benchtop centrifuge. Plasmid preparation from the pelleted *E. coli* was conducted using a QIAGEN mini-prep kit, as detailed in the QIAGEN manual and the DNA pellet recovered was re-suspended in 50 µl Milli Q water. Plasmid DNA was stored at -20 °C.

2.13 DNA sequencing

10 µl of plasmid DNA (prepared 2.12) was placed in an eppendorf and sent to the core genomic faculty in the Sheffield School of Medicine and Biomedical Sciences to be sequenced. Both forward and reverse data was obtained from sequencing from the T7 promoter and terminator respectively. The sequence data was analysed using the software Finch TV.

2.14 Pigment preparation from *Rhodobacter capsulatus* ZY5

2.14.1 Growth of *R. capsulatus* starter cultures

R. capsulatus ZY5 glycerol stock was streaked out on VN Agar (detailed A.1.4), using appropriate selection and left to grow at 34 °C until colonies had formed. Single colonies were picked and inoculated into 125 ml glass conical flasks containing 100 ml VN Media (detailed A.1.5) containing appropriate selection and left to grow overnight, in the dark at 34 °C with agitation.

2.14.2 Large Scale Growth

R. capsulatus starter cultures (prepared 2.14.1) were transferred to 2 litre flasks containing 1 litre of fresh VN medium (detailed A.1.5) and the appropriate selection. The flasks were grown in the dark at 34 °C, in the presence of four polyurethane foam bungs. Every 24 hours the foam bungs containing Pchlide were removed and replaced with fresh bungs; this process was repeated until no further Pchlide was produced.

2.14.3 Pigment Extraction

Foam bungs detected to contain Pchlide (detailed 2.14.2) were removed from the VN medium and placed into 1 litre of 100 % methanol. The Pchlide was washed from the bungs before being dried down completely in a rotary evaporator. The rotary evaporator was set up with the evaporation flask resting in a 30 °C water bath and with water cooled to 4 °C running through the cooling coil. Following evaporation of the methanol, 50 ml of fresh 100 % methanol was placed into the evaporation flask and swirled until the methanol was saturated with pigment. Following saturation the methanol was removed from the evaporation flask and made up to 1 litre with 100 % acetone. This process was repeated until all the previously dried down pigment had been re-suspended.

2.14.4 Pigment Purification

The re-solubilised Pchlide (detailed 2.14.3) was passed through a Buchner funnel using a Whatman paper filter and the filtrate was collected. 600 ml of the prepared filtrate was loaded onto a CM Sepharose column which had been pre-equilibrated in 100 % acetone. The column was washed with three column volumes of 100 % acetone until all the carotenoids appeared to have been washed off. Contaminants were further washed from the column using three column volumes of 5 % methanol in acetone. Following the removal of all carotenoids, Pchlide was eluted from the column using three column volumes of 25 % methanol in acetone. 5 ml fractions were collected and the absorbance measured between 400 and 700 nm; all fractions which had an absorbance at 630 nm were pooled together and stored for final concentration. The process was repeated until all the prepared filtrate had been purified and the Pchlide extracted from it.

All fractions containing Pchlide were pooled together and concentrated to a volume of 20 ml in a rotary evaporator, set up as detailed in 2.14.3. The concentrated Pchlide was aliquoted into 1 ml black eppendorfs and dried down to a powder by passing nitrogen gas over the top of the eppendorfs, before being stored at -20 °C.

2.15 POR Assays

2.15.1 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 13,000 rpm for 5 minutes, to remove any insoluble Pchlide. The absorbance of Pchlide was analysed at 630 nm using a Cary 50 spectrophotometer. The extinction coefficient of 30.4 mM was used to calculate the concentration of Pchlide present in the sample.

2.15.2 Concentration of NADPH

A 1 ml stock solution of 100 mM NADPH was made up with MilliQ H_2O. The absorbance of NADPH was analysed at 340 nm using a Cary 50 spectrophotometer. The extinction coefficient of 6.2 mM was used to calculate the concentration of NADPH present in the sample.

2.15.3 POR Assay

A 1 ml solution of low salt buffer (detailed A.3.7) containing 75 μM Pchlide, 100 μM NADPH, 5 mM DTT, 50 mM BOG and 1 μM POR was incubated for 2 minutes in a Cary 50 spectrophotometer which had been pre-heated to 55 °C. A programme was set up in Scanning Kinetics to scan between 600 and 750 nm at a rate of 100 nm/second for a period of three minutes. Following initiation of the programme and the collection of a base line spectrum, the sample was illuminated with a KL 1500 electronic fibre optic light source with a 400 – 500 nm filter attached. The reduction of Pchlide and the formation of Chlide were measured by the respective reduction in absorbance at 630 nm along with the increase in absorbance at 670 nm. The files were exported from Scanning Kinetics and were further analysed in Microsoft Excel.

2.16 Formation of a POR-Pchlide-$NADP^+$ Ternary Complex

20 ml of SP sepharose resin (GE Healthcare) was packed into a glass chromatography column at a rate of 1 ml/min. Following packing of the column, the resin was buffer exchanged into low salt buffer (detailed A.3.7). A 2 ml solution containing 50 μM POR, 100 μM Pchlide suspended in methanol, 5 mM $NADP^+$, 25 mM DTT and 50 mM BOG made up in low salt buffer (detailed A.3.7) was heated for 15 minutes in a 42 °C water bath. The sample was diluted up to 10 ml using low salt buffer (detailed A.3.7) and centrifuged at 13,000 rpm for 15 minutes in a bench top centrifuge.

The supernatant was loaded onto the pre-equilibrated SP sepharose column at a rate of 0.5 ml/min and the column was attached to an Äkta Prime chromatography machine. The column was washed with 20 ml low salt buffer (detailed A.3.7) before a 2M salt gradient was

applied to the column over 100 ml. The column was further washed with 20 ml of high salt buffer (detailed belowA.3.8). All buffers were passed through the column at a rate of 1 ml/min and all eluant was collected in 5 ml fractions using a fraction collector. The absorbance of the eluant was measured using a UV spectrometer and elution traces were created using the software Primeview.

Fractions which had an absorbance at 280 nm and were green in colour were collected and concentrated using a viva-spin with a 10 kDa PES membrane attached. Viva-spins were centrifuged at 2,000 rpm in a refrigerated centrifuge chilled to 4 °C.

2.17 HPLC

50 µl of POR-Pchlide-$NADP^+$ ternary complex (detailed 2.16) was loaded onto a Phenomenex S2000 HPLC gel filtration column, which had been pre-equilibrated with HPLC buffer (detailed A.3.10). HPLC buffer was passed through the HPLC column at a rate of 1 ml/min and 500 µl fractions were collected. The absorbance of all eluant was measured at 630 and 280 nm to detect for the presence of Pchlide and protein respectively.

2.18 Fluorescence Experiments

All fluorescence experiments were conducted using a Flurolog Jobin Yvonn Fluorimeter, which was connected to a 25 °C water bath.

2.18.1 Tryptophan Fluorescence

Emission spectra of a 1 ml solution of fluorescence buffer (detailed A.3.12) containing 60 µM POR and increasing concentrations of guanidine hydrochloride were recorded between 300 and 400 nm following excitation at 280 nm. Excitation and emission slit widths were set to 1.5 nm and 3 nm respectively. The quenching of the tryptophan signal at 332 nm was used as a measure of protein unfolding.

2.18.2 NADPH Fluorescence

Emission spectra of a 1 ml solution of fluorescence buffer (detailed A.3.12) containing 20 nM POR and increasing concentrations of NADPH were recorded between 400 and 600 nm following excitation at 340 nm. Excitation and emission slit widths were set to 14 nm and 5 nm respectively. The quenching of the signal at 454 nm was used as a measure of NADPH binding to the protein. Results were plotted out in Excel and fitted using Equation 2.2.

$$F_{obs} = F_o + F_{max} \frac{[L]_T + [E]_T + K_d - \sqrt{([L]_T + [E]_T + K_d)^2 - 4[L]_T[E]_T}}{2[E]_T}$$

Equation 2.2: Equation for the binding of a ligand, where a single type of binding site is assumed. F_{obs} is initial fluorescence, F_{max} is the maximum amplitude of fluorescence quenching, $[L]_T$ is the total ligand concentration, $[E]_T$ is the total concentration of POR protein and K_d is the apparent dissociation constant.

2.19 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

The buffer system of (Laemmli 1970) was used to separate protein samples using precast 16 % SDS PAGE (Invitrogen). Gels were set up as detailed by the manufacturers' instructions, with 200 ml of 1 × MES buffer placed between the gels and 500 ml of 1 × MES buffer in the tank surrounding the gels. Between 5 and 20 µl of protein sample were prepared with an equal volume of 2x sample buffer and loaded onto the gel alongside either pre-stained BioRad SDS broad range markers or Sigma low range molecular markers. The gels were run at 180 mV for 60 min at room temperature and protein bands were visualised using Coomassie brilliant blue R250.

2.20 Native Gels

The buffer system of (Laemmli 1970) was used to separate protein samples, based on the Blue Native Polyacrylamide Gel Electrophoresis technique, developed by Schägger (Schägger *et al.* 1991), using precast 4-16 % *NativePAGE NOVEX* (Invitrogen). 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 500 ml of 1 × NativePAGE running buffer in the tank surrounding the gels. 20 µl of protein sample were prepared with an equal volume of 2x 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.

2.21 NMR Experiments

NMR experiments were conducted using 500, 600 800 and 900 MHz Bruker spectrometers. Experiments conducted using the 600 and 900 MHz spectrometers were done using a cryoprobe to improve the sensitivity of the experiments. The parameters of the experiments conducted are detailed in Table 2.3. All data was processed in Felix 2007 using in-house macros, which can be found at /ld10/home1/andy/felix970/macs. All experiments were processed with a 90 ° sine bell window function over the first 1024 points and a 90 ° sine bell window function over all the nitrogen points.

2.22 Asstools

The chemical shift data of all backbone Co, Cα and Cβ as well as the data for their preceding residues were inputted to the simulated annealing program of the Asstools suite of assignment programs. Asstools uses the protein's amino acid sequence and attempts to reorder the spin systems into the correct sequence, based on minimising an energy function related to both 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.

		Experiment						
		HSQC	HNCO	HNCA	HNCACO	CBCACONH	HNCACB	HNCOCA
Proton	Offset (ppm)	4.71	4.71	4.71	4.71	4.71	4.71	4.71
	Spectral Width (ppm)	25.019	25.019	25.019	25.019	25.019	25.019	25.019
	Acquisition Time (ms)	120	120	120	120	120	120	120
Nitrogen	Offset (ppm)	118	118	118	118	118	118	118
	Spectral Width (ppm)	32.002	32.002	32.002	32.002	32.002	32.002	32.002
	Acquisition Time (ms)	23.1	23.1	23.1	23.1	23.1	23.1	23.1
Carbon	Offset (ppm)		174	54	174	40	40	54
	Spectral Width (ppm)		13.999	27.995	13.999	60.023	60.023	27.995
	Acquisition Time (ms)		29.8	11.8	29.8	6.6	6.6	14.2

Table 2.3: Details of the parameters used for NMR experiments.

CHAPTER 3

Optimising the Over-expression and Purification of His_6 – Thermosynechococcus POR

3.1 Introduction

The original aim of this project was to obtain the structure of POR using NMR. Due to the relaxation problems associated with analysing large protein by NMR, triple labelled protein (^{13}C, ^{15}N, ^{2}H) must be grown and used. Unfortunately growing triple labelled protein is expensive and the yield achieved is often less than that obtained from similar nutrient rich growths. It is therefore essential that the growth and purification of POR is optimised thus allowing the production of as much protein as possible from as little medium as possible.

All results, unless otherwise stated, were obtained using small scale growths and unlabelled protein. With regards to the growth conditions, it was hoped that having scaled every element down by a factor of eight it would give a good representation of how the cells would behave when the process was scaled up to a full 500 ml large scale growth. Likewise it was assumed that the purification and behaviour of labelled protein would be the same as that of unlabelled protein.

3.2 Optimisation of the Growth of *E. coli* and protein expression

E. coli BL21pLysS cells were made competent (detailed 2.4) via the Rubidium chloride method and transformed using the Heat Shock method with 50 ng recombinant pET9a + His6-POR plasmid DNA (detailed 2.5). Transformed cells were plated out onto nutrient rich LB agar containing the relevant antibiotics and incubated overnight at 37 °C. 100 ml nutrient rich starter cultures, containing the relevant antibiotics, were inoculated with a single transformed colony, and grown overnight to completion at 37 °C, with agitation (detailed 2.6.1). The following day, 1 ml of the overnight starter culture was spun down and turned into a glycerol stock and stored at -80 °C (detailed 2.5.1).

3.2.1 Growth of *E. coli* BL21pLysS in baffled and non-baffled flasks both prior to and following the induction with IPTG

Four small scale growths containing 62.5 ml nutrient rich LB media were set up (detailed 2.6.2), two in baffled flasks, and two in non-baffled flasks. The relevant antibiotics were added prior to inoculation with 125 µl overnight starter culture (detailed 2.6.1). The flasks were incubated at 37 °C with agitation and the cell densities of all growths were measured at 600 nm every hour using a Cary 50, zeroed with sterile LB media (detailed 2.6.4), and plotted logarithmically. When cells showed evidence of entering into log phase, one baffled and one non-baffled flask were induced with 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (detailed 2.6.5). Following the induction with IPTG, spectroscopic measurements were continued to be taken and were plotted logarithmically (Figure 3.1).

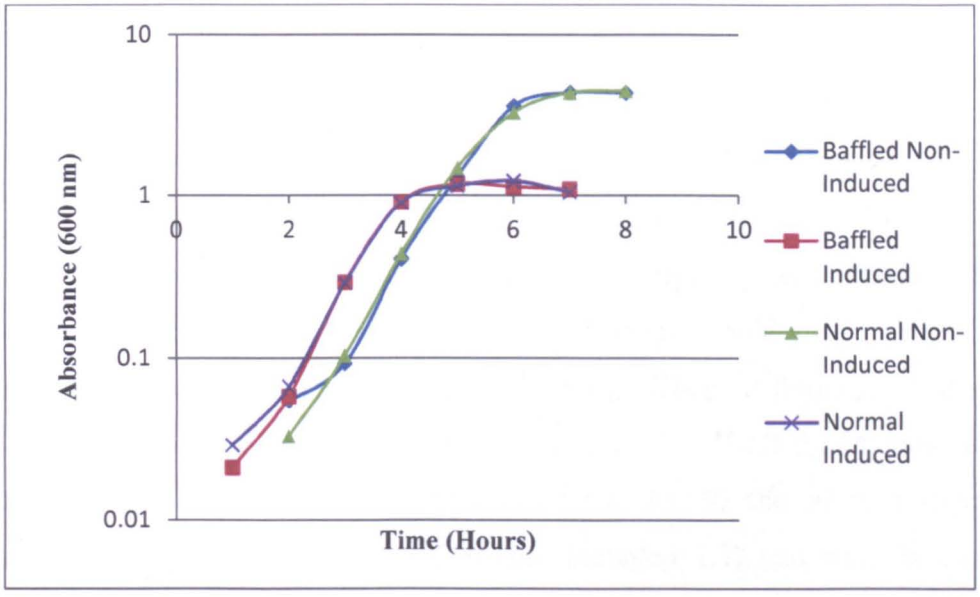


Figure 3.1: Normal and induced growth curves of *E. coli* BL21pLysS in normal and baffled flasks containing LB medium plotted on a log scale.

It is observable from Figure 3.1 that the growth curve of *E. coli* in baffled flasks is the same as that of the *E. coli* in non-baffled flasks. It can therefore be inferred that there is no significant difference between cells grown in baffled and non-baffled flasks. Even though there is no difference in cell growth it is still to be seen whether the use of baffled flasks has an effect on the levels of protein produced following induction.

It is also observable from Figure 3.1 that the growth of *E. coli* is limited following the addition of IPTG. Un-induced *E. coli* will grow to completion following eight hours of incubation, however after induction of protein expression with IPTG, *E. coli* will only grow for a further two hours before entering stationary phase. It is therefore important to optimise the point of induction so that the growth curve of the induced *E. coli* tracks as closely as possible to that of the un-induced *E. coli*. This way, we will have the most cells producing the most amount of protein for the longest amount of time possible.

3.2.2 Temperature optimisation of protein expression in baffled and non-baffled flasks

Six small scale growths containing 62.5 ml nutrient rich LB media (detailed 2.6.2) were set up, three in baffled and three in non-baffled flasks. The relevant antibiotics were added to the flasks prior to inoculation with 125 µl overnight starter culture (detailed 2.6.1). All flasks were incubated at 37 °C with agitation and spectroscopic measurements were taken every hour at 600 nm using a Cary 50 (detailed 2.6.4). When all flasks reached an OD_{600} of 0.6, all were induced with 0.1 mM IPTG (detailed 2.6.5). Baffled and non-baffled flasks were split into three pairs and incubated overnight at 25, 30 and 37 °C respectively with agitation. Cells were harvested at 10,000 rpm (detailed 2.7) and were broken open via sonication (detailed 2.8). Following sonication, the samples were spun at 20,000 rpm for 30 minutes at 4 °C to separate the soluble protein from the cell debris and insoluble protein. Samples were taken of all fractions and analysed using SDS PAGE (Laemmli 1970) (detailed 2.19) to determine the levels of all protein produced by the cells as well as the level of soluble protein present (Figure 3.2).

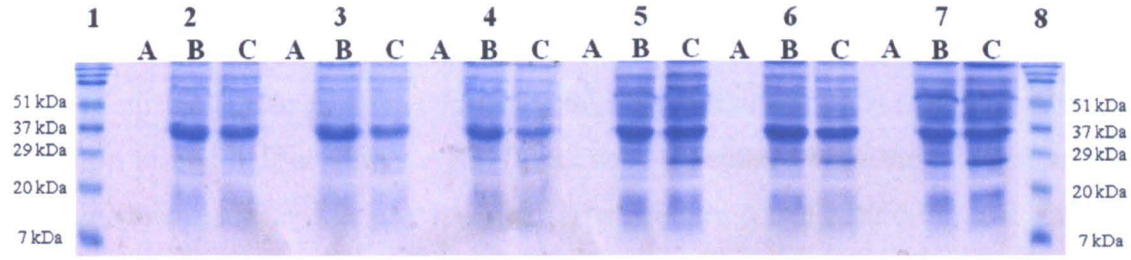


Figure 3.2: SDS PAGE showing the effect of temperature and flask type on the levels of protein present A) in the medium following induction, B) in the soluble fraction following induction and sonication, C) in the insoluble fraction following induction and sonication. Lanes 2, 4 and 6 correspond to normal flasks, whilst lanes 3, 5 and 7 correspond to baffled flasks induced at 37 °C, 30 °C and 25 °C respectively. Lanes 1 and 8 correspond to standard molecular weight markers of a known size.

Following centrifugation, any soluble protein present will be in the supernatant whilst insoluble protein aggregates along with protein present in inclusion bodies will be lost to the pellet. By comparing the intensity of the bands between the sonicated and soluble fractions it is possible to tell which temperature and also which flask type is optimal for the over-expression of His_6-POR.

A strong band corresponding to a protein of 37 kDa was observed in all sonicated and soluble protein lanes (Figure 3.2) which is thought to correspond to POR. As the temperature of induction increases, the intensity of this band decreases whilst all other bands present remain of a comparable intensity. It is also noticeable that as the temperature of induction increases the intensity of the band thought to correspond to POR in the soluble protein fraction decreases. This indicates that not only does less POR get produced at a higher temperature of induction, but also a larger proportion of the protein produced is insoluble or secluded away into inclusion bodies. As a result transformed *E. coli* will be grown at 37 °C to the point of induction when the temperature of the incubator will be reduced to 25 °C.

When comparing the intensity of the protein bands in Figure 3.2, between growths conducted at the same temperature in baffled and non-baffled flasks, no significant difference in intensity is observed. This result coupled to that found from Figure 3.1 led to the conclusion that there was no noticeable benefit to be obtained from using baffled flasks, and that all future growths should be conducted in non-baffled flasks.

3.2.3 Optimising the point of induction of pet9a His_6-POR

Seven small scale growths containing 62.5 ml nutrient-rich LB media (detailed 2.6.2) were set up in non-baffled flasks. The flasks were inoculated with the relevant antibiotics prior to inoculation with 125 µl overnight starter culture (detailed 2.6.1). All flasks were incubated at 37 °C with agitation and spectroscopic measurements were taken every hour at 600 nm using a Cary 50 (detailed 2.6.4). When flasks reached an OD_{600} of 0.3, 0.5, 0.7, 0.9, 1.7 and 1.9, one flask was induced with 0.1 mM IPTG (detailed 2.6.5) and incubated for the remainder of the eight hour experiment at 25 °C with agitation. The growth of *E. coli* prior to and following induction was plotted out logarithmically and compared to the growth found in an un-induced flask (Figure 3.3). Cells were harvested at 10,000 rpm (detailed 2.7) and were broken open via sonication (detailed 2.8). Following sonication, the samples were spun at 20,000 rpm for 30 minutes at 4 °C to separate the soluble and insoluble protein. Samples were taken of all fractions and analysed using SDS PAGE (detailed 2.19) to determine the levels of protein produced by the cells as well as the level of soluble protein present (Figure 3.4).

As observed in Figure 3.1, irrespective of the point of induction with IPTG, Figure 3.3 indicates that the growth of *E. coli* appears to continue for 2 hours before the cells enter stationary phase. It is also apparent from Figure 3.3 that the later the cells are induced the closer the growth of the induced cells is to that of the un-induced cells, to the point that cells induced at OD_{600} of 1.7 and 1.9 track very closely to that of the un-induced *E. coli*. Figure 3.4 indicates that as the point of induction increases, more protein is produced per ml of medium. This is to be expected as there is a greater cell density present, however it also confirms the idea that inducing *E. coli* at an OD_{600} of 1.7-1.9 would increase protein yield.

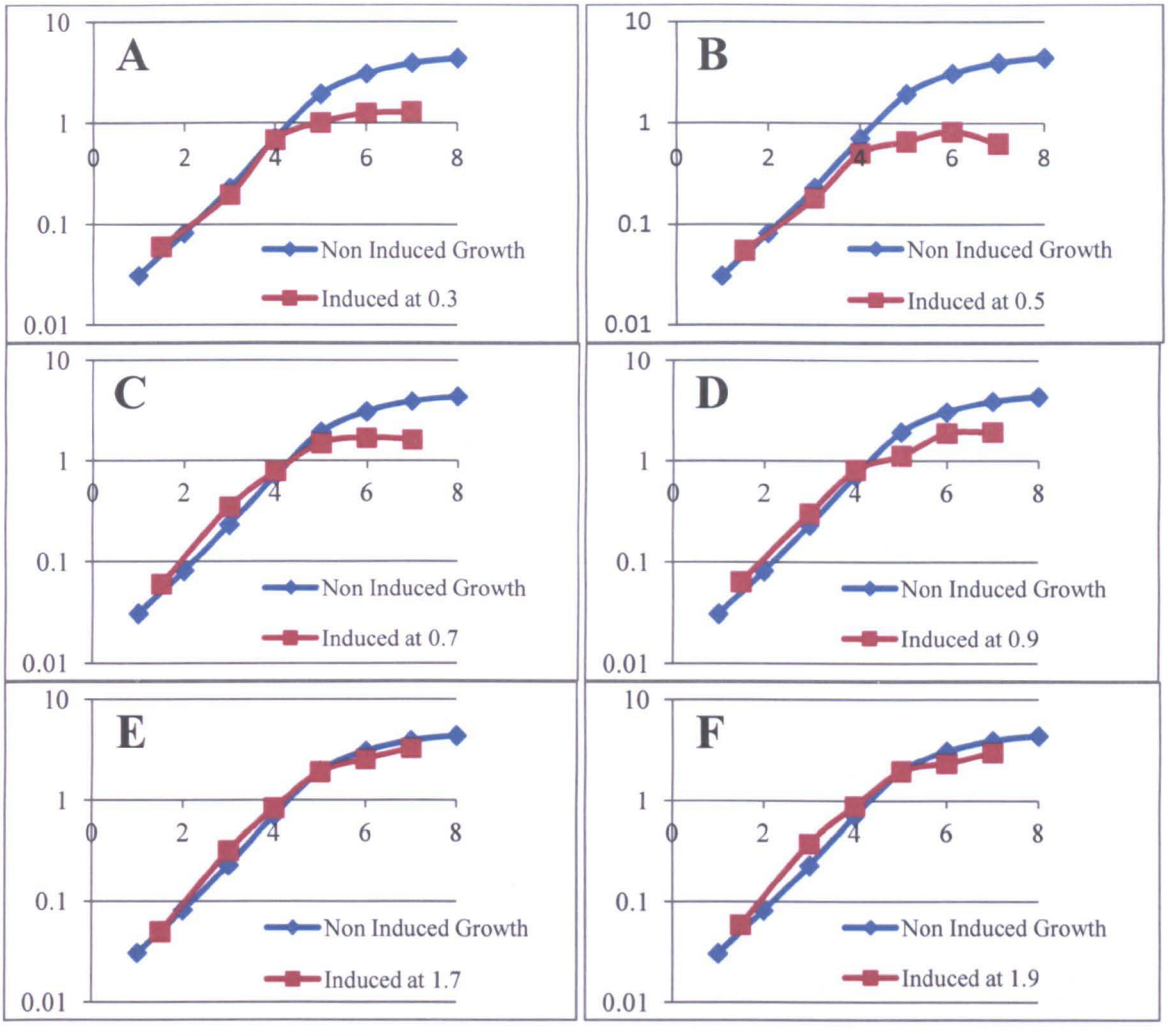


Figure 3.3: Growth curve of non-induced *E. coli* BL 21pLysS compared to *E. coli* cells which are induced at an OD_{600} of A) 0.3, B) 0.5, C) 0.7, D) 0.9, E) 1.7 and F) 1.9. Time in hours is plotted on the x-axis, whilst the absorbance at 600 nm is plotted, on a log scale, on the y-axis.

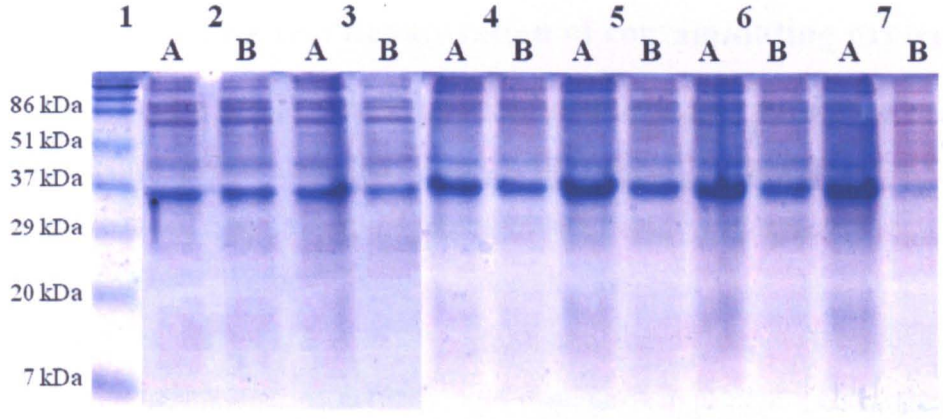


Figure 3.4: SDS PAGE gel of protein present in the A) soluble fraction and B) insoluble fraction following induction at OD_{600} 2) 0.3, 3) 0.5, 4) 0.7, 5) 0.9, 6) 1.7 and 7) 1.9. Lane 1 corresponds to standard molecular weight markers of a known size.

3.3 Cell breakage and protein extraction

Previous work conducted with POR has shown the enzyme to exhibit a self-cleavage property (Dr D.J. Heyes personal communication). Purified POR, when left for a period of time at room temperature, would cleave itself between lysine 250 and lysine 251 producing 31 kDa and 6 kDa fragment, when separated using SDS PAGE gel electrophoresis (Figure 3.5). Studies conducted by Durin indicated that in the presence of EDTA the proteolysis of POR was significantly reduced (Durin 2006). The continuous presence of EDTA however would be a problem since the first step of purification was to pass the cell extract down an immobilised metal ion affinity chromatography column. It was therefore decided that the *E. coli* would be sonicated in the presence of EDTA, which would then be removed before purifying the protein further.

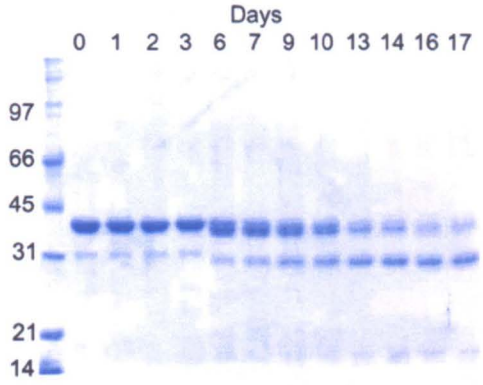


Figure 3.5: SDS PAGE gel showing the levels of degradation of POR following incubation at 25 °C over a period of 17 days. Produced by Dr. D. J. Heyes (University of Manchester).

3.3.1 Sonication and heat denaturation of contaminating proteins

Frozen *E. coli* BL21pLysS cells from 1 litre of growth were thawed and re-suspended in binding buffer plus Roche complete protease inhibitor. The resuspended *E. coli* was sonicated on ice (detailed 2.8). 1.5 ml of sonicated cell extract was removed and placed into a 1.5 ml eppendorf, before being placed into a 35 °C water bath for 15 minutes. Samples were then spun at 13,000 rpm for 30 minutes using a bench-top centrifuge to separate denatured insoluble protein from soluble protein. The supernatant was removed and placed into a fresh eppendorf before being placed back in the water bath at 40 °C, whilst the denatured protein pellet was photographed (Figure 3.6). The process was repeated increasing the temperature of the water bath by 5 °C each time until all protein had

precipitated from solution. Samples of the soluble protein present after each heating step were removed and were analysed using SDS PAGE (detailed 2.19) (Figure 3.7).

As observed in figure 3.7, the levels of all protein present in lanes 3 and 4, corresponding to soluble protein remaining after heating at 35 °C and 40 °C respectively, do not differ that much from those observed in lane 2, corresponding to the soluble protein present after sonication. As the temperature increases to 45 °C, protein bands especially that of POR at 37 kDa are seen to reduce in intensity and continue to do so until 60 °C when there is very little protein remaining. Although figure 3.7 indicates that heating the sample up to 40 °C has no benefit, figure 3.6 clearly shows that heating the sample denatures some material from solution. It was therefore decided to heat the sample at 40 °C following sonication as this would remove a lot of material from solution prior to purification, whilst maintaining maximum levels of POR.

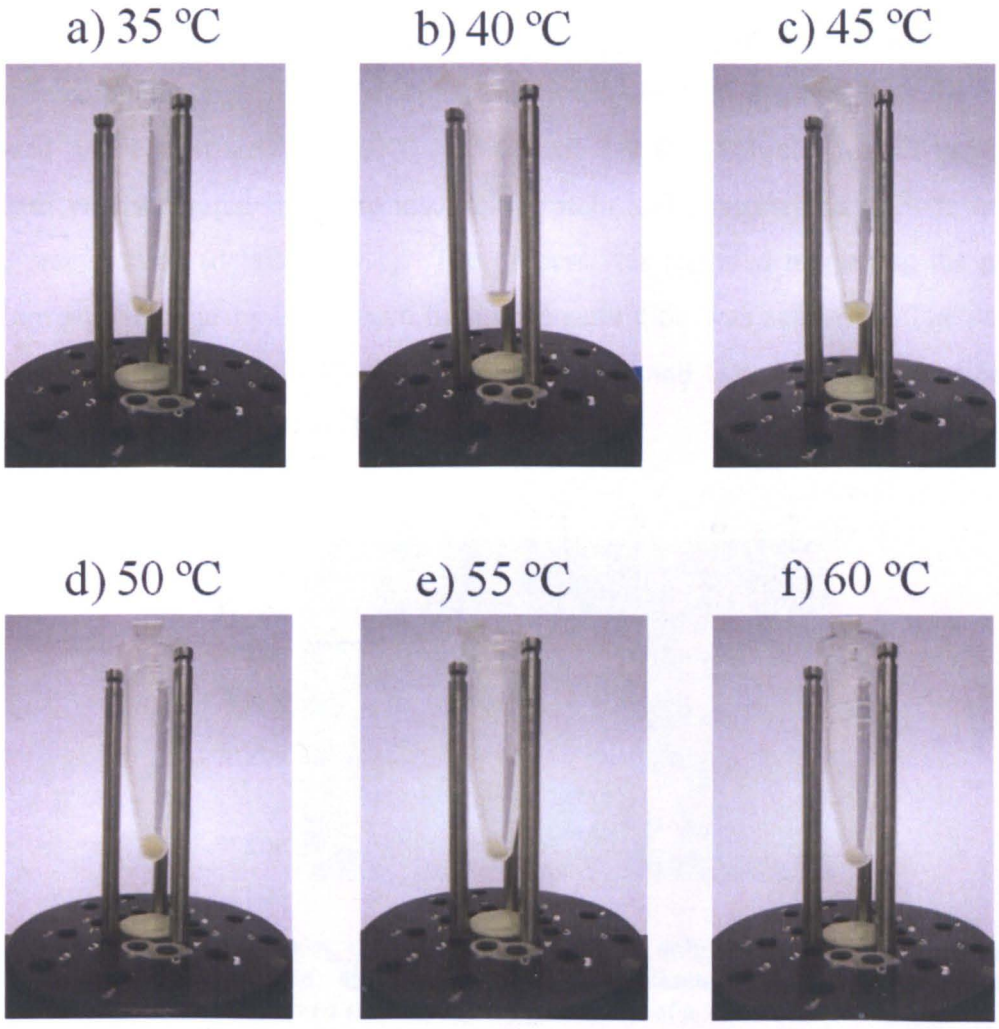


Figure 3.6: Photographs of the thermally denatured protein following incubation at a) 35 °C, b) 40 °C, c) 45 °C, d) 50 °C, e) 55 °C and f) 60 °C.

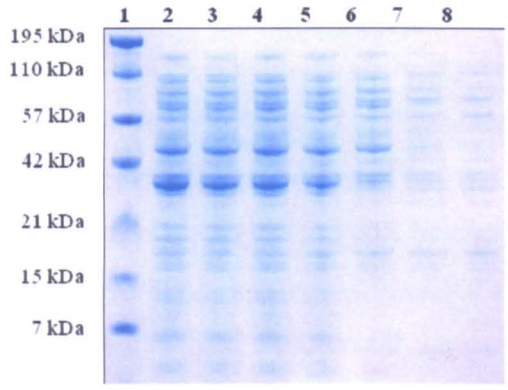


Figure 3.7: SDS PAGE gel showing the levels of soluble protein present in the sample 2) before incubation and remaining following incubation at 3) 35 °C, 4) 40 °C, 5) 45 °C, 6) 50 °C, 7) 55 °C and 8) 60 °C. Lane 1 corresponds to standard molecular weight markers of a known size.

3.3.2 Ammonium Sulphate Precipitation

E. coli BL21pLysS cells from 1 litre of growth were re-suspended in binding buffer, with Roche complete protease inhibitor, and sonicated. Following sonication, the cell extract was heated and centrifuged at 25,000 rpm (detailed 2.8). Soluble protein present in the supernatant was separated from the insoluble protein and subjected to a 20 % ammonium sulphate precipitation (detailed 2.9.1). The process was repeated increasing the percentage ammonium sulphate cut by 20 % each time until saturation was achieved. The precipitated proteins were re-suspended in 10 ml of binding buffer and samples were taken to separate using SDS PAGE (detailed 2.19) (Figure 3.8).

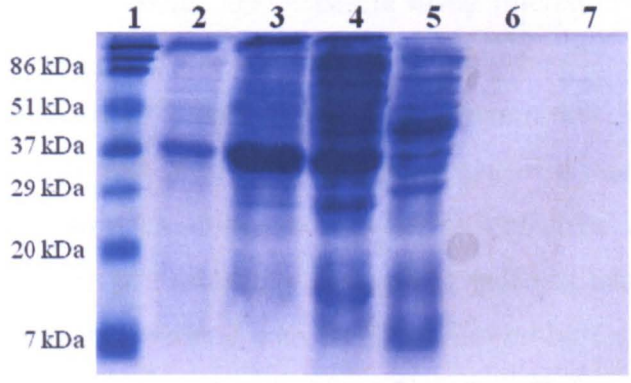


Figure 3.8: SDS PAGE gel showing the proteins remaining in solution following a 2) 20 %, 3) 40 %, 4) 60 %, 5) 60 %, 6) 80 % and 7) 100 % ammonium sulphate precipitation. Lane 1 corresponds to standard molecular weight markers of a known size.

A protein corresponding to a molecular weight of 37 kDa can be observed in the lanes corresponding to the 20, 40, 60 and 80 % precipitations in figure 3.8. There are no protein bands in the lanes corresponding to the 100 % cut and the saturated supernatant indicating that all protein in the sample has precipitated by 80 % saturation. As detailed previously, this step was added to the protocol so that we could essentially separate the protein away from an environment containing EDTA. This stage was never intended to be used as a purification step so, although the majority of protein with a molecular weight of 37 kDa precipitates out between 40 and 60 % saturation, it was decided to perform an 80 % precipitation, to precipitate all protein, then purify POR away from its contaminants by other chromatographic techniques.

3.4 Optimisation of the purification of His_6-POR

The protocol provided by Dr. D. Heyes (University of Manchester) involved purifying His_6-POR using a Talon column followed by a blue sepharose chromatography column. Before starting to optimise the purification protocol it was decided to see how good the existing purification protocol was.

3.4.1 Purification of His_6-POR using Talon resin

Ammonium sulphate pellets (detailed 2.9.1) were thawed and re-suspended in binding buffer with protease inhibitor without EDTA before being filtered through a 0.45 μm filter (Millipore). Talon resin (Clontech) was equilibrated in binding buffer and incubated with the re-suspended protein pellet at 4 °C with agitation. The resin was loaded into a Bio-Rad chromatography column and packed down using binding buffer at a rate of 1 ml / min. The column was connected to an Äkta Prime chromatography machine and was washed with 50 ml of wash buffer at a rate of 2 ml/min to elute off all proteins bound non-specifically to the resin. The column was then washed with 50 ml of elution buffer at a rate of 2 ml/min. The UV absorbance at 280 nm was measured of all the eluant (Figure 3.9) and samples from all fractions of interest were run out using SDS PAGE.

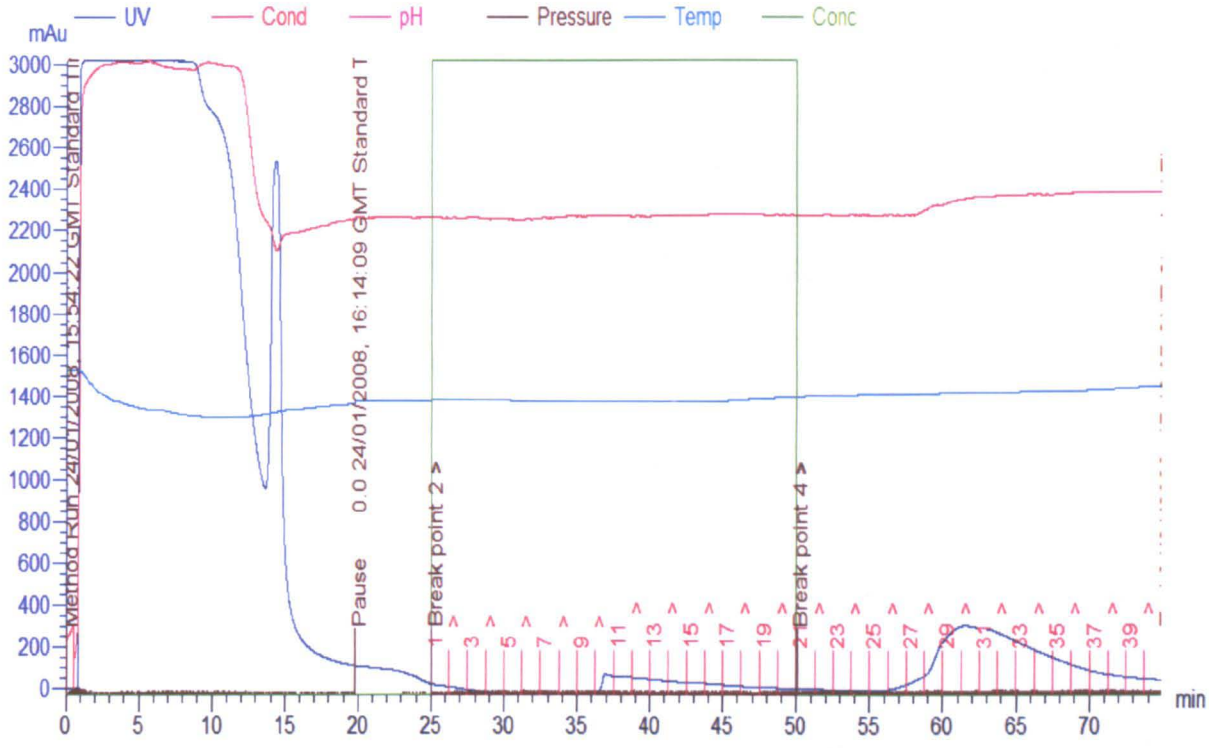


Figure 3.9: Elution profile showing the change in absorbance at 280 nm of the eluant from the Talon column. Red numbered lines appearing from 25 minutes onwards correspond to fractions collected.

As shown in figure 3.9 a significant number of non-specifically bound proteins were washed from the column during the initial wash phase. However when the elution buffer was applied to the column, a very small quantity of protein was eluted. The quantity of the protein present in these fractions was too small to be detected by SDS PAGE. The two reasons for this small yield of protein may be either because the protein didn't initially bind to the column or that it has bound so strongly that 250 mM imidazole does not remove it. The column was washed with a higher concentration of imidazole (50 mM sodium phosphate pH 7, 600 mM NaCl, 500 mM imidazole) yet still no protein of real significance was eluted from the column. This implies that the protein did not initially bind to the resin, which is confirmed by figure 3.10, which clearly shows a distinct band at 37 kDa in the lane corresponding to the protein bound non-specifically which was eluted in the initial wash step.

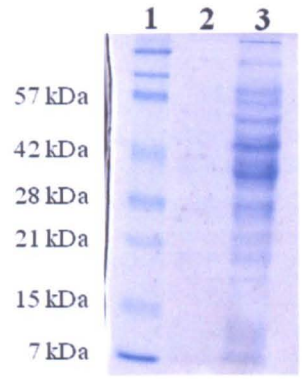


Figure 3.10: SDS PAGE gel of 2) fractions 29 to 33 combined and concentrated along with 3) protein which was eluted from the column in the initial wash phase. Lane 1 corresponds to standard molecular weight markers of a known size.

It is known from sequencing data that the hexa-his tag is present at the N-terminus of the protein so it was thought that the protein did not bind to the resin due to the amendments made to the protocol or due to an issue with the buffers or resin. Instead of wasting more time and trying to identify what the problem was, it was decided to try a more conventional nickel IMAC chromatography column instead.

3.4.2 Purification of His_6-POR using Nickel spin columns

Ammonium sulphate pellets (detailed 2.9.1) were thawed and re-suspended in 20 ml low salt buffer with protease inhibitor minus EDTA before being filtered through a 0.45 μm filter (Millipore). Nickel spin columns were set up as detailed by the manufacturers' instructions and the samples were applied to the columns at 2,000 rpm. The sample application procedure was repeated four times to ensure maximum binding of POR to the resin. 20 ml of 10 mM imidazole wash buffer was applied to the column and spun at 3,000 rpm. The wash process was repeated with increasing concentrations of imidazole, 70 mM, 90 mM and 110 mM before the remaining proteins were eluted with 40 ml of 500 mM imidazole elution buffer. The process was repeated applying the initial protein sample to the column, before being washed and eluted as detailed above. Samples were taken of each sample and were analysed using SDS-PAGE (detailed 2.19).

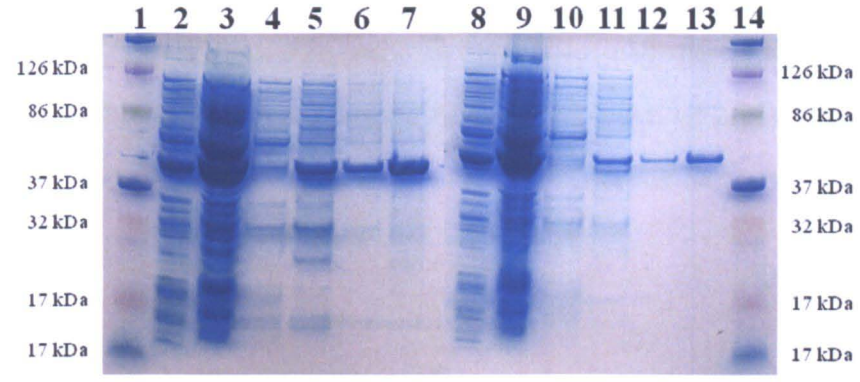


Figure 3.11: SDS PAGE gel of proteins eluted from the Ni spin columns following two rounds of application and elution, lanes 2-7 and 8-13 respectively. Lanes 2 & 8 correspond to samples of the initial protein loaded onto the column, whilst 3 & 9, 4 & 10, 5 & 11, 6 & 12 and 7 & 13 correspond to washes with elution buffers containing 10 mM, 70 mM, 90 mM, 110 mM and 500 mM imidazole. Lanes 1 and 14 correspond to standard molecular weight markers of a known size.

It is observable from Figure 3.11 that a significant proportion of POR does not bind to the column, or binds transiently so is washed off during the initial wash phase. There are also a number of contaminating proteins which co-purify with POR in the 90, 110 and 500 mM wash fractions, far more than what is normally expected from purification of a His-tagged protein. Figure 3.11 also shows that more protein is purified following the second application of the protein sample to the spin column. Although the binding capacity of the column is approximately 60 mg, this result implies that the amount of protein produced and binding to the column is greater than the capacity of the spin column. As a result it was decided to try using a gravity fed Nickel column where the protein can be washed and eluted from the column using an imidazole gradient using a resin which has a larger binding capacity than the spin columns.

3.4.3 Purification of His_6-POR using a Nickel Gravity column

5 ml of Nickel His-60 resin (Generon) was loaded into a glass chromatography column (G.E. Healthcare) and equilibrated with binding buffer. Ammonium sulphate pellets (detailed 2.9.1) were thawed and each re-suspended in 15 ml of binding buffer with protease inhibitor without EDTA. The re-suspended protein pellet was filtered through a 0.45 μm filter (Millipore) before being loaded onto the column at a rate of 0.5 ml/min using a peristaltic pump. The column was connected to an Äkta Prime chromatography machine and

the protein was eluted off the column at a rate of 1 ml/min using a 100ml, 1M imidazole gradient. The UV absorbance at 280 nm was measured of all the eluant (Figure 3.11) and samples from all fractions were run out using SDS PAGE (detailed 2.19) (Figure 3.12).

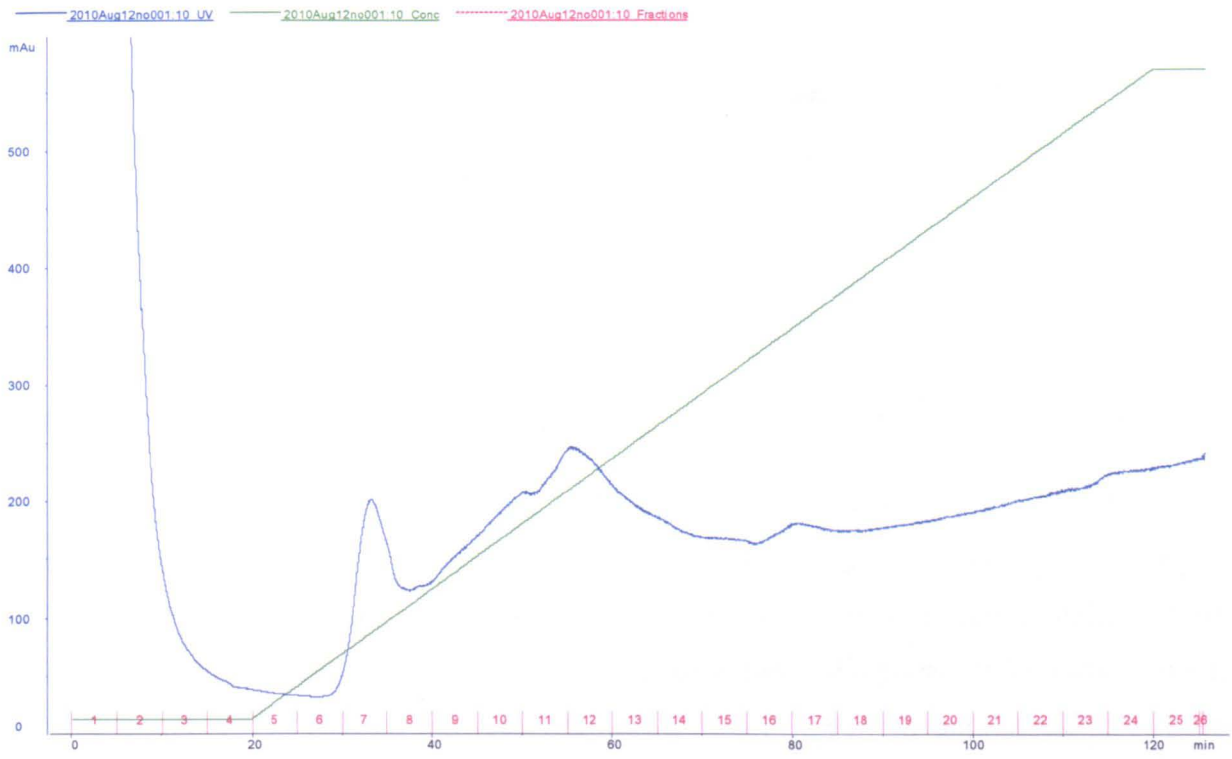


Figure 3.12: Elution profile showing the change in absorbance at 280 nm of the eluant from the Ni-His60 column.

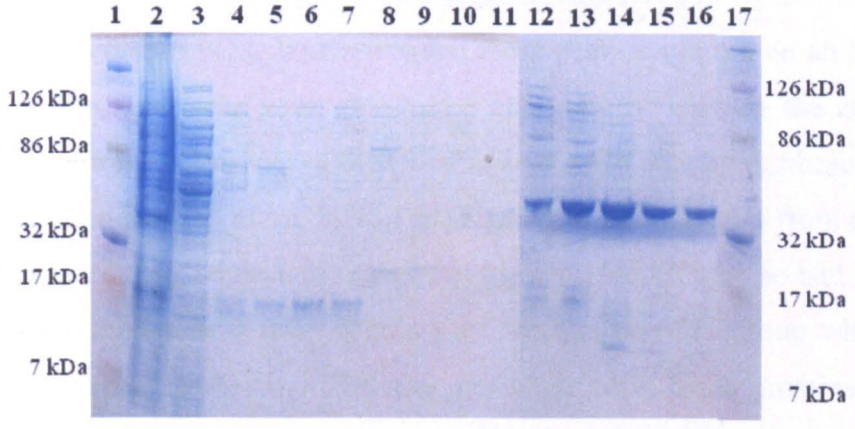


Figure 3.13: SDS PAGE gel showing the proteins present in the 2) loaded fraction, along with those eluted from the column in 3) initial flow through and fractions 4) 1, 5) 2, 6) 3, 7) 4, 8) 5, 9) 6, 10) 7, 11) 8, 12) 9, 13) 10, 14) 12, 15) 13, 16) 14. Lanes 1 and 17 correspond to standard molecular weight markers of a known size.

It is observable in Figure 3.13 that during the initial wash phase, all the unbound protein is washed off the column. Following the application of the imidazole gradient to the column, two protein peaks are observed to elute from the column. Figure 3.13 shows that the first peak, which starts to elute from the column at an imidazole concentration of 100 mM, corresponds to contaminating proteins which have bound to the column. The second peak which starts to elute from the column at a concentration of 200 mM imidazole, contains POR however there is some contamination from the contaminating proteins found in the first peak, which have not finished eluting from the column. If a second wash phase is put in at 180 mM imidazole then this would ensure that all contaminating proteins were washed from the column prior to POR being eluted.

3.4.4 Purification of His_6-POR using blue sepharose resin

Blue sepharose 6 resin (GE Healthcare) was prepared as detailed in the manufacturers' instructions, and loaded into a fast flow glass chromatography column (GE Healthcare). Fractions eluted from the Ni His-60 column, which contained POR (detailed 3.4.3), were loaded onto the blue sepharose column at a rate of 0.5 ml/min using a peristaltic pump. The column was connected to an Äkta Prime purification machine and the protein eluted from it at a rate of 1 ml / min using a 100 ml 1 M, NaCl gradient. The UV absorbance at 280 nm was measured of all the eluant (Figure 3.14) and samples from all fractions were run out using SDS PAGE (detailed 2.19) (Figure 3.15).

Following the wash steps, two protein peaks are observed in figure 3.14. These were initially thought to contain POR, however when these peaks were run on an SDS PAGE gel (Figure 3.15) no protein peaks were observed. This may be because the concentration of POR was too low for the sensitivity of the Coomassie stain to pick up or because POR did not bind to the column in the first place. SDS PAGE gels run with samples from all stages of the purification indicated that the protein was present in the void volume so had never bound to the column. Blue sepharose 6 resin contains a Cibacron blue 3G group which is attached covalently to the sepharose beads. The blue sepharose resin binds proteins which require adenyl-containing cofactors, however if the adenyl binding site on the protein is occluded then binding of the protein to the resin would be significantly hindered. Variability between batches of Blue Sepharose resin purchased seems to be a common problem and due to the

expense and importance of some of the protein samples produced, it was decided to replace this step with a more conventional ion exchange chromatography step.

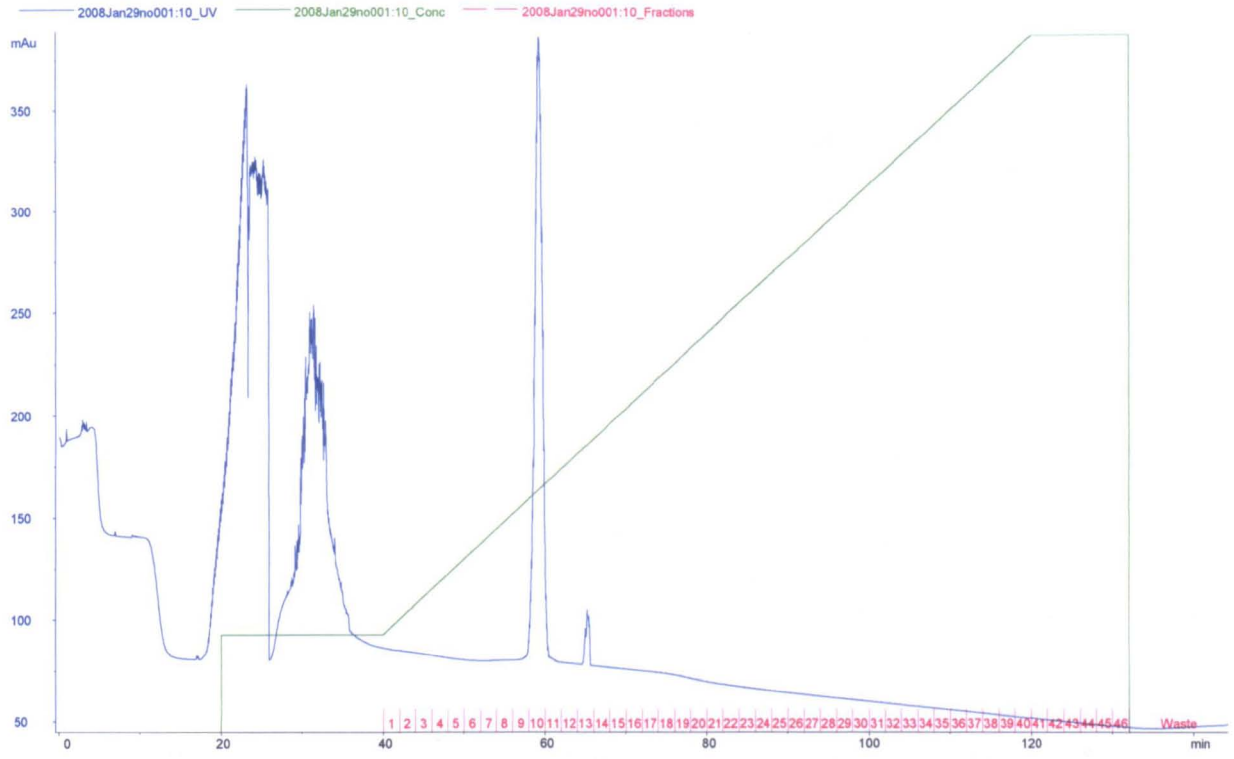


Figure 3.14: Elution profile showing the change in absorbance at 280 nm of the eluant from the Blue Sepharose column.

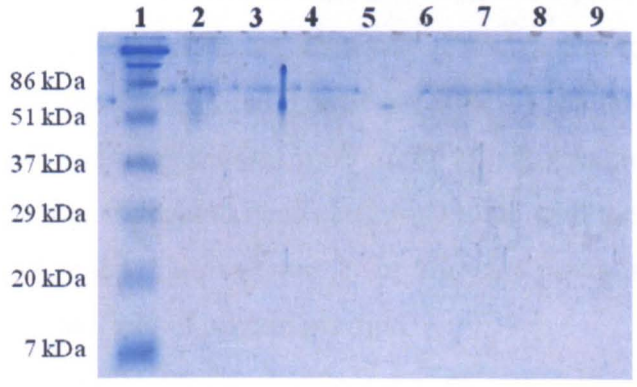


Figure 3.15: SDS PAGE gel showing the proteins present in 2) peak one of the initial wash concentrated, 3) Peak 2 of the initial wash concentrated, and fractions 4) 8, 5) 9, 6) 10, 7) 11, 8) 12 and 9) 13. Lane 1 corresponds to standard molecular weight markers of a known size.

3.4.5 Purification of His_6-POR using a Q sepharose column

The estimated pI of POR was calculated to be approximately 9.4 (Gasteiger E. *et al.* 2005). This is relatively high compared to that of most proteins; therefore at pH 7.4 the protein would be positively charged and will therefore not bind to an anion exchange chromatography column. If all contaminating proteins were negatively charged at the pH then they would bind to the column thus purifying POR further. Since POR would not be binding to the column it was necessary to concentrate the POR containing fractions from the Nickel column prior to loading the sample onto the column. 20 ml of Q sepharose resin (GE Healthcare) was prepared according to the manufacturers' instructions before being loading into a glass chromatography column (GE Healthcare). Fractions from the Ni His-60 column (detailed 3.4.3), which contained POR were concentrated up to 5 ml, using an Amicon stirred cell concentrator, with a 10 kDa membrane (detailed 2.10), before being loaded onto the column at a rate of 0.5 ml/min using a peristaltic pump. The column was connected to an Äkta Prime purification machine and the protein eluted from it at a rate of 1 ml / min using low salt buffer.

In Figure 3.16, a clear protein peak can be observed to elute from the column in fractions 1 and 2. When analysed by SDS PAGE (Figure 3.17) the protein peak was detected to contain POR, however a number of contaminating proteins were still present in the sample. It was therefore decided to purify the sample using cation exchange instead. The contaminating proteins which bound to the Q sepharose resin, would not bind to SP sepharose resin so would elute in the void volume. Whilst the protein which bound to the column would bind with different affinities, so could be purified further on a salt gradient. Also whilst concentrating the sample up to load onto the Q sepharose resin, the sample appeared to aggregate and significant losses were observed. It was unknown as to why this had happened, however it was thought to have something to do with the high concentration of imidazole present. By removing this step from the purification protocol it was hoped that we should be able to increase the yield of protein purified.

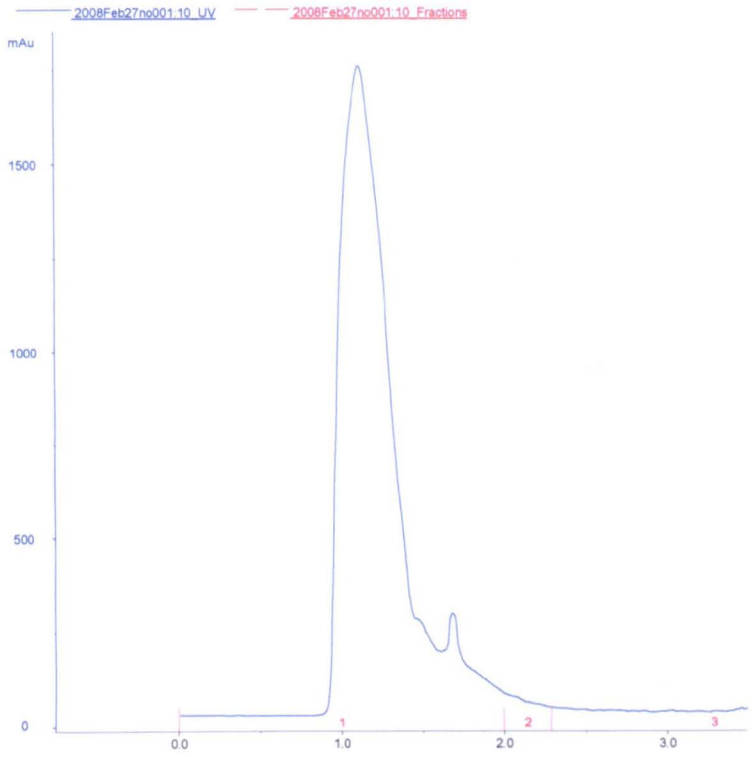


Figure 3.16: Elution profile showing the change in absorbance at 280 nm of the eluant from the Q-sepharose column. The x-axis corresponds to time in minutes.

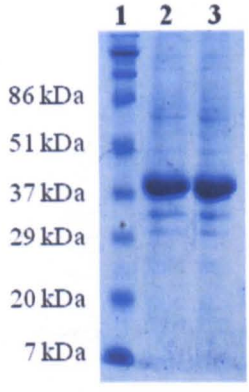


Figure 3.17: SDS PAGE gel of proteins eluted from the Q sepharose column in fractions 2) 1, 3) 2, and 4) 3. Lane 1 corresponds to standard molecular weight markers of a known size.

3.4.6 Purification of His_6-POR using an SP sepharose column

20 ml of SP sepharose resin (GE Healthcare) was prepared according to the manufacturers' instructions prior to loading into a glass chromatography column (GE Healthcare). Fractions eluted from the Ni His-60 column, which contained POR

(detailed 3.4.3), were loaded onto the SP sepharose column at a rate of 0.5 ml/min using a peristaltic pump. The column was connected to an Äkta Prime purification machine and the protein eluted from it at a rate of 1 ml / min using a 100ml 2 M NaCl gradient. The UV absorbance at 280 nm was measured of all the eluant (Figure 3.18) and samples from all fractions were run out using SDS PAGE (Figure 3.19). Following elution from the column, 1 ml of a 10 ml stock of low salt buffer, containing 1 complete protease inhibitor tablet (Roche) and 100 mM DTT, was placed into each fraction containing POR to stabilise it and prevent degradation.

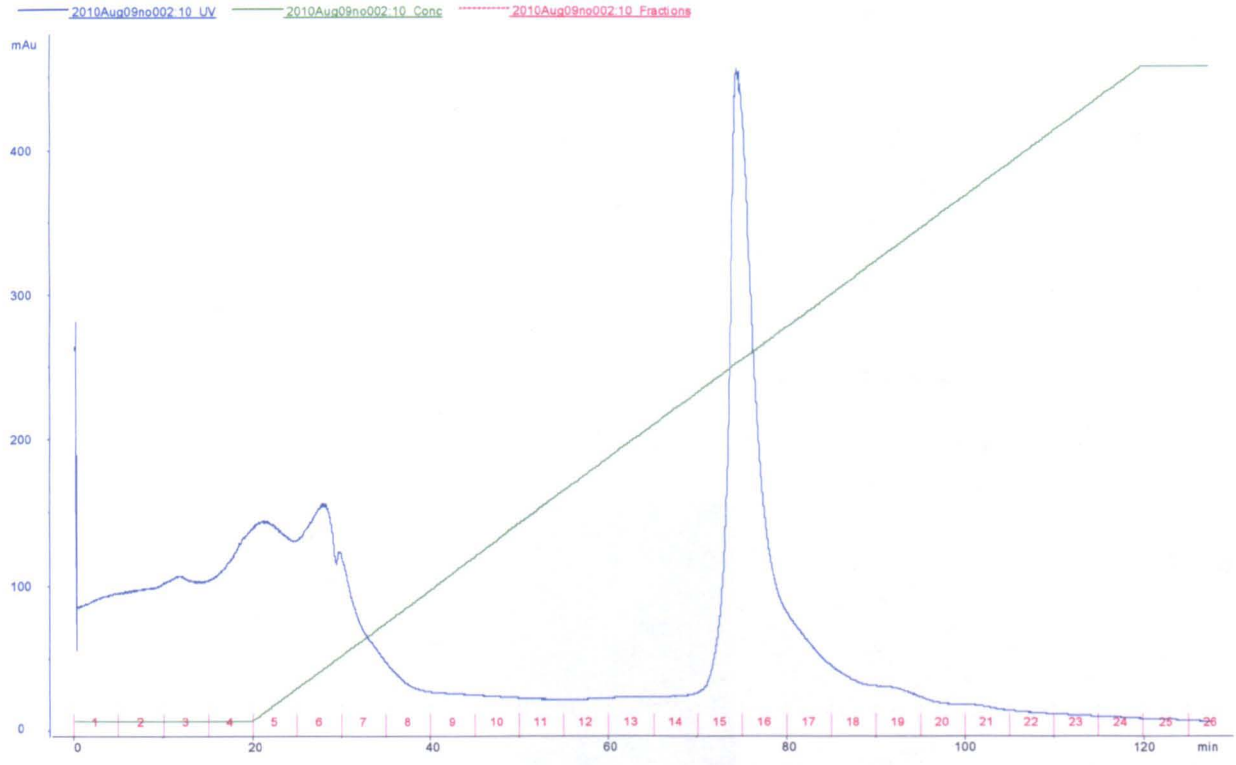


Figure 3.18: Elution profile showing the change in absorbance at 280 nm of the eluant from the SP-sepharose column.

Following the initial wash step, a clear protein peak is observed in Figure 3.18 to elute from the column at a salt concentration of 1M NaCl. A second peak is seen to elute from the column at a concentration of 1.4 M NaCl, however SDS PAGE (Figure 3.19) confirmed the presence of POR in both peaks and the level of purity observed was estimated to be greater than 95 % at this stage. Upon the addition of the reducing stock solution to the POR containing fractions a brown precipitate (Figure 3.20) was observed in fractions 18 to 20 corresponding to the second peak (Figure 3.18). This result was replicated independently by

adding 1 ml of 100 mM $NiSO_4$ solution to 0.1 ml of 1 M EDTA and 0.1 ml of 1 M DTT, so the brown precipitate was attributed to being a nickel DTT salt. Having already confirmed the presence of POR in these fractions (Figure 3.19), the nickel present was attributed to have co-purified with the protein. The obvious place to attribute the binding of the nickel is to the His tag of the protein however, POR does bind a porphyrin ring which has a co-ordinating magnesium ion. It may therefore be possible that the nickel has bound in the Pchlide binding site.

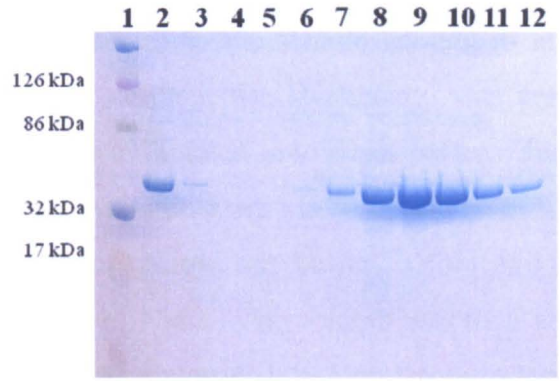


Figure 3.19: SDS PAGE gel of the protein present in the 2) initial loaded fraction along with protein eluted from the column in fractions 3) 6, 4) 12, 5) 13, 6) 14, 7) 15, 8) 16, 9) 17, 10) 18, 11) 19 and 12) 20. Lane 1 corresponds to standard molecular weight markers of a known size.

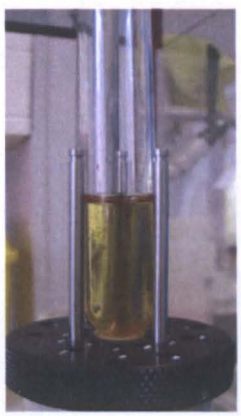


Figure 3.20: Photograph of the brown precipitate formed in fraction 19 following the addition of 1 ml of reducing solution containing 100 mM DTT and 1 mM EDTA.

The presence of the nickel attached to the protein is enough to differentiate the two protein species as independent species and as a result they do purify separately. If a larger resin volume is used (40 ml) then this should allow the column to act as a gel filtration column thus separating the two species so they elute at different times.

3.4.7 Purification of His_6-POR using Mono-S column

Although the purity of the protein following purification through the nickel and ion exchange column was estimated to be greater than 95 %, it was decided to try and purify it further. That way if crystal trials would be laid down again with POR, the protocol would have already been developed.

Following purification from the SP sepharose column, fractions which contained POR (detailed 3.4.6) were pooled together and diluted two fold with low salt buffer. This would reduce the salt concentration present in the sample enough to allow POR to bind to the Mono-S column. The Mono-S column (GE Healthcare) was prepared as detailed by the manufacturers' instructions and equilibrated in low salt buffer. The column was attached to an Äkta FPLC before loading the sample on, via a superloop, at a rate of 0.5 ml/min. The column was washed with 10 ml of low salt buffer, before 190 ml gradient was applied between 400 mM NaCl and 1.7 M NaCl. The column was then washed with 10 ml of 2 M NaCl to ensure all proteins had been removed. 0.5 ml fractions were collected and analysed by SDS PAGE (detailed 2.19) to assess purity.

Following the initial wash phase where contaminating proteins were eluted from the column (Figure 3.22), a protein peak is observed to elute from the Mono-S column at a salt concentration of 500 mM NaCl (Figure 3.21). Analysis by SDS PAGE (Figure 3.22) showed this protein to correspond to POR and the purity at this stage was estimated to be greater than 99 %.

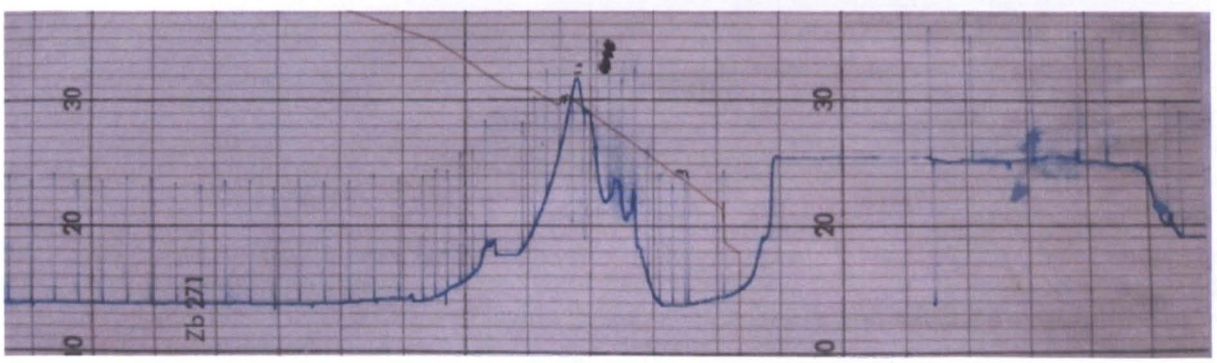


Figure 3.21: Elution profile showing the change in absorbance at 280 nm of the eluant from the Mono-S column. Blue vertical lines correspond to 0.5 ml fractions collected, with fraction 1 on the far right of the trace, and subsequent fractions following to the left.

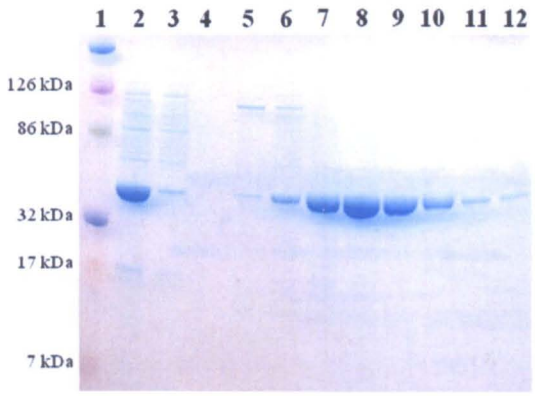


Figure 3.22: SDS PAGE gel showing the proteins present in the 2) initial loaded fraction along with protein present in fractions 3) 6, 4) 11, 5) 12 6) 14, 7) 16, 8) 19, 9) 22, 10) 25, 11) 28 12) 31, eluted from the Mono-S ion exchange column. Lane 1 corresponds to standard molecular weight markers of a known size.

3.4.8 Purification of His_6-POR using Gel filtration

Following purification using a Mono-S column, the fractions detected to contain POR were combined and concentrated down to a volume of 1 ml using an Amicon stirred cell (detailed 2.10). A pre-equilibrated gel filtration column (GE Healthcare) was attached to an Äkta Prime protein purification machine and the sample was applied using a super-loop at a rate of 0.5 ml / min. The gel filtration column was then run at 1 ml / min with 0.5 ml fractions being collected, which were analysed by SDS PAGE (detailed 2.19).

Figure 3.23: Elution profile showing the change in absorbance at 280 nm of the eluant from the Gel-Filtration column. Blue vertical lines correspond to 0.5 ml fractions collected, with fraction 1 on the far right of the trace, and subsequent fractions following to the left.

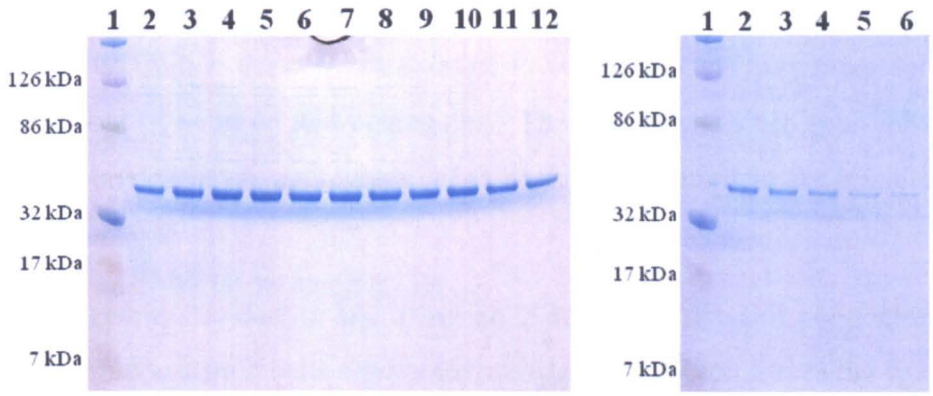


Figure 3.24: SDS PAGE gel showing A) the proteins present in the A) fractions 85 to 89, corresponding to the high molecular weight peak and B) fractions 91 to 101, corresponding to the lower molecular weight peak eluted from the gel filtration column. Lane 1 in both gels corresponds to standard molecular weight markers of a known size.

Two mono-dispersed protein peaks were observed to elute from the gel filtration column (Figure 3.23) both of which were observed to contain POR (Figure 3.24). Following calibration of the gel filtration column, the first peak was noted to correspond to monomeric POR, whilst the second peak eluted at a volume which corresponded to dimeric POR. This was a surprise as although all other members of the SDR family are documented to exist as dimers or even larger oligomeric units (Oppermann *et al.* 2003), all previous work conducted on POR has suggested that POR is present as a monomer.

To see if the dimerisation of POR was concentration dependant the gel filtration column was run again, loading a sample which was half the concentration of that previously applied. Following gel filtration however two protein peaks were observed to elute from the column and were present in exactly the same ratio as that previously seen with the higher concentration. Since the proportion of dimer made up less than 5 % of the sample it was decided that this would not affect the NMR measurements.

3.5 Concentration of His_6-POR

The initial protocol provided used a vivaspin column with a 30 kDa cut off filter spun at 30,000 rpm, to concentrate the POR sample. Whilst concentrating POR in a vivaspin, even though the volume of the sample was decreasing the concentration of POR did not seem to increase. Whilst using a vivaspin, the set up of the concentrator produces a concentration

gradient across the sample, which is detrimental if the protein of interest has a relatively low maximum solubility. When the concentration of POR exceeds the maximum solubility, the POR precipitates out of solution and aggregates. These aggregates are spun down onto the concentrator membrane but the aggregates act as seeds and precipitate the remaining protein from the sample.

It was therefore decided to try using an Amicon stirred cell concentrator instead (detailed 2.10). Amicon stirred cells work under nitrogen gas which forces the solute through the 10 kDa PES membrane. They also have a magnetic stirrer bar inside them which stirs the sample and prevents a concentration gradient occurring across the sample. Whilst using the Amicon stirred cell the amount of protein lost was significantly reduced, compared to when the protein was concentrated in a vivaspin. However once the concentration of the sample reached 0.3 mM, aggregates as previously observed in the vivaspin, appeared and similar losses to those previously seen occurred. It is therefore fair to assume that the maximum solubility of POR *in vitro* is 0.3 mM.

3.6 Activity assay using His_6-POR

As previously detailed, in the presence of NADPH and light, POR catalyses the reduction of a double bond in the D ring of Pchlide forming Chlide (Lebedev *et al.* 1998). This can be monitored spectroscopically as both Pchlide and Chlide absorb at distinct wavelengths. Reaction assays containing 10 μM Pchlide, 10 μM NADPH, 50 mM BOG and 10 mM DTT were set up (detailed 2.15). The samples were placed in a Cary 50 spectrophotometer and incubated at 55 °C for 15 minutes prior to addition of the enzyme. 20 μM POR was added to the reaction and spectra were acquired in the dark between 600 nm and 720 nm. The sample was then illuminated using a KL 1500 electronic fibre optic light source with a 400-500 nm filter attached, whilst spectra continued to be taken (Figure 3.19).

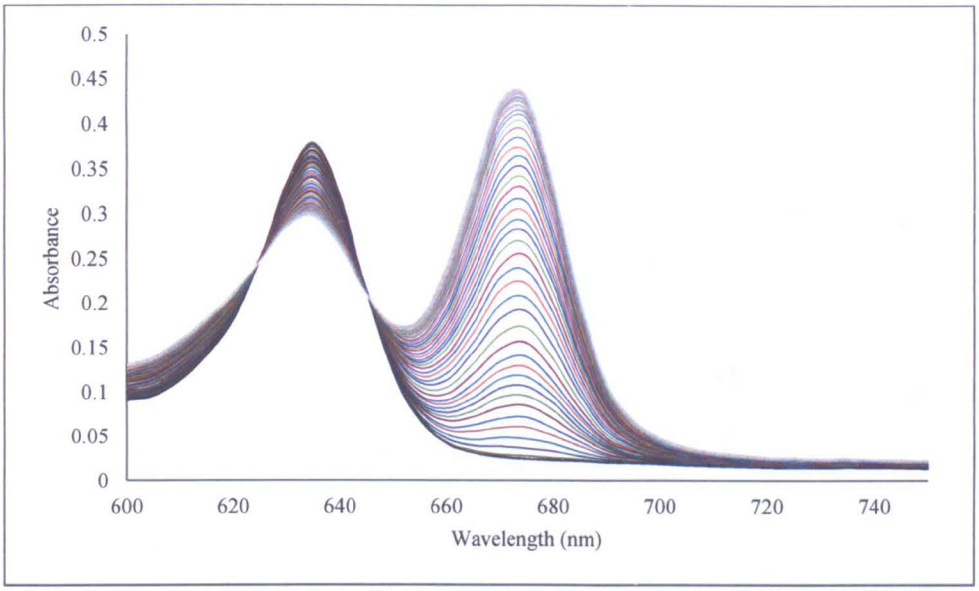


Figure 3.25: POR assay. Upon illumination, the peak at 635 nm, corresponding to Pchlide, decreases whilst the peak at 670 nm, corresponding to Chlide, increases.

Prior to illumination the Pchlide absorbs at 635 nm, however following the illumination of the sample, a decrease in intensity of the Pchlide peak is observed, along with the increase in intensity of a peak at 670 nm, corresponding to Chlide. A perfect isobathic point is prevented from forming however due to the degradation of Chlide under high light intensity. After three minutes under light saturation conditions, the reaction has reached completion. This not only confirms the purified protein to be POR but also confirms that the purified protein is active.

3.7 Production of isotopically labelled protein

In order to analyse POR using NMR, labelled protein must be grown where ^{15}N ammonium sulphate is used as the sole nitrogen source, and ^{13}C glucose is used as the sole carbon source. This process ensures that all nitrogen and carbon atoms are 100 % isotopically labelled. It was previously important to ensure that the protocol developed allowed the maximum amount of POR to be over-expressed and purified, as the incorporation of labels and the growth in minimal media is noted to affect the final concentration of *E. coli* thus producing less protein.

3.7.1 Over-expression of labelled proteins

E. coli BL21pLysS cells containing the pET9His *Thermosynechococcus* POR plasmid were plated out on LB agar plates (detailed A.1.1), containing the relevant selection, and incubated at 37 °C overnight (detailed 2.5.1). A single colony was picked and inoculated in 100 ml of M9 minimal media containing ^{15}N labelled ammonium sulphate and either $^{12}C_6$ D-glucose or $^{13}C_6$ D-glucose (detailed A.1.3), depending on whether a single or double labelled growth was to be conducted and was incubated overnight at 37 °C (detailed 2.6.1).

1 ml of overnight starter culture was placed into 500 ml of M9 minimal media containing the relevant labels and was incubated at 37 °C (detailed 2.6.3). Following 4 hours of exponential growth the *E. coli* were induced with 0.1 mM IPTG and incubated for a further 2 hours at 25 °C (detailed 2.6.5). The *E. coli* were harvested by centrifuging at 10, 000 rpm at 4 °C (detailed 2.7) and were re-suspended in 10 ml of binding buffer prior to being stored at -20 °C.

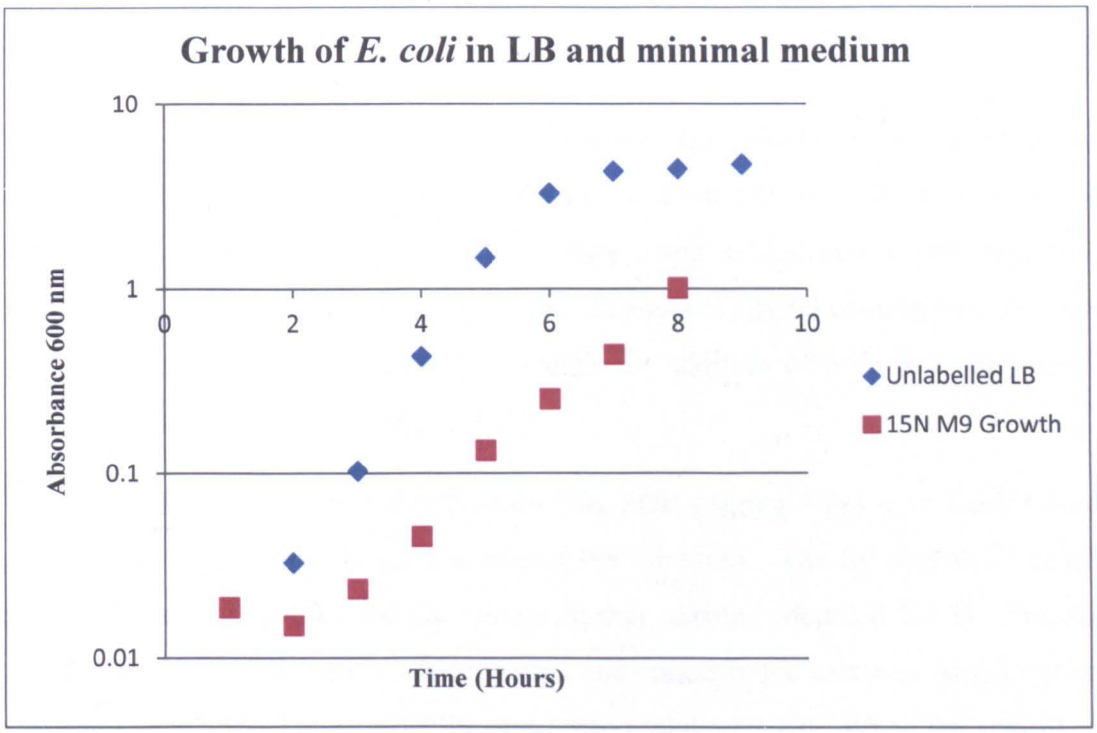


Figure 3.26: Growth curve of un-induced *E. coli* BL21pLysS grown in unlabelled LB medium and ^{15}N labelled M9 minimal medium, plotted on a log scale.

Since the *E. coli* cells were growing slower in the labelled medium it would not be possible to induce the cells at an OD_{600} of 1.7. Using the observation that *E. coli* grow for a maximum of 7 hours before reaching stationary phase (Figure 3.1), it was decided to let the cells grow for 4.5 hours, the time normally taken for *E. coli* to reach an OD_{600} of 1.7 in LB medium, prior to induction with 0.1 mM IPTG. The cells were then incubated at 25 °C for a further two hours (detailed 2.6.4) to allow protein expression.

3.7.2 Purification of labelled proteins

Cell pellets containing isotopically labelled protein (detailed 3.7.1) were defrosted at room temperature prior to being sonicated. The sonicated samples were incubated at 40 °C for 15 minutes before being centrifuged at 24,500 rpm for 30 minutes at 4 °C (detailed 2.8). The supernatant was removed from the samples and an 85 % ammonium sulphate precipitation was conducted, whilst the pellets were discarded. The ammonium sulphate precipitation was spun down at 20,000 rpm for 30 minutes at 4 °C (detailed 2.9.1) and the supernatant was discarded. Precipitated protein could be stably stored at -20 °C awaiting further purification.

Protein precipitates were defrosted and each was resuspended in 10 ml of binding buffer, containing Roche complete protease inhibitor without EDTA. The protein samples were filtered through a 0.45 μm syringe filter before being loaded onto a pre-equilibrated nickel-His 60 IMAC column a rate of 0.5 ml / min. The nickel His-60 column was connected to an Äkta prime and purified (detailed 2.9.2) with the addition of a second wash step at 180 mM imidazole (detailed 3.4.3) (Figure 3.27).

Fractions which were detected to contain His_6-POR (Figure 3.28) were loaded onto a pre-equilibrated SP sepharose column at a rate of 0.5 ml / min. The SP sepharose column was connected to an Äkta prime and the protein further purified (detailed 2.9.3). Fractions detected to contain His_6-POR were pooled together and concentrated using an Amicon stirred cell concentrator with a 10 kDa cut off PES membrane (detailed 2.10). When the sample was concentrated to a volume of 1 ml, a further 10 ml of NMR buffer (detailed A.3.10) was added to the stirred cell and the sample was once again concentrated until a concentration of 0.3 mM was achieved (detailed 2.11).

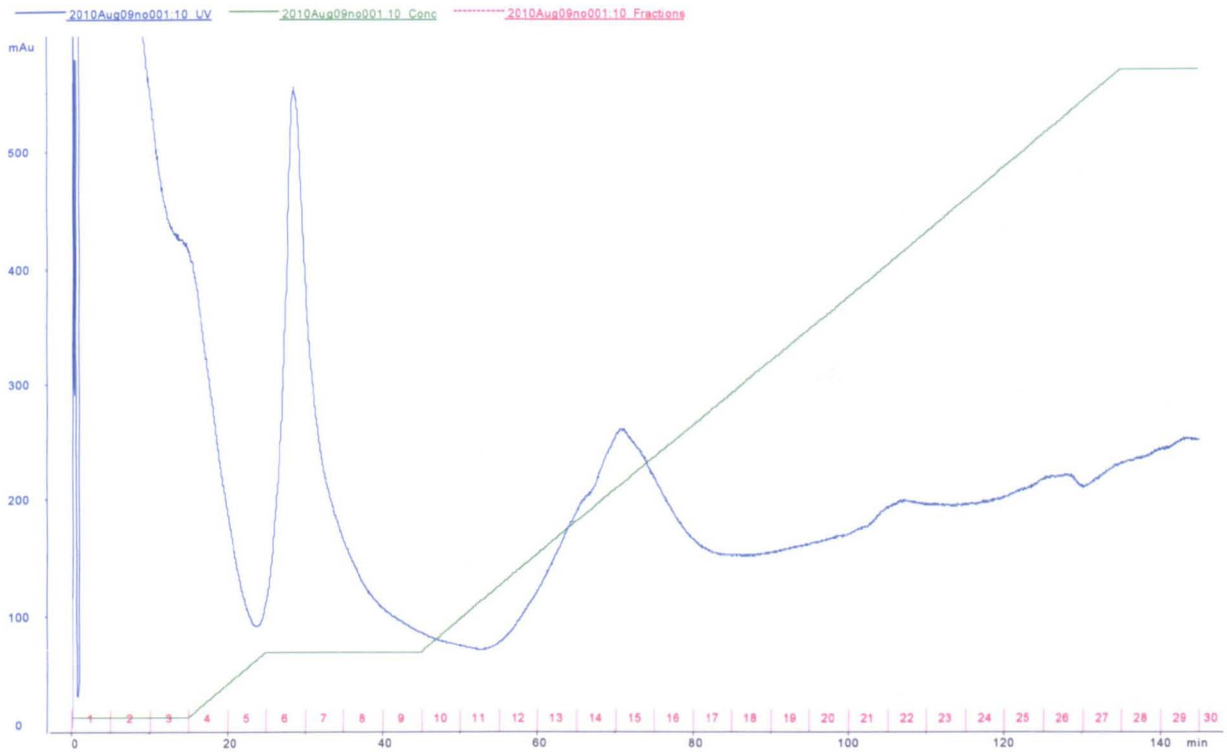


Figure 3.27: Elution profile showing the change in absorbance at 280 nm of the eluant from the Ni-His 60 column following addition of an extra wash step at 180 mM imidazole.

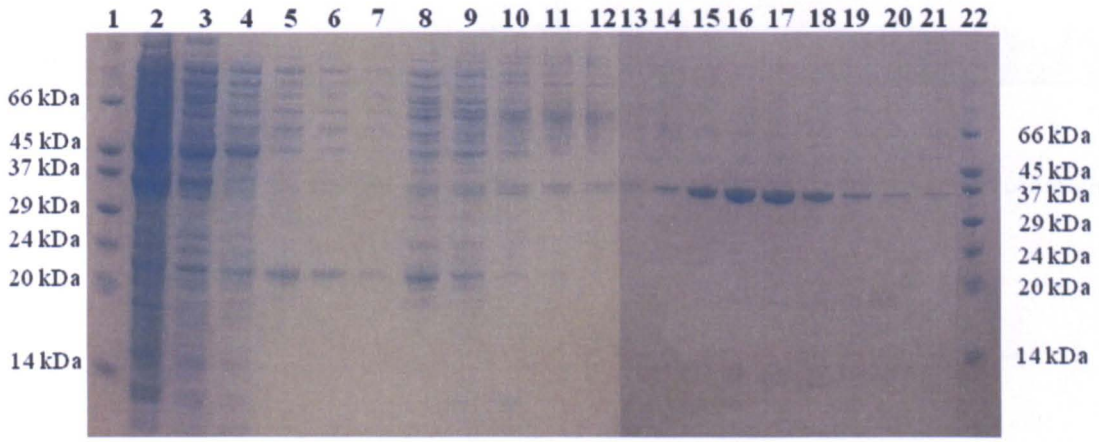


Figure 3.28: SDS PAGE gel showing the proteins present in the 2) loaded fraction along with proteins present in fractions 1 to 19 in lanes 3 to 21 respectively. Lanes 1 and 22 correspond to standard molecular weight markers of a known size.

3.8 Conclusion

The protocol for the over-expression and purification of un-labelled POR provided by Dr. D. Heyes (University of Manchester) originally took five days and produced a yield of 10 mg protein per litre of medium used. Although this is a reasonable amount of protein to be produced, when you take into account the losses observed whilst growing cells in labelled minimal medium, the cost of medium required to produce a 0.5 ml sample of concentration 0.3 mM would not be viable. Following the optimisation of both the over-expression and purification of POR, the length of time taken for the whole preparation was cut down to two days and increased the yield of un-labelled protein produced to 30 mg protein per litre of medium used. The improvements made to the protocol have resulted in enough protein being produced from one litre of labelled growth to conduct both single and double labelled NMR experiments as summarised in Table 3.1.

Type of Protein Produced	Yield of Protein Achieved / Litre of Medium (mg)
Un-labelled protein	30
^{15}N labelled protein	11
^{15}N, ^{13}C labelled protein	5

Table 3.1: Table detailing the amount of protein produced from 1 litre of growth medium containing no, one and two labels.

CHAPTER 4

Optimisation of the NMR Conditions and Production of a ^{13}C, ^{15}N, Deuterated Sample

4.1 Introduction

One of the aims of the PhD was to solve the structure of POR using NMR. In order to do that, a peak corresponding to every amide in the protein needed to be seen in the standard POR HSQC. Figure 4.1 shows the initial spectrum of POR acquired with 0.3 mM ^{15}N labelled POR, in Tris buffer pH 7.5, 1 mM DTT and Roche Complete Protease Inhibitor acquired at 25 °C. At this stage there were nowhere near the 322 peaks expected from the backbone amides. Naturally before any structural studies could be conducted on POR the NMR conditions need to be optimised to allow the visualisation of all of the backbone amide groups. The main areas where improvements in spectral quality were thought to come were from adjusting the pH of the sample, increasing the temperature of acquisition and also optimising the concentrations of salt and ligand present.

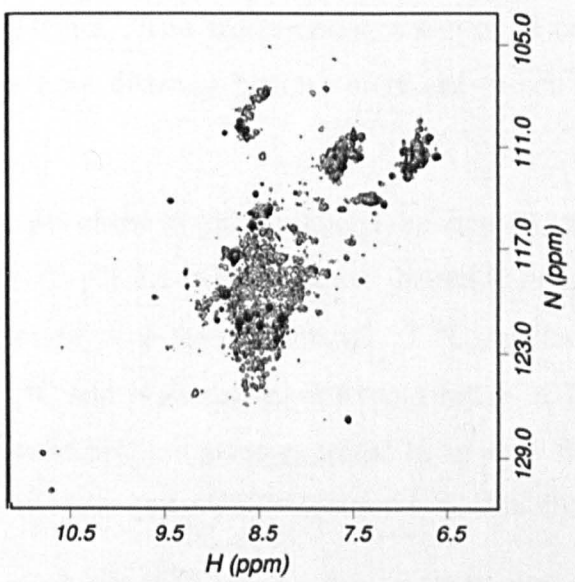


Figure 4.1: The initial HSQC acquired prior to optimisation of the NMR conditions.

4.1 Optimising the conditions for NMR

4.1.1 Broad Range Screen using Sypro Orange

Sypro Orange is a fluorescent dye which binds to hydrophobic regions of proteins (Steinberg *et al.* 1996). Following thermal denaturation, the hydrophobic core of the protein becomes exposed and the SYPRO orange dye binds and becomes unquenched. This process results in an increased fluorescence signal and can be monitored using a real time PCR machine. The reactions can be carried out in 96-well plates allowing a multitude of different buffer conditions to be screened at once, and the observed increase in fluorescence signal can be used as an indicator as to how the different buffers affect the thermal stability of the protein (Steinberg *et al.* 1996; Lo *et al.* 2004).

50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer, pH 5.5 and 50 mM sodium phosphate pH 7.5 were tested with sodium chloride, concentrations ranging between 100 and 300 mM, and ammonium sulphate, concentrations ranging between 0 and 200 mM. 20 μg of POR was loaded into each well along with the relevant buffer and 5 μl of 4 X SYPRO orange before plastic caps were placed on each well. The plate was centrifuged at 2,000 rpm for 1 minute to ensure all reactants were mixed together before being placed in a Stratagene MX3005P real time PCR machine. The plate was heated from 25 °C to 98 °C at a heating rate of 1 °C / minute, and the fluorescence intensity was measured at 530 nm, following excitation at 490 nm. The fluorescence was plotted out against a function of temperature to determine how different buffer conditions affect POR's thermal stability (Figure 4.2).

As expected as the pH of the buffer is reduced the thermal stability of the protein also decreases (Figure 4.2 A). At pH 7.5 POR begins to thermally denature at a temperature of 62 °C and is fully denatured at a temperature of 73 °C, whilst at pH 5.5 the thermal denaturation begins at 57 °C and is completed at a temperature of 73 °C. Although POR is less thermally stable at lower pH, the gains expected to be seen from reducing the rate of amide exchange is expected to far outweigh the lower thermal stability.

The incorporation of salt, both ammonium sulphate (Figure 4.2 B) and sodium chloride (Figure 4.2 C) at any concentrations do not increase the thermal stability of POR. In the presence of sodium chloride, thermal denaturation begins at 60 °C and is completed by 72 °C whilst in the presence of ammonium sulphate thermal denaturation begins at 59 °C and

is completed by 75 °C. Both of these ranges are well within the error of those seen with just altering pH alone so the decision not to include any additional salt was taken. By keeping the salt concentration as low as possible, we will reduce the proton 90 ° pulse length which will maximise the sensitivity of NMR experiments.

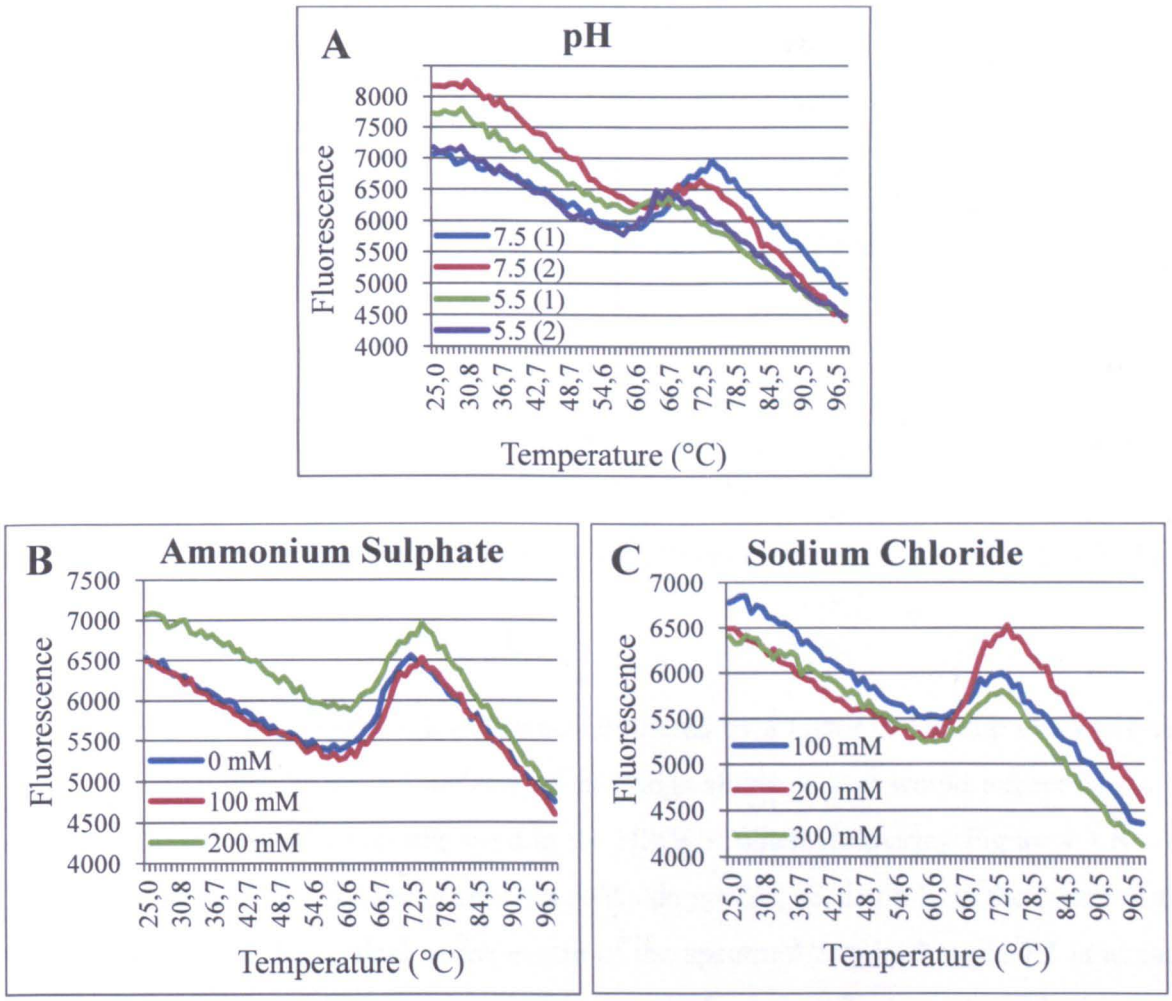


Figure 4.2: SYPRO Orange fluorescence experiments which show the effect of A) pH, B) Ammonium Sulphate and C) Sodium Chloride on the thermal stability of POR.

4.1.2 pH

Two 0.3 mM samples of POR were buffer exchanged into 50 mM sodium phosphate pH7.5 (Figure 4.3 A) and 50 mM sodium phosphate pH 5.5 (Figure 4.3 B). Both buffers contained 100 mM NaCl, 1 mM DTT along with Roche Complete protease inhibitor. 500 µl of each sample was placed into a 5 mm NMR tube and analysed using a 500 MHz

spectrometer pre-equilibrated at 50 °C. The sample was left to equilibrate in the spectrometer for 15 minutes before the spectrometer was set up (detailed 2.21) and the experiments run (detailed Table 2.3).

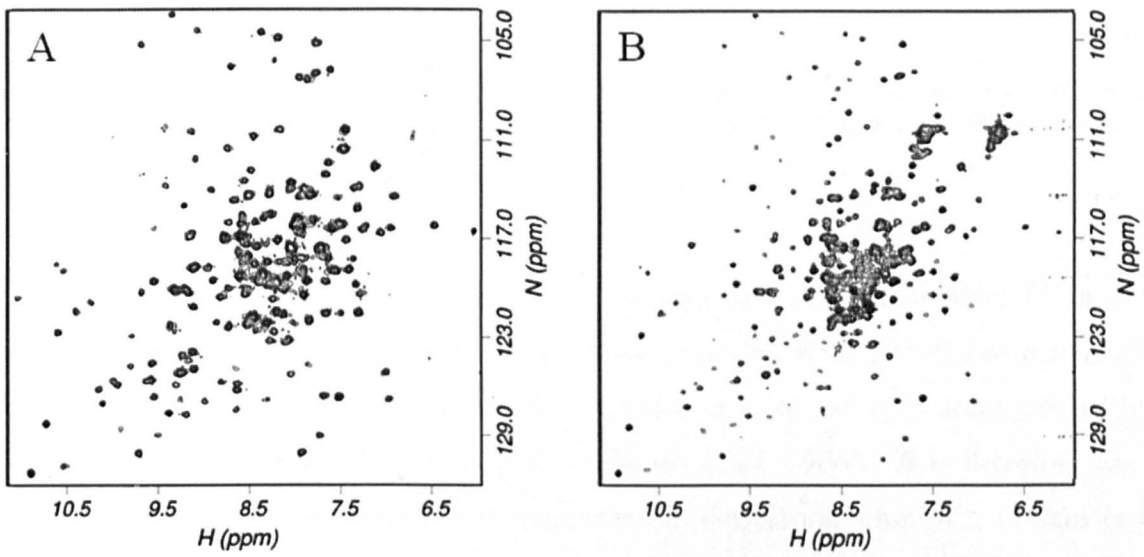


Figure 4.3 HSQC spectra acquired using ^{15}N POR at 600 MHz in sodium phosphate buffer A) pH 7.5 and B) pH 5.5.

The rate of amide exchange with water is altered by a factor of 10 with every pH unit changed. As you reduce the pH the rate of exchange slows so you would expect to see an increase in the number of peaks observed in the HSQC. When comparing Figure 4.3 B with Figure 4.3 A, as predicted a number of extra peaks do appear, particularly in the centre of the spectrum. This is far from ideal as the centre of the spectrum acquired at pH 7.5 is already very crowded with the peaks present being heavily overlapped. It is also noticeable that some peaks which previously appeared as strong peaks at pH 7.5 have decreased in intensity when at pH 5.5. However since those peaks are still present in the spectrum the increased number of peaks observed from working at a lower pH do outweigh the draw backs.

4.1.3 Temperature of Acquisition

The over-expressed POR originates from the thermophilic organism *Thermosynechococcus elongatus* which is natively found at temperatures of 55 °C in hot

springs in Japan (Nakamura *et al.* 2002). By using a thermophilic version of POR this allows us to work at temperatures above 298 K, which will decrease the viscosity of the solvent and decrease the correlation time of the protein (Equation 4.1).

$$\tau_c = \frac{4\,\pi\eta a^3}{3\,kT}$$

Equation 4.1: Equation used to calculate the correlation time of proteins. η is the viscosity of water, a^3 is the hydrodynamic radius of the protein, k is the Boltzmann constant and T is the temperature in Kelvins.

In Equation 4.1 two values change with temperature, the temperature (T) and the viscosity of water (η). The increase in temperature from 298 K to 323 K has a negligible effect on the temperature value; however this increase in temperature is accompanied by a decrease in viscosity from 0.9 to 0.5 kg.s/m^2 (Kestin *et al.* 1966). It is therefore fair to assume that the major component in determining the correlation time of a protein is the solvent viscosity. As we increase the temperature of acquisition, the correlation time of the protein should decrease and we should see a significant improvement in the quality of the spectrum acquired, both in peak line shape and intensity (Figure 4.4).

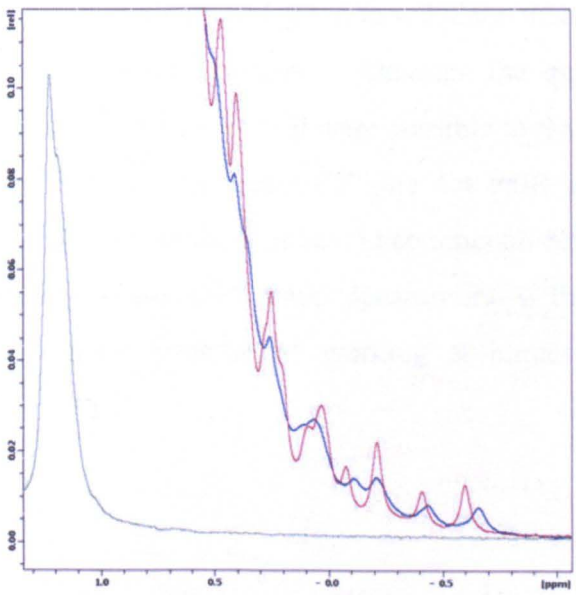


Figure 4.4: 1 D spectra of unlabelled POR between -1.0 and 1.5 ppm acquired at 25 °C (Blue), 55 °C (Red) and 65 °C (Green) using a 600 MHz spectrometer.

Figure 4.4 shows a 1 D spectrum of 0.3 mM POR acquired at 25 °C (Blue), 55 °C (Red) and 65 °C (Green). As the temperature of acquisition increases, the peaks below 0 ppm corresponding to methyl groups packed against aromatic side chains, not only increase in intensity but also get sharper up to a point, 60 °C, where protein denatuaration occurs and the peaks disappear completely.

A 0.3 mM ^{15}N POR sample in 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 1 mM DTT and Roche Complete Protease Inhibitor was placed into a 5 mm NMR tube and analysed using a 600 MHz spectrometer with cryoprobe. The spectrometer was set up as detailed 2.21 and HSQC experiments were run (detailed Table 2.3) between 25 °C and 60 °C at 5 °C intervals (Figure 4.5). As the temperature of the spectrometer was increased the spectrometer was left to equilibrate, with sample in, for 30 minutes before the next experiment was set up and run.

As expected, as the temperature of acquisition increases the number, intensity and line widths of peaks appearing in the spectrum increases, with significant improvements in spectral quality being observed between 40 °C (Figure 4.5 B) and 45 °C (Figure 4.5 C). Above 45 °C (Figure 4.5 D, E & F) no extra peaks seem to appear however the peaks present do appear to get sharper and increase in intensity.

The number of peaks observed in Figure 4.5 D, E & F is somewhere near the 322 expected backbone amide peaks, however it is hard to tell since there is a significant level of overlap observed in the centre of the spectrum. Although the quality of the spectrum is unlikely to increase further, it would be nice if it were possible to resolve the peaks present in the centre of the spectrum further. As seen in Figure 4.4 POR appears to denature at a temperature of 65 °C, as a result it was thought best to conduct all future work at 55 °C. That way the sample would not be at risk of thermal denaturation if there were fluctuations in spectrometer temperature yet the benefits of working at higher temperature would be observed.

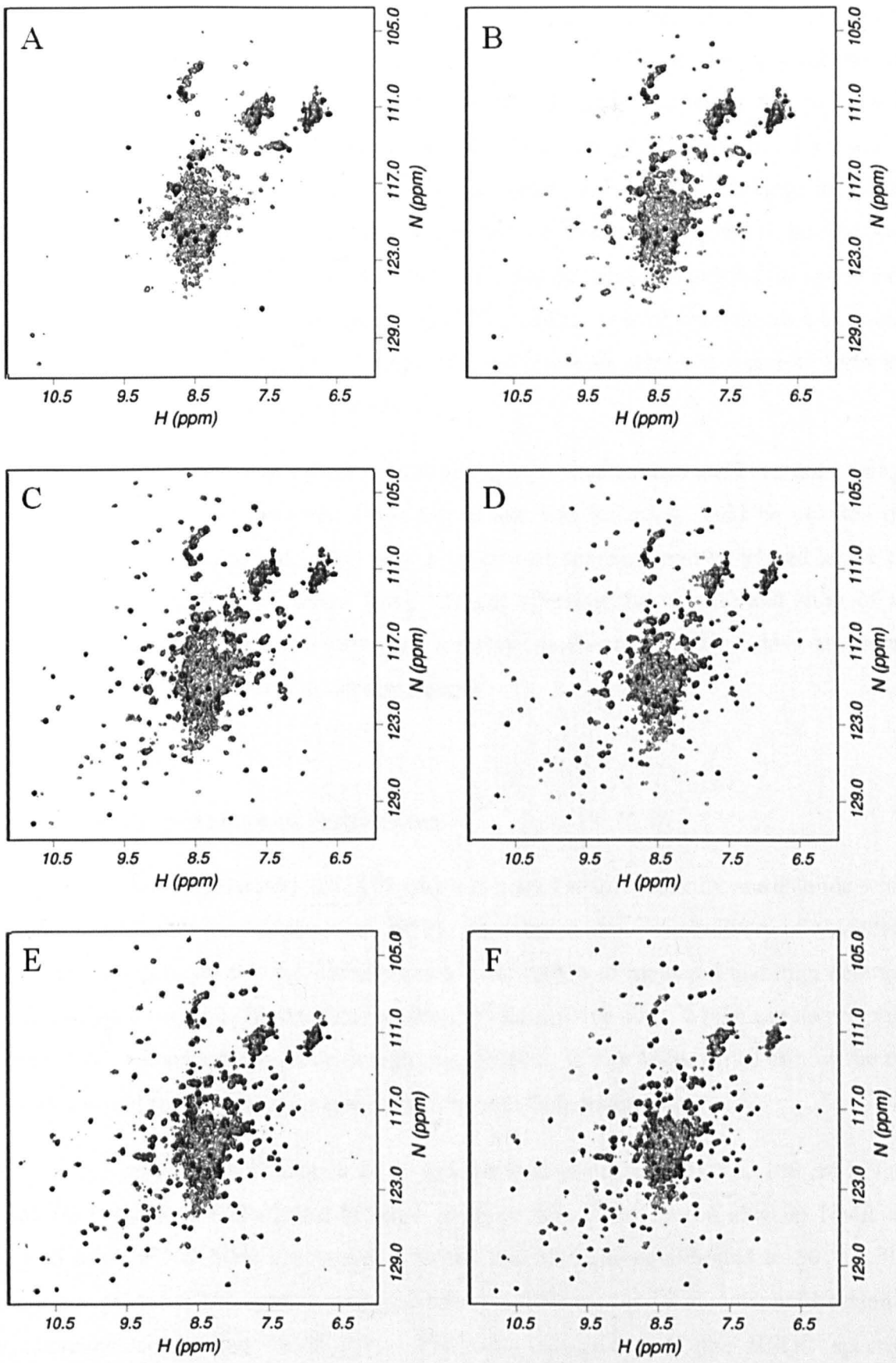


Figure 4.5: HSQC spectra acquired using ^{15}N POR at 600 MHz in sodium phosphate buffer pH 5.5 at A) 30 °C, B) 40 °C, C) 45 °C, D) 50 °C, E) 55 °C and F) 60 °C.

It was observed that following the completion of the experiments at 60 °C the sample lock had dropped significantly. When the sample was removed from the spectrometer the sample volume had noticeably decreased in volume and significant levels of condensation had occurred in the part of the tube which was held within the spinner. Evidently a temperature gradient exists across the spectrometer and whilst working at higher temperatures the sample was evaporating off and re-condensing when it hit the cold environment of the spinner. Once sufficient sample had accumulated around the spinner the weight of the droplet formed was greater than the surface tension and the droplet would rejoin the sample. This continual refluxing was sufficient to disrupt the sample shim and affect the quality of the spectrum obtained.

All previous NMR experiments conducted at high temperatures had been done using a normal 5 mm NMR tube however it was hoped that this refluxing could be avoided if a shigemi tube was used instead. That way a volume of sample could be placed above the shigemi plunger which could reflux away without affecting the volume and shim of the sample being analysed. This was tested and samples could sustain multiple days of analysis at high temperature without losing the sample lock.

4.1.4 Incorporation of Substrates

POR has two substrates; NADPH which is bound at (nM) affinity and Pchlide which is bound at (μM) affinity (Menon *et al.* 2010). As detailed above the delivery of NADPH is in aqueous solution, yet the Pchlide requires a combination of methanol and high detergent concentrations in order to be delivered to the enzyme's active site. Whilst the thermophilic form of POR is known to be stable at high temperatures, it was unknown as to how the two substrates would respond to prolonged periods at high temperature.

A 0.3 mM ^{15}N POR sample in 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 1 mM DTT and Roche Complete Protease Inhibitor was placed into a shigemi NMR and analysed using a 600 MHz spectrometer, which had been pre-equilibrated to 50 °C. The sample was placed in the spectrometer and left to equilibrate for 15 minutes before running the experiments (detailed Table 2.3). Following acquisition of the HSQC spectrum (Figure 4.6 A) the sample was removed from the spectrometer and 0.3 mM NADPH was added to the sample; this was a significant excess however it was done so to ensure that all

protein molecules bound NADPH and remained bound. The sample was returned to the spectrometer and left to equilibrate for 15 minutes before running the experiments (detailed Table 2.3).

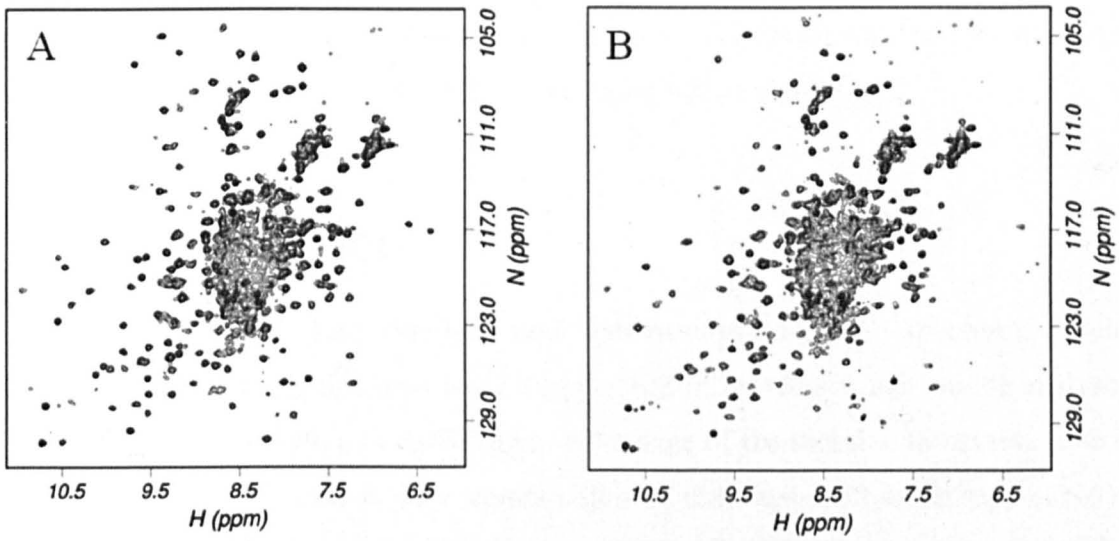


Figure 4.6: HSQC spectra acquired using ^{15}N POR at 600 MHz in sodium phosphate buffer pH 5.5 in the A) absence of NADPH and B) presence of 0.3 mM NADPH.

Upon addition of NADPH to the sample (Figure 4.6 B) some peaks on the outside of the spectrum are observed to shift, which would indicate that the protein has indeed bound NADPH. However there appears to be no significant improvement in the dispersion of peaks in the centre of the spectrum. 1 D spectra acquired before and after the HSQC spectrum indicated significant changes had occurred in the amide envelope during the course of the experiment. Peaks corresponding to folded protein however had not changed in intensity or line shape, so it was thought that the changes observed in the amide envelope were due to NADPH degradation.

Following acquisition of the HSQC spectrum in the presence of NADPH (Figure 4.6 B) the sample was removed from the spectrometer and 0.3 mM Pchlide, re-suspended in methanol, was delivered to the sample in 50 mM BOG. The sample was returned to the spectrometer and left to equilibrate for 15 minutes however after this incubation time there was a problem whilst trying to shim the sample. The sample was removed from the spectrometer and a green precipitate was observed to have formed at the

bottom of the NMR tube. It was unknown as to why this had happened but it was thought best to develop a new protocol for the delivery of both substrates, where the NMR could be conducted in the absence of both solvent and detergent.

Since the incorporation of substrates had not improved the quality of the spectrum, the decision was made to conduct structural studies with apo POR. Although this is not a complex which is found in the catalytic cycle (Heyes *et al.* 2004) it was thought that NADPH would not survive 10 days at 55 °C without becoming fully oxidised.

4.1.5 TROSY vs HSQC

The Transverse Relaxation-Optimised Spectroscopy (TROSY) spectrum, developed by Pervushin *et al*, has significantly increased the size of proteins which can be analysed by NMR (Pervushin *et al.* 1997). TROSY takes advantage of the fact that transverse relaxation of ^{1}H and ^{15}N nuclei is caused by a combination of chemical shift anisotropy (CSA) and dipole-dipole (DD) interactions, and that a component of the HSQC spectrum can be selected in which these two relaxation mechanisms oppose each other. This leads to greatly reduced relaxation rates and thus much sharper and more intense signals, although at the expense of selecting one quarter of the total signal. This effect is predicted to be greatest at frequencies close to 1 GHz. The T_2 relaxation within an amide group can be almost completely cancelled out for one of the four components and it is this signal which TROSY observes. Here the residual line width is almost solely a result of DD interactions with remote hydrogen atoms, which can be further suppressed following deuteration (Ikura *et al.* 1990). This technique has been further developed for 3D spectroscopy, allowing the gains observed to be extended to structure assignment experiments (Salzmann *et al.* 1998).

0.3 mM ^{15}N labelled POR in 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 1 mM DTT and Roche complete protease inhibitor was placed in a shigemi tube and placed in a 600 MHz spectrometer which had been pre-equilibrated to 50 °C (described 2.21). HSQC and the equivalent TROSY experiment were acquired using the same sample and processed in an identical way (Figure 4.7)

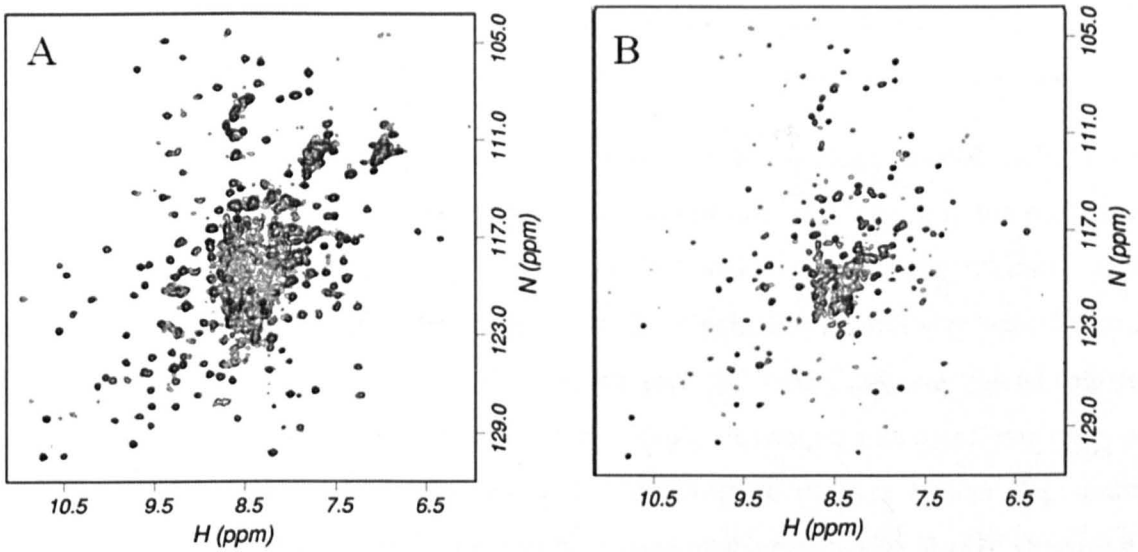


Figure 4.7: Comparison of a A) HSQC and B) TROSY spectrum acquired using ^{15}N POR at 600 MHz, in sodium phosphate buffer pH 5.5.

When comparing Figure 4.7 A and B it is noticeable that there is significantly more signal present in the HSQC than in the TROSY spectrum. This is to be expected as each peak present in Figure 4.7 A is the summation of all four signal components, achieved from running the spectrum with decoupling in both the t_1 and t_2 dimensions, whilst the peaks present in Figure 4.7 B are a result of the single component, for which the CSA T_2 relaxation is at its minimum. Unfortunately the further away from the optimal spectral frequency you get the less effective the TROSY effect is. This can also be seen as although the peaks which are present in Figure 4.7 B are sharper, their intensity is significantly less and in some instances are not even present, than their corresponding peaks in Figure 4.7 A. As a result at these lower field strengths it would be more beneficial to run HSQC versions of the experiments required.

4.2 Long Term Stability of POR

For many years POR has been observed to degrade into a 34 kDa fragment and a 3 kDa fragment if left for prolonged periods of time at room temperature (Figure 4.8 A). The cleavage site has been mapped to occur between lysine 250 and lysine 251. This proteolytic activity has been proposed to act as an auto-regulatory mechanism for the enzyme, as in high light conditions, the enzyme will be working at maximum capacity. However under these

conditions the production of chlorophyll is actually detrimental to the organism. This proteolysis is thought to be one of the reasons as to why crystal trials to date have failed.

Work conducted by Durin (Durin 2006) indicated that in the presence of EDTA the rate of proteolysis was reduced. Standard procedure in our lab was to work in the presence of protease inhibitors to stabilise the proteins. Roche produce a protease inhibitor which contains EDTA so based on the observations made by Durin, the natural step was to include the Roche Complete Protease Inhibitor into the sample. 0.3 mM POR was placed into four different buffers and left at room temperature for 10 days. A sample was taken every day and placed in SDS PAGE running buffer before being frozen to prevent further degradation. Once all ten samples had been taken, the samples were analysed by SDS PAGE (Figure 4.8) (detailed 2.19).

The incorporation of the complete protease inhibitor (Figure 4.8 B) has completely eradicated any signs of proteolysis when compared to Figure 4.8 A. Also contradictory to the results seen by Durin, the incorporation of EDTA alone (Figure 4.8 D) not only does not prevent the degradation of POR but actually speeds it up. The concentration of EDTA added to the sample was 1 mM, the same as what is found in the Roche Complete Protease Inhibitors, however it is unknown how much EDTA Durin added to the samples. There is however a potential draw back in using the complete protease inhibitor: Figure 4.9 A shows a 1D spectrum of buffer containing Roche Complete Protease Inhibitor whilst Figure 4.9 B shows a 1 D spectrum of buffer containing Roche Complete Protease Inhibitor which has been filtered through a 10 kDa PES membrane.

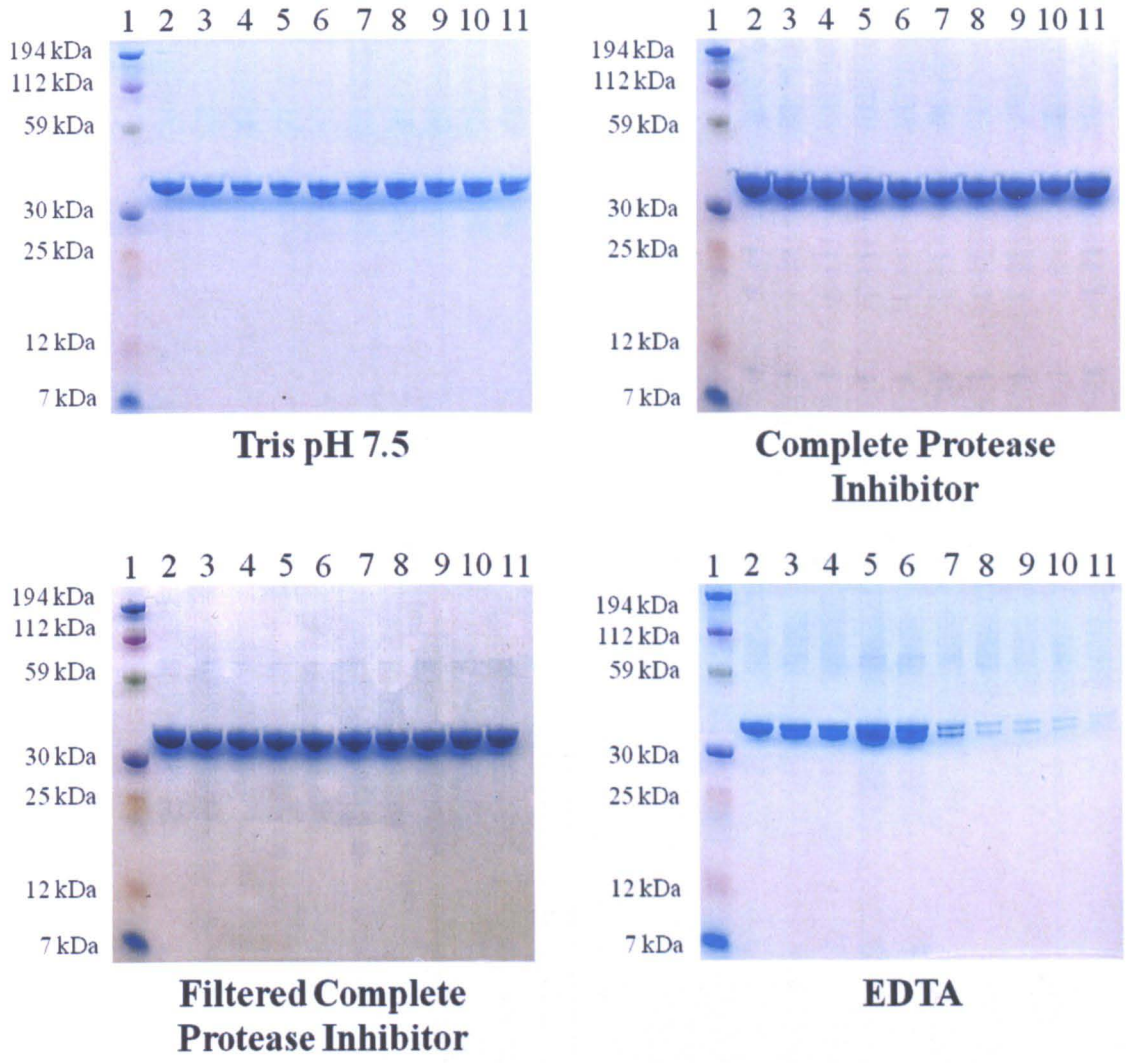


Figure 4.8: Gels showing stability of POR over a period of 1 to 10 days, lanes 2-11 respectively in A) Tris buffer pH 7.5, B) Tris buffer pH 7.5 with Roche complete protease inhibitor, C) Tris buffer pH 7.5 with complete protease inhibitor filtered through a 10 kDa membrane and D) Tris buffer pH 7.5 with 1 mM EDTA. Lane 1 in all gels corresponds to standard molecular weight markers of a known size.

Prior to filtration of the NMR buffer five broad peaks are observed between 1.5 and 2.5 ppm. This is generally the region where methyl bound proton peaks appear so removal of these peaks, if possible, is a benefit. Following filtration of the buffer these peaks disappear (Figure 4.9 B). It was wondered what was present in the protease inhibitor which correspond to the observed broad peaks, however it was assumed that they were due to a large polysaccharide 'packing' compound which was used to form the tablet. Roche were contacted but were reluctant to divulge what was present in their protease inhibitors which could correspond to these peaks.

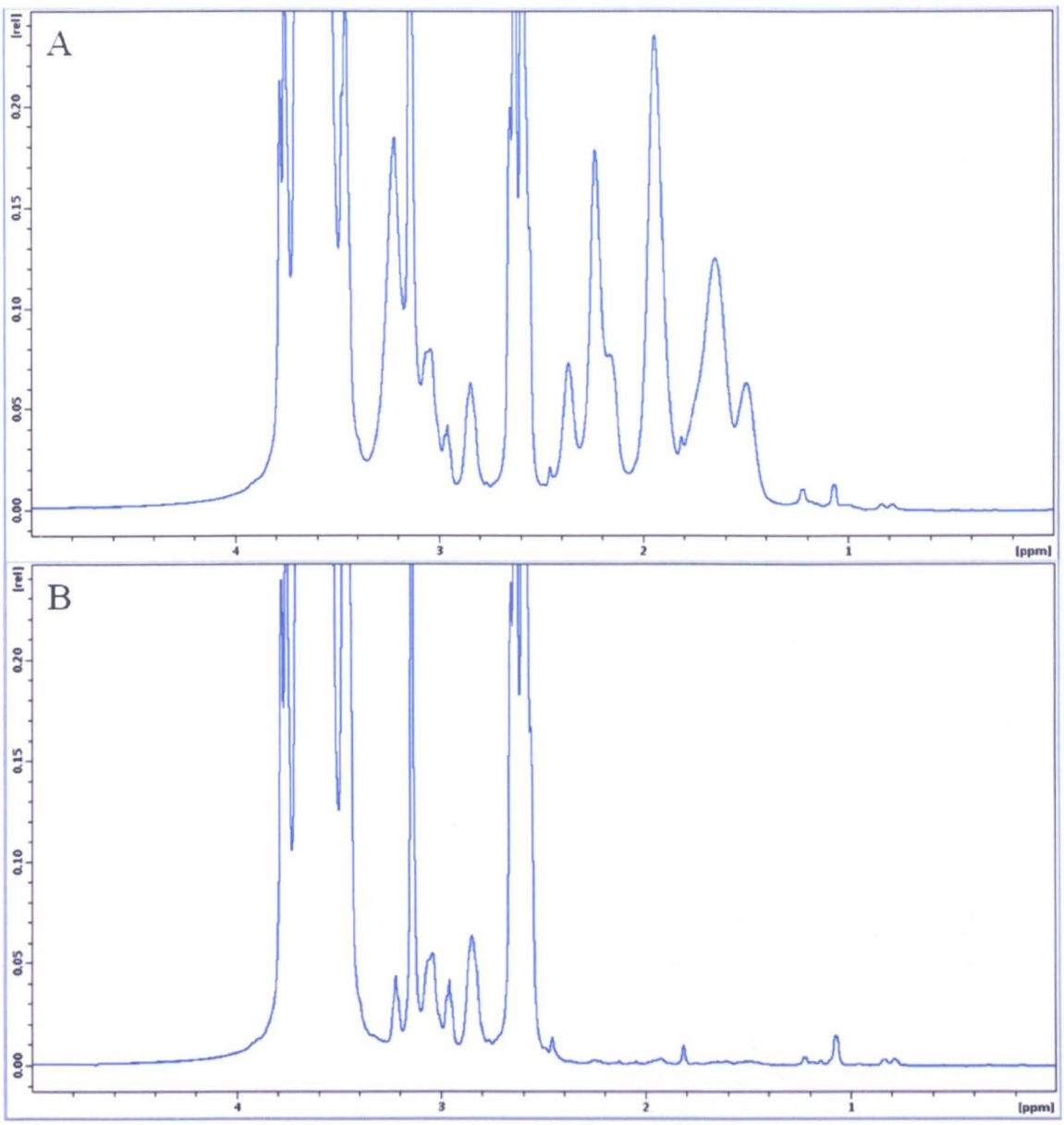


Figure 4.9: 1D spectra of A) TRIS buffer containing Roche complete protease inhibitor and B) the same buffer but filtered through a 10 kDa membrane acquired at 25 °C using a 500 MHz spectrometer.

A sample of POR was exchanged into buffer containing filtered complete protease inhibitor and analysed for stability at room temperature (Figure 4.8 C). Whilst with the un-filtered protease inhibitor no degradation products were observed, following 5 days of incubation with the filtered protease inhibitor a degradation product started to appear. This confirmed that whatever these peaks corresponded to they were vital in the prevention of the degradation of POR and thus needed to be included into the sample. Ultimately the backbone experiments will be conducted on a sample which is ^{13}C, ^{15}N labelled so since these peaks

will not be labelled the presence of these broad peaks should not affect the quality of the spectrum achieved.

4.2.1 Long Term Stability of POR at high temperature

Having established the conditions required to stabilise POR and prevent degradation at room temperature it was important to see if these conditions could also be used at higher temperature to prevent degradation. Although SDS PAGE gels acted as a good indicator as to how different conditions affected the enzymes stability, it was thought that if we repeated these experiments at higher temperatures in the NMR spectrometer, then we could accurately quantify how much degradation had occurred if any. Two 0.3 mM unlabelled POR samples were prepared and placed in 50 mM sodium phosphate buffer pH 5.5, 100 mM NaCl, 1 mM DTT and Roche complete protease inhibitor. The samples were placed in two shigemi tubes and incubated at 45 °C and 55 °C. Every 24 hours the samples were analysed using a 500 MHz spectrometer which had been pre-equilibrated to the desired temperature (Figure 4.10).

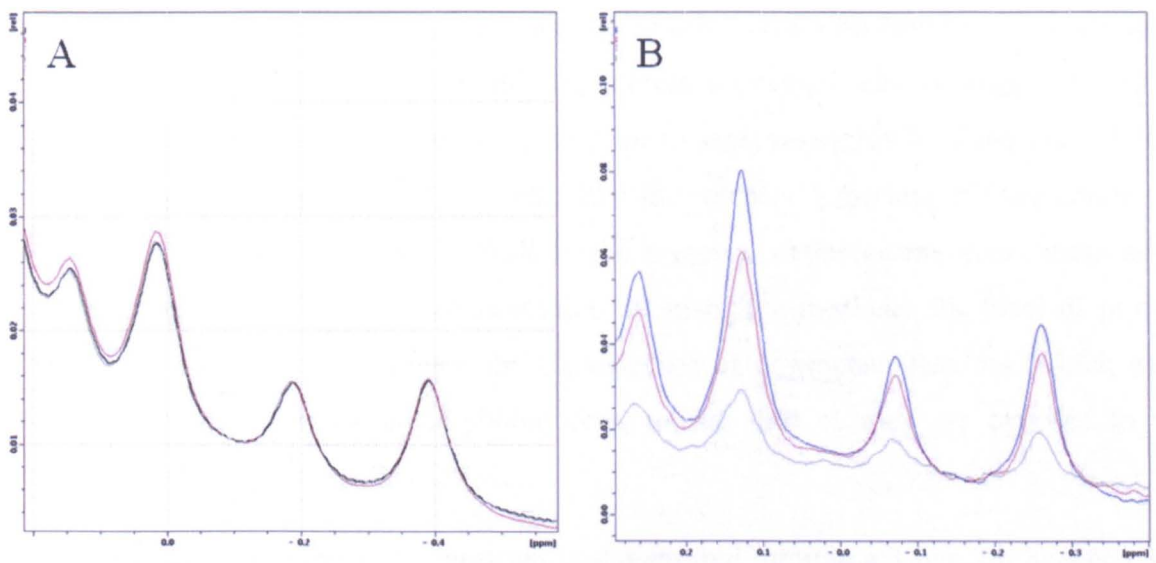


Figure 4.10: 1D spectra acquired using unlabelled POR at 600 MHz in sodium phosphate buffer pH 5.5 at A) 45 °C and B) 55 °C at time 0 (blue) and 24 (red), 60 (green) and 120 (black) hours of incubation.

Following incubation at 55 °C for 30 hours the intensity of the folded protein peaks have decreased by 8 % and after 60 hours of incubation, only 24 % of the initial protein remains in solution. The sample incubated at 45 °C however shows no sign of degradation even after 120 hours of incubation. This does concur with data obtained in Figure 4.2, which implies that at approx 57 °C thermal denaturation of POR begins. Whilst working with a deuterated sample it would be preferable to see no degradation so it was thought prudent to acquire all data at 45 °C. That way any fluctuations in temperature observed would not be enough to affect POR.

4.3 Back exchange of amide protons

Following the deuteration of proteins, all protons present in the protein will be replaced with deuterons. Whilst the presence of carbon bound deuterons is essential in reducing the DD T_2 relaxation (Pervushin *et al.* 1997) all ^{15}N bound deuterons require to be back exchanged to protons, as amide bound protons are the initial starting point of magnetisation for almost all backbone NMR experiments. Amide protons readily exchange with the solvent and although this is fine for solvent exposed groups, amide groups which are present in tight secondary structure in the hydrophobic core of the protein, generally do not readily get exposed to the solvent. Previous studies conducted with deuterated protein have used chemical means to completely unfold their protein, allowing back-exchange of all amide groups to occur. The protein is then refolded prior to analysis by NMR (Reed *et al.* 2003). To see if this was possible with POR, tryptophan fluorescence experiments were conducted with a 60 μM sample of unlabelled POR in the presence of increasing concentrations of guanidine hydrochloride. As the concentration of guanidine increases the level of protein unfolding will increase. This causes the fluorescence of tryptophan residues, which were previously packed within the hydrophobic core, to red shift as they are exposed to the aqueous environment of the solvent.

The fluorescence emission spectrum was measured through a 3 mm slit between 310 and 400 nm, following excitation through a 1.5 mm slit at 280 nm. The change in fluorescence observed at 332.4 nm was used to measure the level of protein unfolding (detailed 2.18.1) (Figure 4.11).

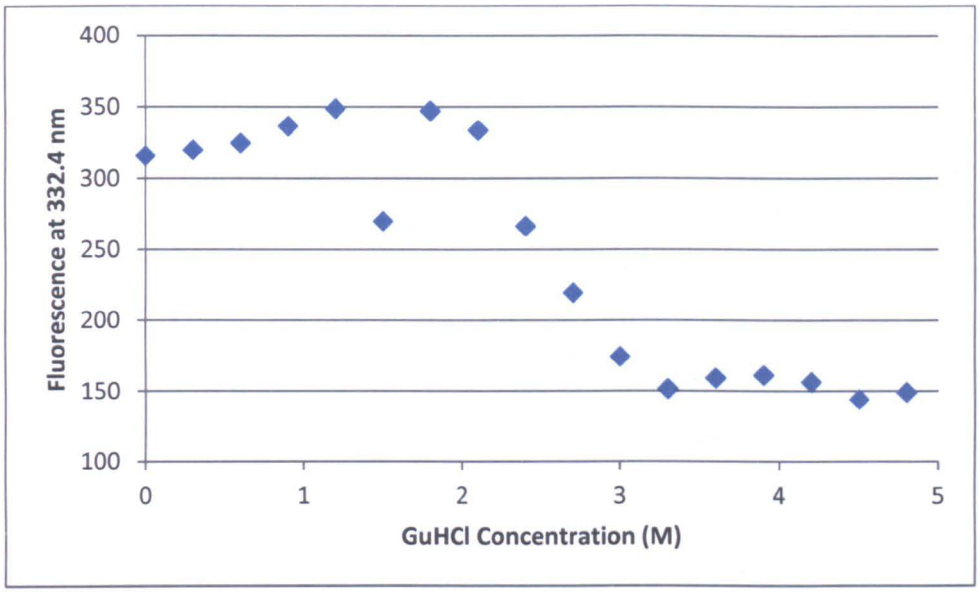


Figure 4.11: Graph showing the changes in tryptophan fluorescence of POR, measured at 332.4 nm, with increasing concentrations of guanidine hydrochloride.

As the concentration of guanidine increases in the sample, the level of protein unfolding increases. The level of tryptophan fluorescence increases slightly between 0 and 1.2 M before the signal rapidly drops off between 1.8 and 3.3 M. From 3.3 to 4.8 M there is no significant change in the level of protein fluorescence, so it is fair to assume that at 3.3 M guanidine the protein is fully unfolded. There is one anomalous point at 1.5 M however this point appears in this position in all repeats conducted and in other tryptophan fluorescence experiments conducted with different concentrations of enzyme at different temperatures.

Following complete protein unfolding with 4.1 M guanidine, the sample was loaded into dialysis tubing and dialysed overnight against 5 litres of low salt buffer at 4°C. Following dialysis the sample was concentrated (described 2.10) and the fluorescence spectrum was taken as described above. The tryptophan fluorescence was the same as that previously observed for fully folded protein, however when the protein sample was assayed, there was no enzyme activity. It was therefore decided to try and avoid the use of guanidine and to try and back exchange all amide protons with temperature alone.

As described above, the rate of amide exchange increases by a factor of 10 for every pH unit dropped. Whilst when improving the quality of the spectrum it was beneficial to work at a lower pH to decrease the rate of amide exchange, for this purpose it is more beneficial to work at a higher pH to increase the rate of amide exchange. The pI of POR had

previously been calculated to be approximately 9.4 so it was thought best to work at 1 pH unit below this, that way avoiding any loss and precipitation of protein.

A 250 μM ^{15}N sample of POR which had previously been used for analysis at pH 5.5 was split into five samples and diluted five fold with sodium phosphate buffer pH 7.5 in D_2O, 100 mM NaCl, 1 mM DTT and complete protease inhibitor. The buffer pH was uncorrected for the presence of deuterium prior to the dilution of the H_2O sample; as a result it was difficult to accurately determine the pH of the sample, however it was estimated to be around pH 6.8. The sample was placed in a 500 MHz spectrometer, which had been pre-equilibrated to 45 °C and HSQC spectra were taken to measure the rate of exchange (Figure 4.12).

It was hoped that as the experiment progressed and more amide protons were exchanged to deuterons the number of peaks present in the spectrum would decrease. This was certainly true however following 48 hours incubation at 45 °C (Figure 4.12 B) a significant number of peaks remained. The temperature of the spectrometer was increased to 55 °C for 24 hours and another HSQC spectrum was acquired (Figure 4.12 C). Although there were less peaks remaining in the spectrum than had previously been there, following the incubation at 45 °C, there was still a significant number of peaks remaining. It was realised that because the sample had been in 80 % D_2O, complete exchange would never occur. It was therefore decided to prepare a fresh ^{15}N labelled POR sample and buffer exchange it into a 100 % D_2O sodium phosphate buffer pH 7.5, 100 mM NaCl, 1 mM DTT and complete protease inhibitor. The sample was placed into the spectrometer which had been pre-equilibrated to 55 °C and a HSQC spectrum was acquired (Figure 4.12 D)

Following the 30 minutes incubation time at 55 °C, there were no peaks present in the HSQC spectrum acquired. This confirmed that heat, along with a higher pH was sufficient to back exchange the amide protons. Since all the purification of POR is conducted at pH 7.5 it was thought easy enough to incorporate an additional heating step at higher temperatures, thus allowing the back exchange process to occur. Since it had been previously shown that incubating the sample at high temperature in the presence of other proteins was detrimental to the sample (Figure 3.6 & Figure 3.7) it was decided to put this additional heating step in following elution of the protein from the SP sepharose column. That way the protein will be pure and less affected by the precipitation of other denatured proteins.

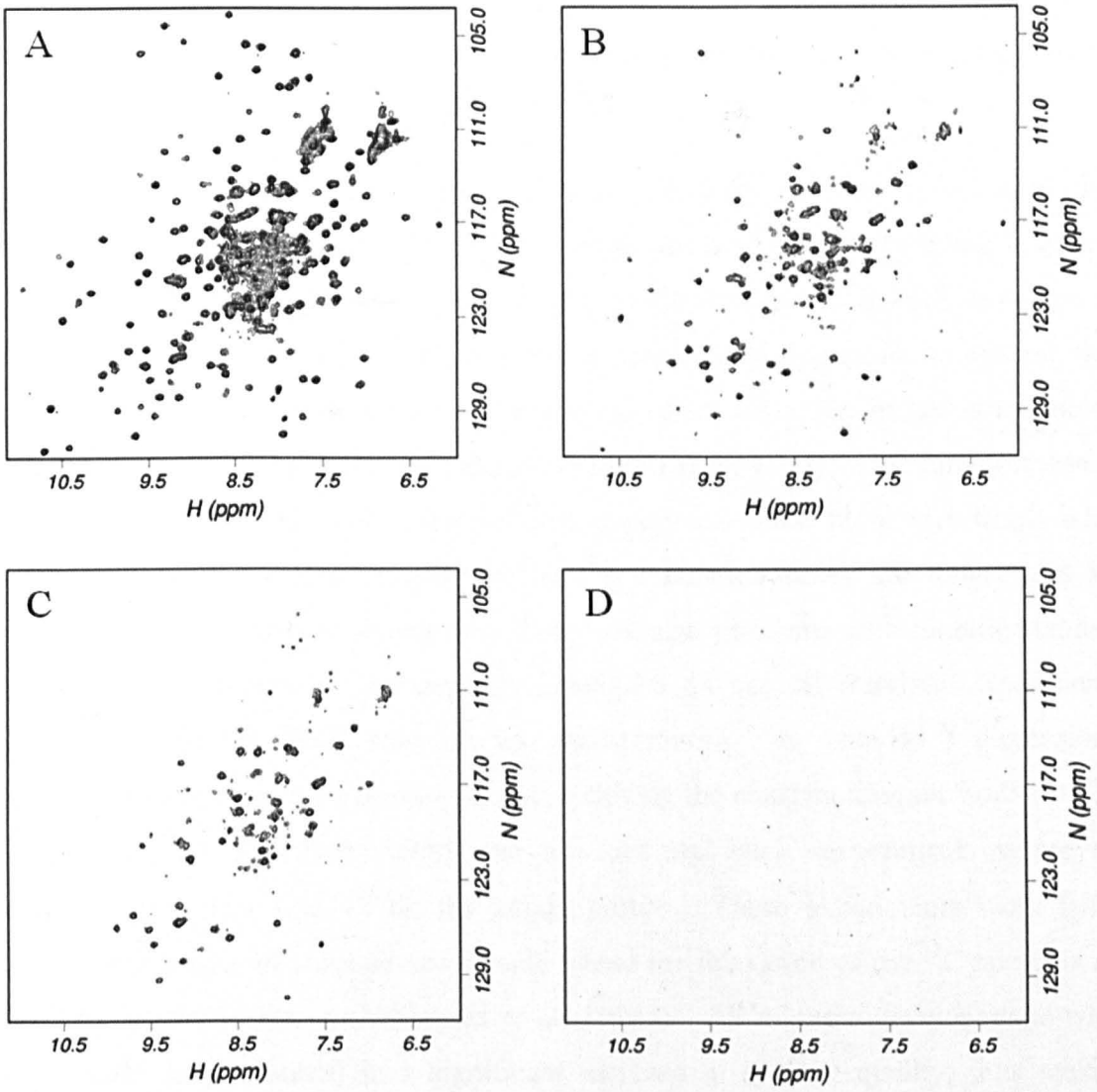


Figure 4.12: HSQC spectra acquired using ^{15}N POR at 600 MHz in 80 % D_2O sodium phosphate buffer pH 6.8, showing the level of deuterium exchange achieved at A) time 0 B) 48 hours incubation at 45 °C C) 24 hours incubation at 55 °C. D) Shows the level of exchange achieved using a ^{15}N sample in 100 % D_2O sodium phosphate buffer pH 7.5.

4.4 Production of a per-deuterated sample

NMR studies have primarily concentrated on proteins which have had a molecular weight of less than 25 kDa however, developments in samples and pulse programs used have extended the range of sizes which can be analysed by NMR up to approx 40 kDa. One of the major reasons why larger proteins can not be analysed using NMR is due to the dipole-dipole relaxation which is experienced, specifically from methyl protons (Pervushin *et al.* 1997). Deuteration should reduce the R_2 relaxation rates of ^{13}C nuclei, particularly for C-H nuclei, and thus make the line widths of peaks sharper as the gyromagnetic ratio of ^{2}H is

approximately one sixth that of protons. This effect can be significantly reduced by using a deuterated protein sample, that way the rates of relaxation are decreased by approximately a factor of 7 (Grzesiek *et al.* 1993a; Kushlan *et al.* 1993).

Increases in spectral quality are also achieved from the use of constant time experiments. In isotopically enriched proteins, the level of homonuclear coupling in normal pulse programmes is relatively high, which results in the decrease in resolution of spectra acquired. This is more prominent when using a per-deuterated sample. Constant time experiments modify the normal pulse programme to reduce the effect of the homonuclear coupling, thus improving the spectral resolution (Santoro *et al.* 1992). The improvements in spectral resolution observed, with constant time experiments, are more significant when working with a per-deuterated sample, as the ^{13}C relaxation times are longer and the experiments take a shorter time to acquire. There are also problems with running standard 3 dimensional experiments on per-deuterated samples as not all standard experiments, specifically the CBCA(CO)NH start on the amide proton. A suite of 3 dimensional experiments have therefore been developed, incorporating the constant time methods detailed above (Yamazaki *et al.* 1994b) which are all 'out and back' experiments where the magnetisation starts and finishes on the amide proton. These experiments work better however in the presence of a deuterated sample where the relaxation of the ^{13}C nuclei is not influenced by attached protons (Yamazaki *et al.* 1994a). All of these developments when coupled together have resulted in a significant increase in spectral quality, thus making analysis of larger proteins by NMR possible.

4.4.1 Acclimatisation of *E. coli* for growth in D_2O

D_2O is toxic to *E. coli,* as a result when the dominating solvent becomes D_2O instead of H_2O significant lag times occur. Previous work conducted on per-deuterated proteins (Yamazaki *et al.* 1994a; Reed *et al.* 2003) have expressed the benefit of acclimatising a strain of *E. coli* to growth in D_2O, thus reducing the lag time before growth starts. This acclimatised strain can then be used for growth and expression in triple labelled media, with minimal lag time.

Competent BL21pLysS *E. coli* were transformed with His_6 Thermo POR plasmid DNA and plated out on LB agar plates containing the relevant selection (described 2.5).

Single colonies were picked from the plate of transformed *E. coli* and inoculated into 100 ml of LB medium, made up with 100 % H_2O, containing the relevant selection. A Cary 50 spectrophotometer was used to measure the absorbance of the growth at 600 nm (described 2.6.4). Once the cells had reached an optical density of 0.6, 1ml of growth was removed from the medium and used to inoculate 100 ml LB medium, made up with 25 % D_2O. The process was repeated with LB medium made up with 50 %, 75 % and 100 % D_2O. Once the 100 % D_2O LB medium had been inoculated, the cells were left overnight to grow to completion before a glycerol stock was made and stored at -86 °C (detailed 2.5.1).

E. coli cells grew without any lag time or delay when in medium where H_2O was the dominating solvent however when the 75 % D_2O LB growth was inoculated, the bacteria took over 24 hours before they entered into log phase. After this point the *E. coli* grew normally and following inoculation in the 100 % D_2O LB, the cells followed the growth curve of normal non-induced *E. coli* (as observed in Figure 3.1).

4.4.2 Over-expression and Purification of a per-deuterated sample

To maximise the amount of protein produced from 1 litre of triple labelled medium, it was decided to bulk grow the *E. coli* in unlabelled LB medium. The cells would then be spun down in sterile Beckmann centrifuge flasks before being washed with sterile Milli-Q H_2O and finally D_2O. The *E. coli* could then be stored overnight in the fridge before being re-suspended in the triple labelled medium and grown further before being induced.

A sample of D_2O resistant *E. coli* were streaked out on an LB agar plate containing the relevant selection and incubated overnight at 37 °C. Single colonies were picked and used to inoculate 100 ml LB medium containing the relevant selection, prior to incubation overnight at 37 °C (detailed 2.6.1). 1 ml of overnight starter culture was used to inoculate two large scale LB growths (detailed 2.6.3) which were grown at 37 °C for 4 hours. Once a dense culture had been established, the cells were removed from the incubator and centrifuged at 10,000 rpm for 15 minutes, in sterile Beckmann centrifuge tubes (detailed 2.7). Following pelleting of the cells, the medium was removed and the cells were re-suspended in 50 ml of sterile Milli-Q H_2O. The cells were mixed gently before being centrifuged again at 10,000 rpm for 15 minutes (detailed 2.7). The supernatant was once again removed and the bacteria were re-suspended in 10 ml sterile D_2O, before being transferred to sterile 50 ml

falcon tubes. The *E. coli* were once again centrifuged at 3,000 rpm using a refrigerated bench-top centrifuge. Once again the supernatant was removed and the cells were stored at 4 °C overnight.

1 litre of per-deuterated medium was made up (detailed A.1.3) using uniformly deuterated ^{13}C glucose (Goss Scientific) as a carbon source and ^{15}N ammonium sulphate (Goss Scientific) as a nitrogen source. The *E. coli* prepared previously were removed from the refrigerator and re-suspended in 20 ml of triple labelled medium before being inoculated into the large scale growths. The growths were placed into the 37 °C incubator and the optical density was checked every hour (Figure 4.13).

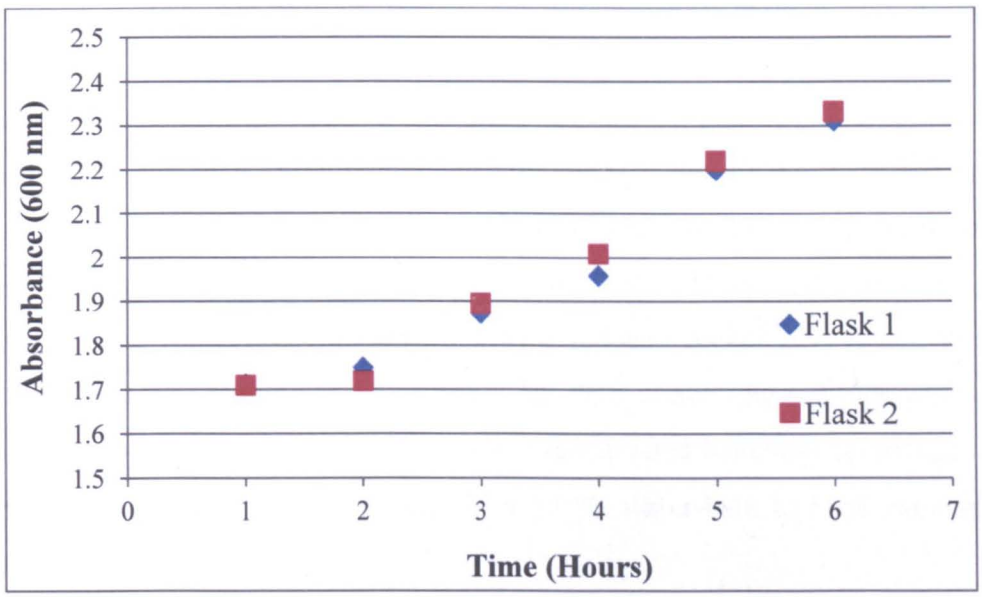


Figure 4.13: Growth curve of *E. coli* BL21pLysS following re-suspension in triple labelled M9 minimal medium.

Following re-suspension of the *E. coli* in the per-deuterated medium, the rate of growth was slow; as a result it was difficult to spot when the cells entered into log phase. Figure 3.3 E shows that following 4.5 hours of growth at 37 °C, *E. coli* growth in LB medium reach an OD of 1.7, the optimal point for induction. The concern was that if the cells were left to grow for any longer without being induced, then the *E. coli* would grow past the optimal point of induction and protein over-expression would be hindered. After 4.5 hours of growth at 37 °C, the *E. coli* were induced with 0.1 mM IPTG (detailed 2.6.5) and transferred to a 25 °C incubator for 2.5 hours. Following protein induction, the *E. coli* were removed

from the incubator and harvested (detailed 2.7) before being re-suspended in binding buffer and storage, in universal tubes, at -20 °C.

E. coli containing the triple labelled protein were removed from the freezer and defrosted. One Roche complete protease inhibitor tablet was added to the cell samples prior to sonication. The cell mixture was heated at 42 °C for 15 minutes prior to being centrifuged at 25,000 rpm for 30 minutes (detailed 2.8). The supernatant was removed from the centrifuge tubes and an 85 % ammonium sulphate precipitation was conducted (detailed 2.9.1). The precipitated protein was centrifuged at 20,000 rpm for 30 minutes and the supernatant was discarded.

Ammonium sulphate protein pellets were re-suspended in 20 ml of binding buffer and filtered through a 0.45 μM syringe filter. The re-suspended protein was loaded onto a pre-equilibrated Ni His-60 IMAC column at a rate of 0.5 ml/min and purified (detailed 2.9.2). All fractions detected to contain POR were loaded straight onto a pre-equilibrated SP sepharose ion-exchange chromatography column, at a rate of 0.5 ml/min, and purified (detailed 2.9.3). All fractions detected to contain POR had 1 ml of low salt buffer containing 100 mM DTT and 5 X Roche complete protease inhibitor. Upon addition of the DTT and EDTA to the POR containing fractions, a brown reduced nickel EDTA salt immediately formed in all fractions (detailed Figure 3.20). All POR containing samples were combined in a falcon tube and centrifuged at 3,000 rpm in a refrigerated benchtop centrifuge to remove the salt prior to incubating for 15 minutes in a 55 °C water-bath to back exchange all the amide protons.

The sample was then concentrated in an Amicon stirred cell concentrator using a 10 kDa PES membrane and buffer exchanged into 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 20 mM DTT and Roche complete protease inhibitor (detailed 2.10). Following concentration, the protein concentration was estimated using the Bradford assay (detailed 2.11) and the amount of protein produced from 1 litre of triple labelled medium was estimated to be 2.5 mg. The sample was concentrated down to a volume of 300 μl and placed in a shigemi tube for analysis by NMR. The sample was placed in a 600 MHz spectrometer which had been pre-equilibrated to 45 °C before 1D and HSQC spectra were acquired (Figure 3.14).

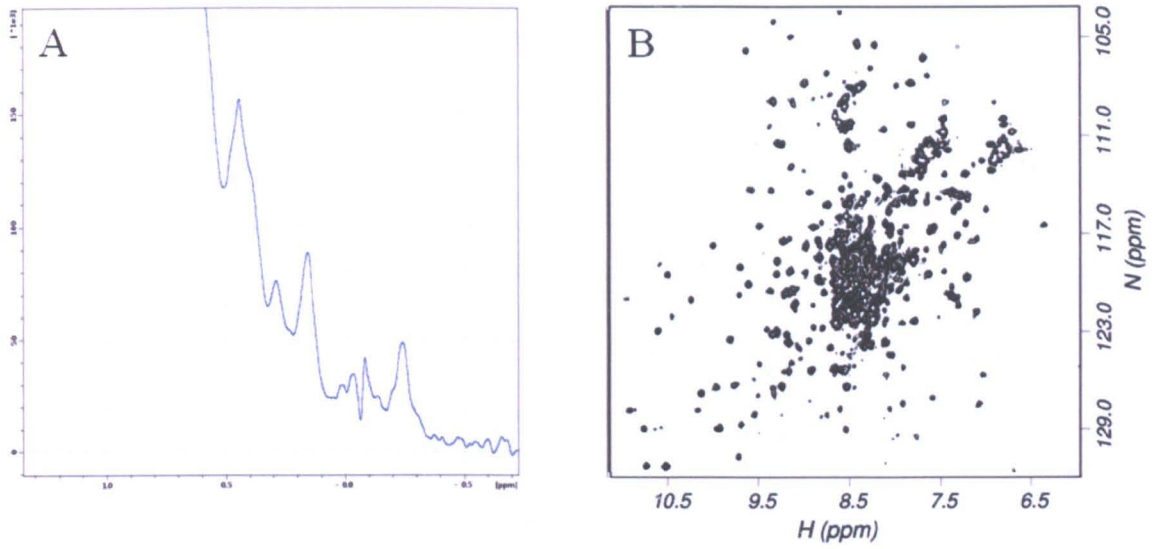


Figure 4.14: A) 1D and B) 2D HSQC spectra acquired using ^{13}C ^{15}N ^{2}H POR at 600 MHz in sodium phosphate buffer pH 5.5.

The sample produced good spectra with the line-widths of well resolved peaks in the HSQC spectrum (Figure 4.14 B) looking significantly sharper than previously observed in HSQC experiments acquired. This led us to believe that the sample was fully deuterated with all amide peaks successfully being back-exchanged. The 1D spectrum acquired however implied otherwise. Proton peaks were present in the 1D spectrum below 0 ppm indicating that methyl groups packed against aromatic side chains were still protonated. A sample of the protein was provided to Mark Dickman (Bio-incubator, Sheffield University) to analyse by mass spectrometry. The data shown in (Figure 4.15) is for a 6 residue fragment, AEQAAK which is located in the 2nd α-helix.

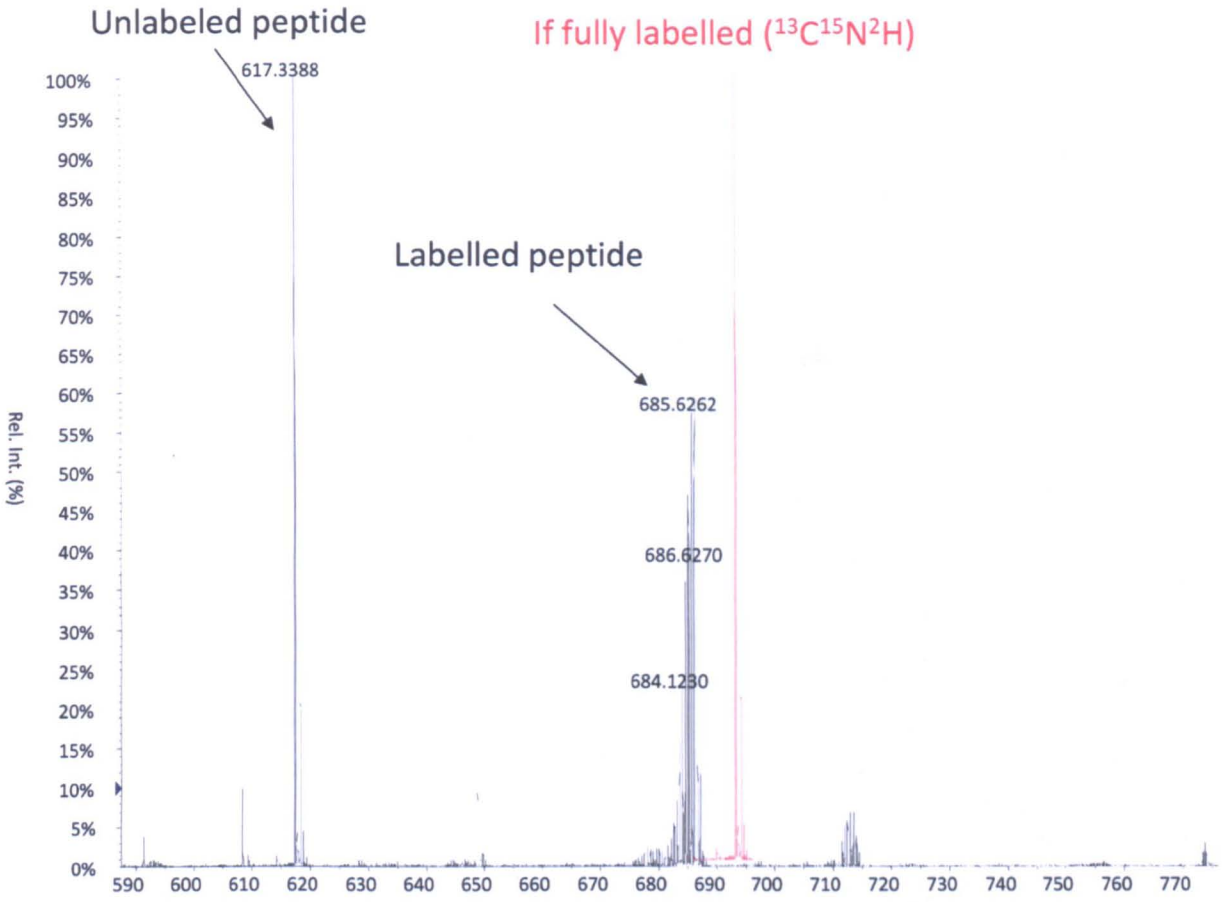


Figure 4.15: Tryptic digest Mass Spectrum of triple labelled POR showing the species present in the sample prepared (black) and the position of a fully labelled, unexchanged fragment (red).

Figure 4.15 shows the molecular weights of the two species present in the sample along with the predicted molecular weight of the peptide if fully ^{13}C, ^{15}N, ^{2}H labelled. The two species present in the sample correlate to a fragment which is fully unlabelled and a fragment that is ^{13}C, ^{15}N, ^{2}H labelled that has successfully had its nitrogen bound deuteriums back exchanged for protons. Since the plasmid is under pLysS control there should have been no expression of POR during growth of the cells in the LB medium and the presence of per-deuterated protein confirmed that protein expression had occurred properly. The inference was therefore that the *E. coli* had been induced too early following growth in the per-deuterated medium. If the *E. coli* had produced stocks of amino acids and precursors whilst being grown in the unlabelled medium, it would make sense that these would be used preferentially before more are made from the labelled sources. This would therefore explain the presence of the two protein species. To test this, a fresh triple labelled sample was prepared in exactly the same way however following re-suspending of the *E. coli* in the

per-deuterated medium, the *E. coli* were grown for 5.5 hours at 37 °C before induction with IPTG. 1D spectra acquired of the purified protein showed no evidence of peaks below 0 ppm (Figure 4.16 B), however good signal was still observed in the amide envelope thus implying that there was only one protein species present which was fully deuterated.

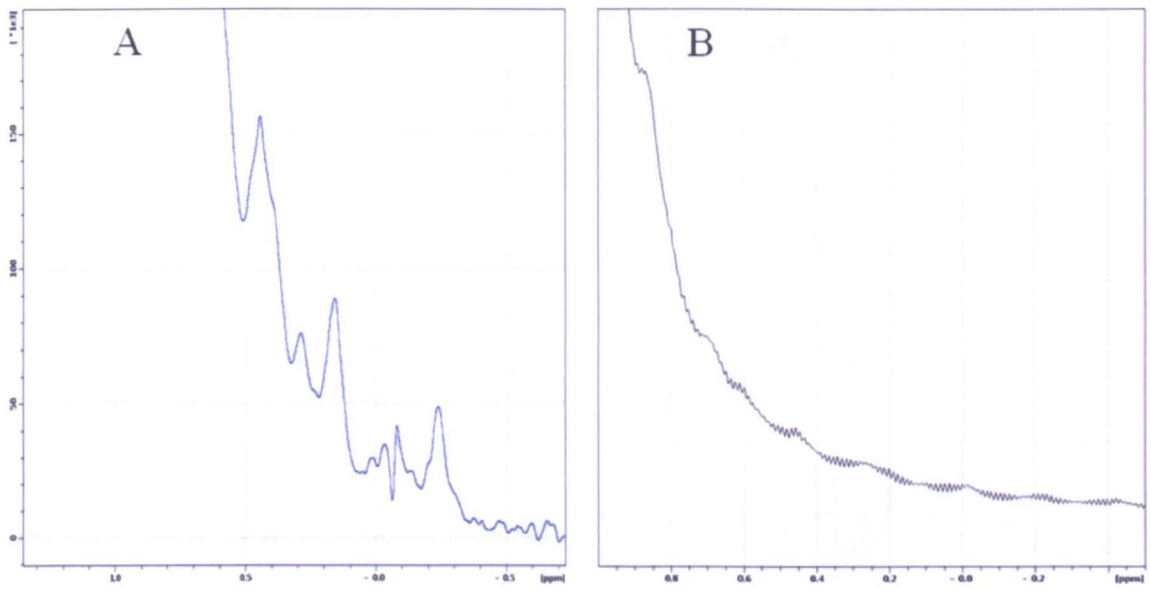


Figure 4.16: 1D spectra acquired HSQC spectra acquired at 600 MHz in sodium phosphate buffer pH 5.5 using ^{15}N ^{13}C ^{2}H POR over-expressed in *E. coli* which were grown for A) 3.5 hours prior to induction and B) 4.5 hours prior to induction.

4.4.3 NMR experiments acquired with a per-deuterated sample

Having confirmed that the triple labelled sample produced was fully labelled and all carbon atoms had deuterons attached (Figure 4.16), the sample was buffer exchanged into 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 20 mM DTT and Roche complete protease inhibitor using an Amicon stirred cell concentrator, with a 10 kDa PES membrane. The sample was concentrated to a volume of 300 µl (detailed 2.10) before being placed in a shigemi tube and the concentration of the sample was estimated to be 225 µM using the Bradford assay (detailed 2.11). The sample was placed in a 600 MHz spectrometer which had been pre-equilibrated to 45 °C before HSQC and TROSY spectra were acquired (detailed 2.21) (Figure 4.17).

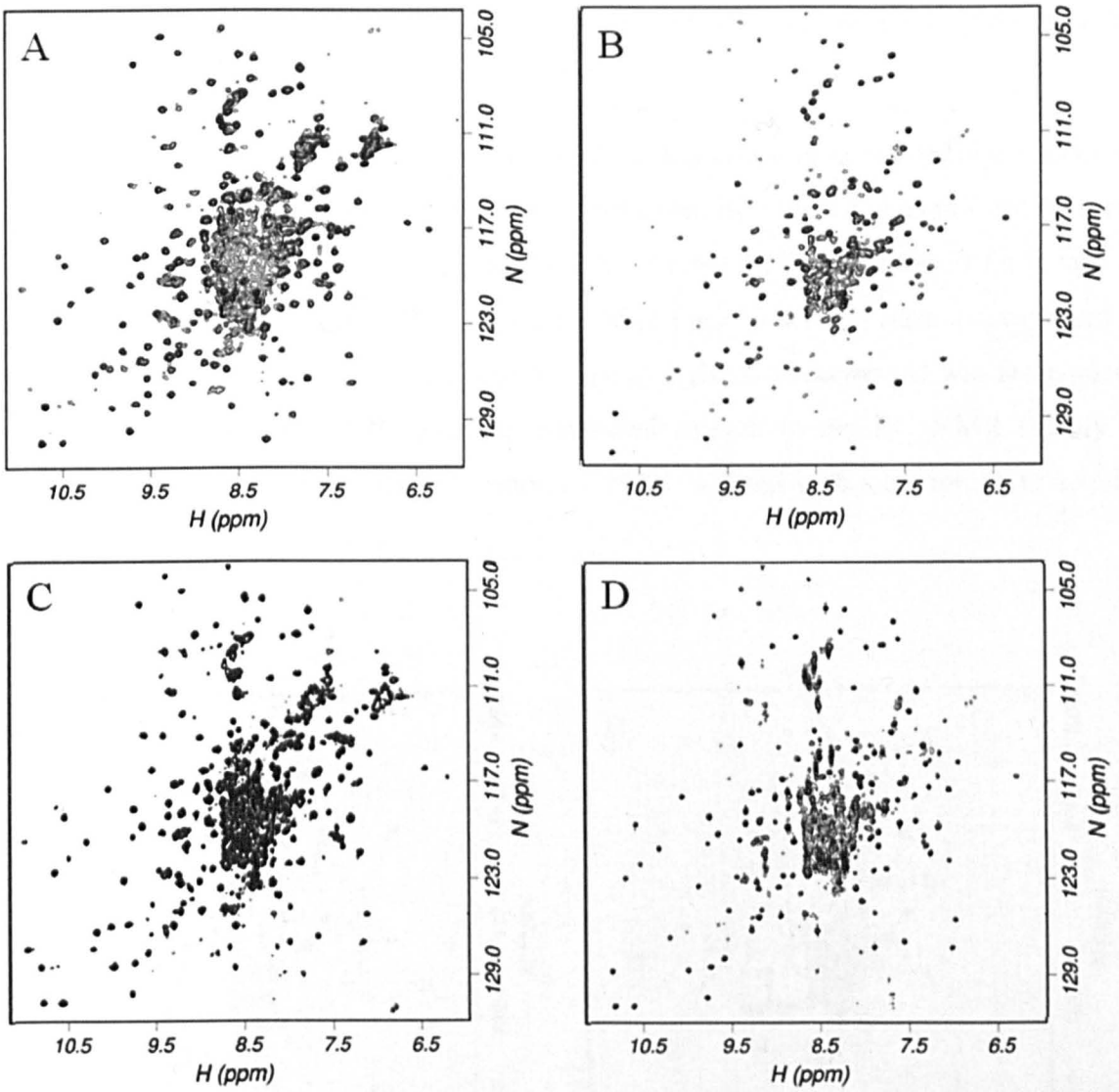


Figure 4.17: A) HSQC and B) TROSY spectra acquired with a ^{15}N labelled POR sample along with C) HSQC and D) TROSY spectra acquired with a ^{15}N, ^{13}C, ^{2}H labelled POR in sodium phosphate buffer pH 5.5 at 600 MHz.

The peaks present in the HSQC and TROSY spectra acquired using the deuterated sample (Figure 4.17 C and Figure 4.17 D) appeared to be sharper than the equivalent peaks present in the HSQC and TROSY spectra acquired using a single ^{15}N labelled sample (Figure 4.17 A and Figure 4.17 B). As expected there is also significantly more signal present in the TROSY spectrum acquired using the deuterated sample, than with that acquired using the single labelled sample (Figure 4.17 D and Figure 4.17 B). There does appear however to be less signal present in both deuterated spectra acquired than in the HSQC spectrum acquired with the single labelled sample. Also there do appear to be changes in the chemical shifts of some peaks present in the deuterated spectra along with some peaks which

are clearly not present that were present in the spectra acquired with the single labelled protein.

Taking all of the above data into consideration it is concerning that there are so many problems associated with the deuterated sample, as the improvements expected were far more significant than those observed; however as detailed above the TROSY effect is far greater at higher field strength than at lower field strength. Since the 800 MHz spectrometer present in Sheffield is not stable for prolonged periods of time at high temperature and was temporarily out of action, it was decided to take the deuterated sample to the EU NMR facility at Birmingham and use their 900 MHz spectrometer which is fitted with a cryoprobe to acquire a TROSY spectrum (Figure 4.18).

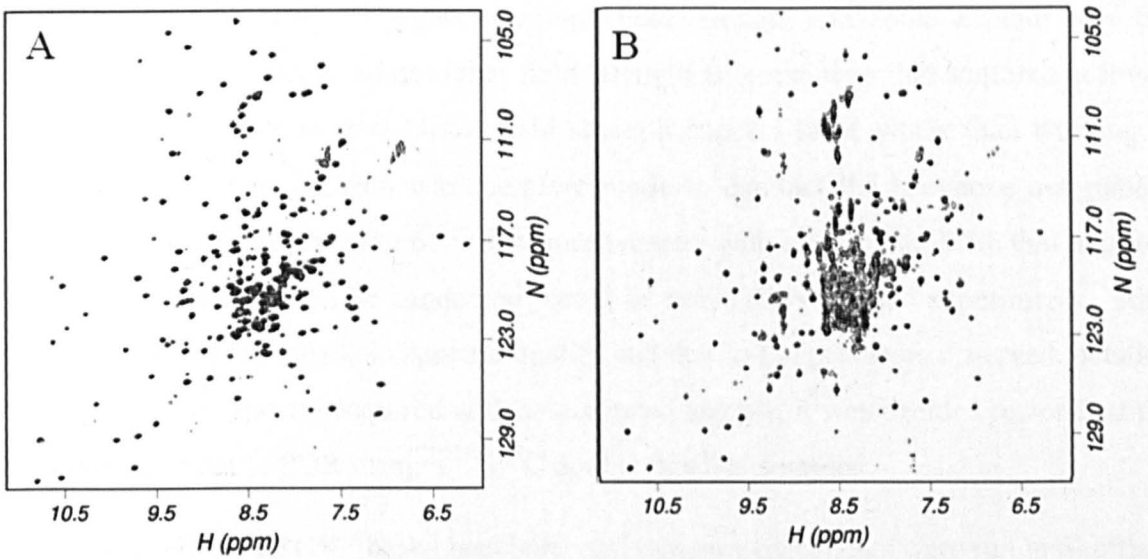


Figure 4.18: TROSY experiment acquired using a ^{15}N, ^{13}C, ^{2}H POR sample in sodium phosphate buffer pH 5.5 at (A) 900 MHz in Birmingham (B) 600 MHz in Sheffield.

Again the quality of the TROSY spectrum acquired at 900 MHz (Figure 4.18 A) did not show the significant improvements expected from running TROSY experiments at higher magnetic field. In fact the number of signals present in the centre of the spectrum appeared to have decreased significantly. Since it was the same sample run in both spectrometers it was decided that the reduction in quality of the spectrum was a sample specific problem. As detailed above whilst purifying the triple labelled protein a significant brown precipitate was observed upon the addition of EDTA and DTT. The implication of this was that the

deuterated protein produced had different purification properties, as the protein had evidently bound and leached nickel from the Ni-His 60 column (detailed 2.9.2). Where and how the protein has bound the nickel is unknown; since the protein still has its hexa-his tag attached it is possible that the nickel is bound to that. However this is the case for all POR proteins produced so this does not explain why the formation of the reduced nickel EDTA salt should be worse with the deuterated protein purified. Since POR binds a porphyrin ring which is co-ordinated by a central divalent magnesium metal ion it is possible that the nickel is bound to the enzyme active site. This may explain why the formation of the nickel salt is worse with the deuterated protein as the environment of the substrate binding pocket will have altered significantly following the replacement of all carbon bound protons with deuterons.

If some residual nickel is bound to POR, then this could result in a degree of paramagnetic relaxation associated with the signals in close proximity to the bound metal ion. This effect would be worse at higher magnetic field strength and could explain why the quality of the spectrum acquired at higher field strength is worse than that acquired at lower field strength. Since working at higher field strength appears to be worse than working at lower field strength the decision was therefore made to conduct the backbone assignment experiments in Sheffield using the 600 MHz spectrometer with cryoprobe. With that in mind, the backbone experiments to be conducted would be non-TROSY based experiments. Also with the lack of improvements in spectral quality and due to the problems observed, detailed above, in the HSQC spectra acquired with a deuterated sample, it was decided to conduct the backbone assignment of POR using a ^{15}N ^{13}C double labelled sample.

A suite of non-TROSY based backbone assignment experiments were run in Sheffield as described above, however whilst collecting the data for the last experiment, the HNCACB, the cryoprobe on the spectrometer failed and the sample was ejected. With the HNCACB being one of the least sensitive experiments the use of a cryoprobe whilst collecting this data was viewed as a necessity. The sample was therefore taken to the EU NMR facility in Birmingham, where a non-TROSY HNCACB experiment was going to be run on their 600 MHz spectrometer with cryoprobe. Following processing the quality of the data acquired was worse than that acquired during the limited running on the spectrometer in Sheffield. A TROSY based HSQC experiment was also run on the sample at Birmingham, prior to the collection of the HNCACB data. Upon processing, the quality of this spectrum compared to that acquired with the same sample in Sheffield appeared again to be worse (Figure 4.19). Upon return to Sheffield the sample was once again analysed and an identical spectrum to

that achieved prior to taking the sample to Birmingham was achieved. This implied that there was nothing wrong with the sample however the difference in the quality of the spectrum achieved was a result either of the hardware at Birmingham or how they run their experiments. With this in mind, an element of doubt was cast over the decisions made about working at higher field strength and the deuteration of the sample to be used for the backbone assignment.

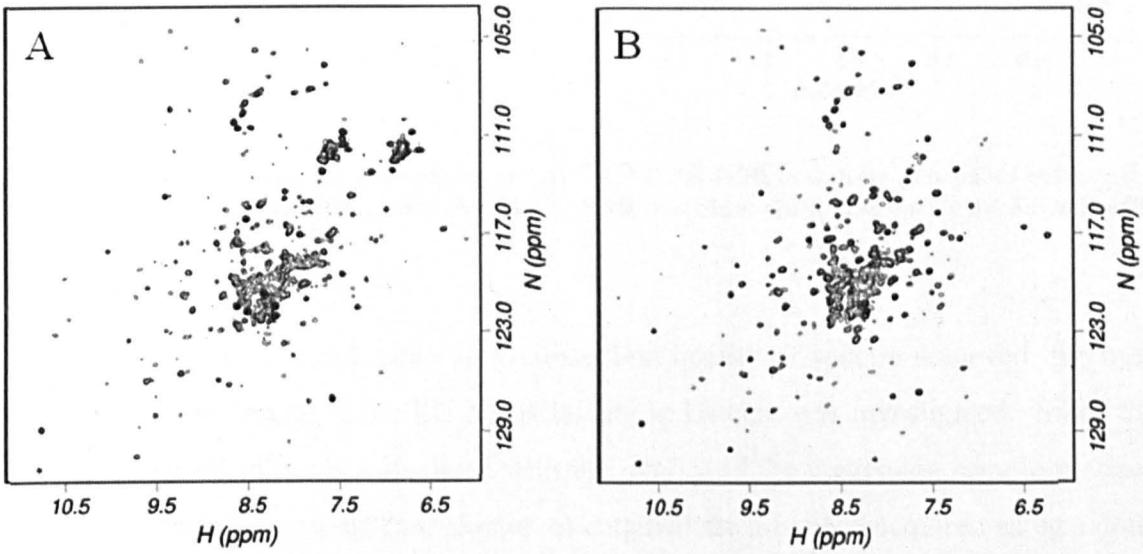


Figure 4.19: TROSY experiment acquired at 600 MHz using a ^{15}N, ^{13}C POR sample in sodium phosphate buffer pH 5.5 in (A) Birmingham (B) Sheffield.

The 800 MHz spectrometer in Sheffield had been repaired so a TROSY based HSQC spectrum was acquired at 45 °C (Figure 4.20). Upon processing of the spectrum, the increases in quality of the spectrum achieved were in line with that initially expected from conducting a TROSY experiment at higher field strength. Also when comparing the quality of the spectrum achieved at 800 MHz in Sheffield (Figure 4.20 B) to that acquired in Birmingham at 900 MHz with a cryoprobe (Figure 4.20 A) the differences observed confirmed the assumptions made about the quality of the hardware or the way experiments are run at Birmingham.

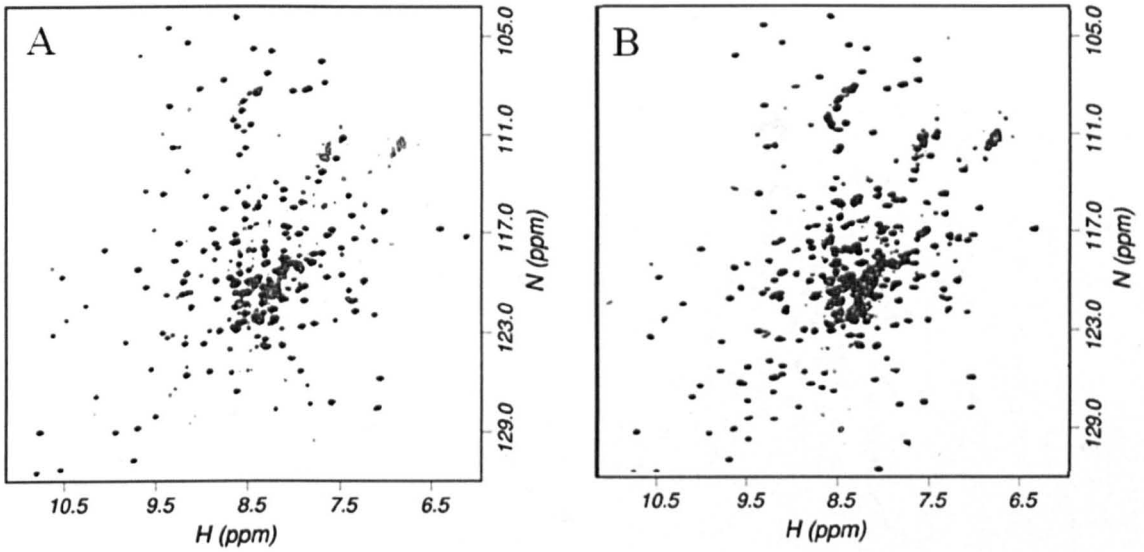


Figure 4.20: TROSY experiment carried out on (A) ^{15}N, ^{13}C, ^{2}H POR in sodium phosphate buffer pH 5.5 in Birmingham at 900 MHz (B) ^{15}N, ^{13}C POR in sodium phosphate buffer pH 5.5 in Sheffield at 800 MHz.

Following recommendations on the excellent quality of spectra achieved, the use of the 900 MHz spectrometer in the EU NMR facility in Utrecht was investigated. Since there was still an element of doubt associated with the quality of the deuterated sample produced, along with the fact that a significant amount of data had already been acquired using a double labelled sample as well as the time invested in analysing the data, it was decided to take a double labelled sample to Utrecht to acquire the final HNCACB experiment. Before acquiring the HNCACB data a HSQC TROSY experiment was run with the same parameters used to acquire the data in Birmingham. Upon processing the Utrecht data it was evident that working at higher field strength was not an issue as the quality of the spectrum achieved using a non-deuterated, ^{15}N, ^{13}C labelled sample (Figure 4.21 B) was significantly better than that obtained previously in Birmingham using a deuterated ^{15}N, ^{13}C sample (Figure 4.21 A).

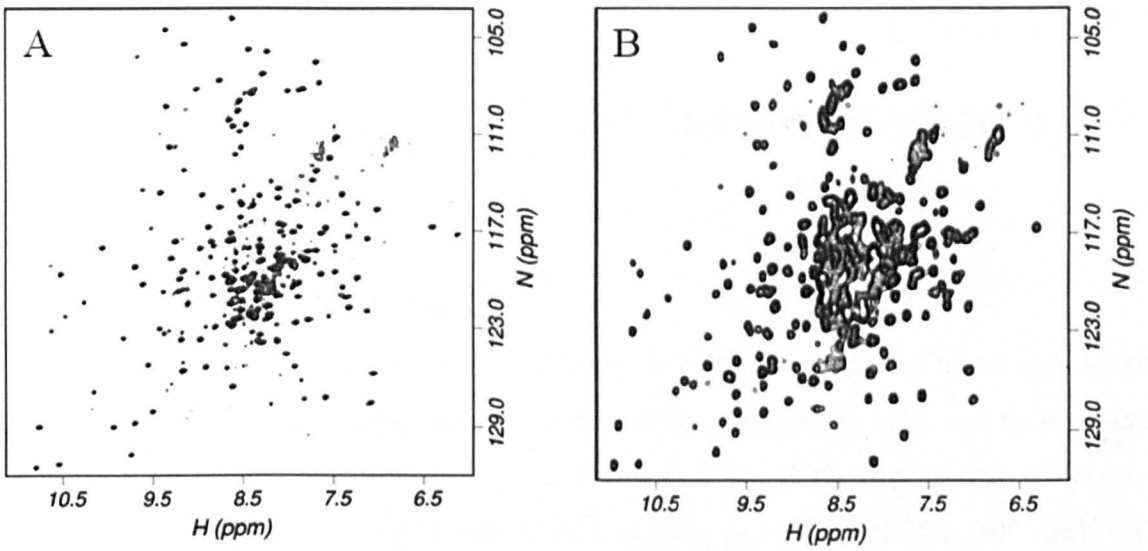


Figure 4.21: TROSY experiment carried out on (A) ^{15}N, ^{13}C, ^{2}H POR in sodium phosphate buffer pH 5.5 in Birmingham at 900 MHz (B) ^{15}N, ^{13}C POR in sodium phosphate buffer pH 5.5 in Utrecht at 900 MHz.

In summary, there remain unexplained problems with a deuterated protein sample, which currently limit the quality of spectra that can be achieved, and mean that there is little if any advantage to be gained by using a deuterated sample. If these problems could be solved, analysis of the NMR spectra would be much simpler. However there has not been time to do this during the PhD, and therefore the spectral analysis reported in Chapter 5 has been carried out on a non-deuterated, ^{15}N, ^{13}C labelled sample.

CHAPTER 5

NMR Studies of Thermosynechococcus POR

5.1 Introduction

To date all crystal trials set up with the enzyme POR have failed. The reasons for this remain unknown, however it has been proposed that the structure of POR can be elucidated using NMR. To do this, a 0.3 mM ^{13}C ^{15}N double labelled POR sample was prepared as detailed above and buffer exchanged into 50 mM sodium phosphate pH 5.5, 100 mM NaCl, 20 mM DTT with Roche complete protease inhibitor. Backbone assignment experiments (detailed Table 2.3) were conducted using a 600 MHz spectrometer except the HN(CA)CB experiment which was acquired using a 900 MHz spectrometer at the EUNMR facility in Utrecht. Both spectrometers were fitted with a cryopropbe and were preheated to 50 °C prior to acquisition of the dataset. All experiments acquired at 600 MHz were done over a time period of 24 hours, with the HN(CA)CB being acquired over a period of 4 days.

The signal obtained for the least sensitive experiments, specifically the HN(CA)CO and HN(CA)CB, was poor. Even at 900 MHz all peaks were still not present in the HN(CA)CB experiment acquired. A HN(CA)CO experiment was also acquired at 900 MHz using the equivalent parameters and pulse programme used to acquire the HN(CA)CB experiment, but still the level of signal achieved was poor. The assignment of POR was therefore done primarily using the Cα and Cβ data alone with the carbonyl data being used when present.

5.2 Peak Picking

All spectra were processed using Felix 2007 and in-house macros. All peaks that could be automatically picked in the HSQC spectrum 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 a database and the peaks were given a spin system number. This list, along with in-house macros, was used to go through all other backbone assignment spectra and select all peaks which corresponded to a single spin system (Figure 5.1). Proton carbon

planes were loaded for all spectra in the nitrogen plane, which corresponded to the nitrogen chemical shift of the spin system in question. 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.

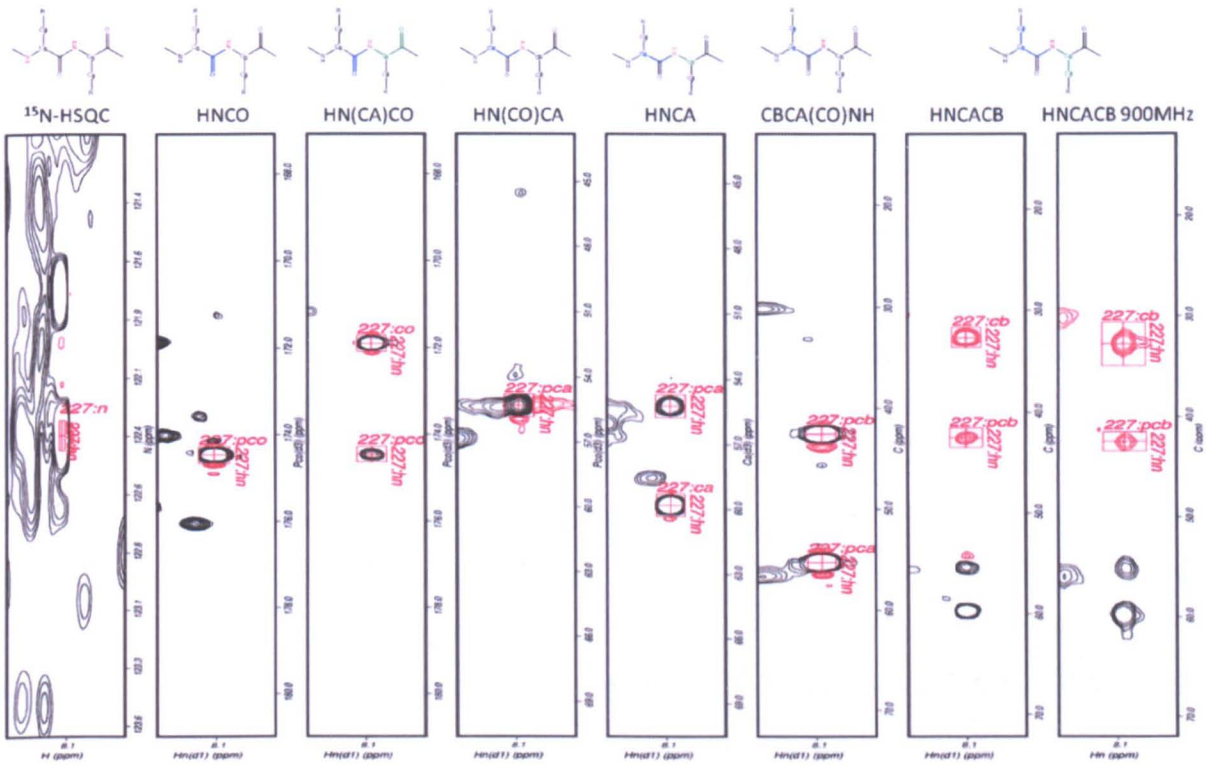


Figure 5.1: Proton Carbon planes from all spectra acquired for spin system 227: black peaks correspond to positive peaks, whilst red peaks correspond to negative peaks.

Although peaks were heavily overlapped in the HSQC spectrum the resolution of the peaks was somewhat improved in the HNCO spectrum. As a result the HNCO spectrum was used as the reference spectrum for characterisation of the spin systems, instead of the HSQC spectrum which is normally used. The referencing of the HSQC was adjusted so that well resolved peaks aligned with their corresponding peaks in the HNCO. Heavily overlapped peaks in the HSQC spectrum, which had a well resolved peak in the HNCO spectrum but had not been selected by the automatic peak picking process, were picked manually in the HSQC spectrum and added to the entity list previously created. The spin systems observed and picked in each spectrum are detailed in Table 5.1.

	Co	pCo	Cα	pCα	Cβ	pCβ
HNCO		X				
HN(CA)CO	X	X				
HN(CO)CA				X		
HNCA			X	X		
CBCA(CO)NH				X		X
HNCACB			X	X	X	X

Table 5.1: Table detailing the spin systems present and selected (Black), along with the spin systems present and not selected (Red) in each spectrum acquired.

As detailed in Figure 5.1 and Table 5.1 the peaks present in the HNCO and HN(CO)CA experiments correspond to the carbonyl and Cα respectively of the amino acid which is directly before the residue corresponding to the spin system in question. In all experiments acquired and analysis conducted in Sheffield, peaks which correspond to the previous amino acid are prefixed with a p, whilst peaks which correspond to the intraresidue are labelled without a p. Two peaks, corresponding to the previous amino acid and the intraresidue amino acid, are present at the same proton chemical shift in the HN(CA)CO and HNCA experiments acquired, whilst the two peaks present in the CBCA(CO)NH experiment acquired, correspond to the Cα and Cβ residue of the previous residue. Four peaks are present at the same proton chemical shift in the HNCACB experiment acquired; two are positive and correspond to the Cα and pCα whilst two are negative and correspond to the Cβ and pCβ. Only the negative peaks corresponding to the Cβ and pCβ are selected in the HNCACB spectrum as the Cα data present in the HN(CO)CA, HNCA and CBCA(CO)NH experiments have significantly more signal and are easier to unambiguously pick.

5.3 Assignment of POR

The chemical shifts of the spin systems selected were outputted to a spin system list. As detailed in Table 5.1, information for each nucleus can be obtained from more than one experiment. Normally whilst producing the spin system list information for each nucleus is taken from one experiment i.e. the pCo information is taken solely from the HNCO spectrum

even though information about the pCo can be found in the HN(CA)CO. However in the case of POR information for different nuclei could be found in different spectra. The spin system list was therefore generated using the information from all spectra but by specifying a hierarchy for selection of the values (Table 5.2). The hierarchy was decided according to the sensitivity of the spectrum acquired along with the number of peaks present in it. A pre-requisite was also set that for a picked spin system to be included in the spin system list, the spin system must contain information for the pCα nucleus from any or all of the three experiments which provide such information. A total of 418 spin systems were picked across all experiments, of which 275 fulfilled the criteria specified above and were used in the final spin system list. Excluding proline residues, this was 28 spin systems short of the number required to fully assign POR. Since the pCα peaks in the CBCACONH were very well resolved, the absence of the 28 spin systems was assumed to be a result of multiple spin systems overlapping each other in the experiments acquired.

	HNCO	HN(CA)CO	HN(CO)CA	HNCA	CBCACONH	HNCACB	HNCACB (900 MHz)
Co		1					
pCo	1	2					
Cα				1			
pCα			3	2	1		
Cβ						1	2
pCβ					1	3	2

Table 5.2: Table detailing the spin systems present in each experiment run along with the hierarchy in which the chemical shifts are selected to form the spin system list.

Even though information was taken from all spectra acquired to generate the spin system list, there was still a significant proportion of information missing from the list. A summary of the information present and missing for each spin system is detailed in Table 5.3.

CO		Cα		Cβ	
Criteria	**% spin systems**	**Criteria**	**% spin systems**	**Criteria**	**% spin systems**
Has a CO and a pCO	23	Has a Cα and a pCα	5	Has a Cβ and a pCβ	9
Has a CO and no pCO	0	Has a Cα and no pCα	9	Has a Cβ and no pCβ	21
Has no CO and a pCO	65	Has no Cα and a pCα	86	Has no Cβ and a pCβ	34
Has no CO and no pCO	12	Has no Cα and no pCα	0	Has no Cβ and no pCβ	36

Table 5.3: Table detailing the proportion of data present and missing in the spin system table for the carbonyl, Cα and Cβ nuclei.

The majority of information missing in the spin system list is about the Cβ and Co for the attached residue. This was not a surprise as the experiments which yield this information, the HNCACB and HN(CA)CO, are the least sensitive experiments, due to the relaxation experienced whilst passing the magnetisation through the attached Cα nucleus. There is also a proportion of spin systems which are lacking information for the attached Cα but have information for the preceding Cα residue; this may well be due to overlap in the spectrum. The HNCA is a fairly sensitive experiment but since the J coupling of the amide nitrogen is stronger to its own Cα residue (11 Hz) than that of its previous Cα residue (7 Hz) (Bax *et al.* 1993) you would expect that if a peak was present in the HNCA spectrum it would correspond to the Cα of the attached residue. The presence of pCα peak in the HN(CO)CA spectrum made the selection of the two Cα peaks in the HNCA spectrum slightly easier, however when the two peaks in the HNCA spectrum were overlapped, I could only confirm the presence of and pick the pCα peak in the spectrum.

A number of amino acids, specifically glycine, alanine serine and threonine residues have characteristic Cα or Cβ chemical shifts. The spin system list was also analysed to see how many of these amino acids had had their characteristic chemical shift selected (Table 5.4).

Amino Acid Sequence	Number Expected	Number Observed	Amino Acid Sequence	Number Expected	Number Observed
A	29	20	G	26	28
AA	2	1	GG	2	1
AG	1	3	GA	1	0
A[ST]	5	3	G[ST]	2	1

Amino Acid Sequence	Number Expected	Number Observed
[ST]	35	20
[ST][ST]	4	1
[ST]A	3	3
[ST]G	7	4

Table 5.4: Table detailing the number of amino acids present in POR which have unique chemical shifts, along with the number of amino acids observed in the spin system table.

Serine and threonine residues have Cβ chemical shifts which are generally greater than 60 ppm whilst alanine residues have Cβ chemical shifts which are generally smaller than 25 ppm. The Cα chemical shifts of glycine residues are generally smaller than 50 ppm, but characteristically due to the glycine side chain consisting of a proton, there is no Cβ chemical shift associated with glycine resides. As shown in Table 5.4 there are a number of characteristic chemical shifts missing from the spin system list. Although this ultimately means that these spin systems have not been selected from the data, it does not necessarily mean that data for that specific spin system is missing. Peaks which were present in the spectra acquired which could not be confidently assigned to a particular nucleus in a specific spin system, were not selected in an attempt to avoid incorrect assignments. Likewise when two sequential spin systems correspond to the same amino acid type, there is a distinct possibility that the information for these two spin systems is overlapped. At this stage, it is therefore very difficult to confidently pick the two peaks present.

The spin system list created along with the POR sequence was run through the Asstools simulated annealing assignment programme (Reed *et al.* 2003) 30 times, with a new randomly generated starting seed each time (detailed 2.22). Asstools uses a list of chemical shift ranges expected for each amino acid, derived from the data published in (Grzesiek *et al.*

1993b) along with the spin system list to assign each spin system to an amino acid. The system does not take into account the effect that the secondary structure has on the chemical shift of the amino acid and the amino acid range fed into Asstools was broad enough to cover the chemical shift of each amino acid in all secondary structures. Asstools uses a simulated annealing Metropolis Monte Carlo algorithm to minimise an energy function. The equation is made up of two components; one refers to the linking energy and the other refers to the binding energy. The linking energy describes how well two sequential chemical shifts match, whilst the binding energy describes how well the chemical shifts of the assigned spin systems match that which is predicted in the list of chemical shifts detailed above. Asstools uses a power law with 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, making 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. If this new output has a lower total energy function than the previous then it is accepted. 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 initial output from Asstools was not very encouraging as only 20 out of the 322 amino acids had been assigned to the same position 30 times (Table 5.5). There is reluctance in taking any assignment which has not been put in the same position 30 times as there is a risk that the wrong spin system would be assigned to the wrong amino acid. There was confidence however in the 20 assignments that Asstools had made, as all assignments made were sequential assignments scattered throughout the protein and mainly contained the characteristic amino acids detailed in Table 5.4. It was evident however that new strategies needed to be developed, other than the standard ones already in use in the lab, to use alongside Asstools to assist with the assignment of POR.

Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System
49	C	198	70	T	127	255	L	414
50	R	395	71	I	297	256	F	409
59	A	373	168	K	188	302	G	30
60	K	314	169	I	191	303	R	309
67	E	101	226	S	41	304	K	208
68	A	204	227	T	13	305	A	258
69	Y	85	228	G	71			

Table 5.5: Table detailing the spin systems which were confidently assigned to their corresponding residue (i.e. 30/30) in the first Asstools output.

5.4 Simulated Data

5.4.1 Homology models of POR

A homology mode of S*ynechocystis* POR was created by Townley *et al.* in 2001 using the crystal structures of the enzyme 7-α-hydroxysteroid dehydrogenase from *E. coli* as a structural template (Townley *et al.* 2001). The model consisted of 7 β-strands surrounded by 9 α-helices with the 33 residue insertion being modelled in at the top of the structure between the fourth β-strand and the sixth α-helix as a helix turn strand motif (Figure 5.2). The loop was modelled into this position to form an extension to the face of the substrate binding cleft containing the catalytic diad.

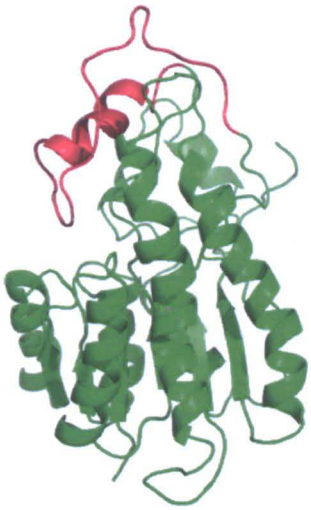


Figure 5.2: Original homology model of POR as detailed in (Townley *et al.* 2001).

Although there is a significant homology between the S*ynechocystis* and the *Thermorsynechocystis* POR sequences, structure prediction technology has advanced over the past decade and more structures now also exist. It was therefore decided to create a homology model for *Thermorsynechocystis* POR using two online structure prediction services; Phyre (Kelley *et al.* 2009) and I-Tasser (Zhang 2008; Roy *et al.* 2010).

Phyre uses known protein structures from the Structural Classification of Proteins (SCOP) database, along with new depositions in the Protein Data Bank (PDB) to generate a profile of the sequence in question. The system runs PSI-Blast five times to gather both close and remote homologous enzymes to the sequence in question, before combining all alignments into a single sequence using the sequence in question as the master. The secondary structure is then predicted using three secondary structure prediction programmes, with each programme providing a confidence value for the secondary structure predicted for each position of the sequence in question. The confidence values are averaged and used to generate a final consensus sequence. In addition to this another programme is run to predict which of the regions of the sequence are structurally ordered or disordered (Kelley *et al.* 2009).

I-TASSER (iterative threading assembly refinement) uses a four-step process to generate predicted secondary structures. Initially the programme identifies possible template proteins from databases of known structures, which have a similar structure or structural

motifs to that of the sequence in question. As with Phyre, a PSI-Blast is then run to identify evolutionary relatives before generating a sequence profile, which is used to predict a secondary structure using PSIPRED. The sequence is then compared to a representative PDB structure library with the top template hits being selected for further consideration. The model generated undergoes several rounds of optimisation prior to the production of a final prediction. I-TASSER also infers the function of the sequence in question, by matching the structure of the predicted model against that of proteins with known structure and function in the PDB (Zhang 2008; Roy *et al.* 2010).

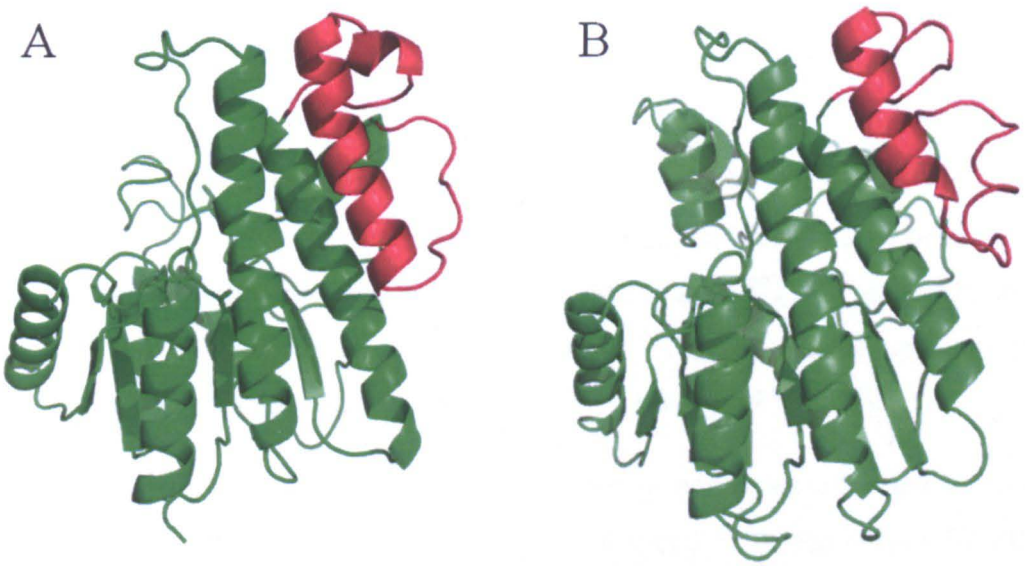


Figure 5.3: Models of POR generated using the secondary structure prediction software A) Phyre and B) I-Tasser.

As expected the models produced by Phyre and I-Tasser for *T. elongatus* POR resemble that produced by Townley and Sessions. Within reasonable error, the 7 β-sheets have all been placed in the same place and orientation, along with 6 of the predicted α-helices. There are however a couple of places where the models generated disagree with the original model created. I-Tasser and Phyre have both positioned the 33 residue insertion loop adjacent to the hydrophobic faces of the sixth and fourth α-helices. This makes more sense as it seems unlikely that the loop would position itself at the top of the protein exposing the large hydrophobic surface found on helix four and six. There are also regions of the models generated where there is no consistency in the structures generated. These regions

fall between Asn-115 to Gly-118, His-236 to Thr-252 and the C-terminus Gly-290 to Ala-322. The Phyre model has omitted these stretches of sequence in order to provide a better model, whilst I-Tasser has kept the residues in a similar secondary structure, but placed them in a complete different orientation at the top of the model.

It was decided that for future analyses we would take the consensus sequences of the three models generated and would therefore leave out the stretches of amino acids detailed above which have been placed in a different orientation each time. The homology model created by Phyre had essentially done this, so it was decided to use this model of *Thermosynechococcus* POR for all future analysis.

5.4.2 ShiftX 2 Chemical Shift Predictions

ShiftX 2 is a piece of internet available software which uses a PDB file along with information about the level of deuteration, as well as the pH and temperature of acquisition to predict the chemical shift of all backbone atoms (Han *et al.* 2011). The homology model generated for POR as detailed above was used as the PDB file, and chemical shifts were generated for a spectrum acquired at pH 5.5 at a temperature of 50 °C (Table 5.6).

The chemical shifts generated for the PDB file provided are expected to have RMS error values of 1.1 (^{15}N), 0.4 (^{13}Cα), 0.5 (^{13}Cβ) and 1.0 ppm (^{13}Co). However whilst this may be correct under normal circumstances, the values generated for our model are likely to have a larger degree of error, as the PDB file provided was a model. Even so the chemical shifts derived for the established secondary structure should be fairly accurate as there is quite a high level of confidence in those structural regions.

Residue Number	Residue	Cα	Cβ	Co
5	P	61.7	33.31	174.41
6	R	56.03	31.22	175.61
7	P	62.42	32.51	172.77
8	T	61.96	71.86	172.88
9	V	60.19	34.41	173.63
10	I	60.26	42.01	173.08
11	I	60.14	41.09	174.32
12	T	60.77	69.99	173.78
13	G	46.89		174.11
14	A	51.59	19.95	176.74
15	S	60.12	63.85	173.41
16	S	56.46	65.62	173.77
17	G	45.87		173.25
18	V	65.99	31.82	178.62
19	G	46.88		173.85
20	L	57.94	41.39	177.72
21	Y	61.86	38.62	175.36
22	A	55.34	18.54	177.19
23	T	67.29	68.69	173.51
24	K	60.42	32.84	178.02
25	A	55.39	18.84	177.93
26	L	57.44	41.8	178.35
27	A	55.54	19.51	178.02
28	N	55.05	38.44	175.68
29	R	56.13	30.94	175.04
30	G	45.73		171.31
31	W	56.52	31.6	173.96
32	H	56.78	31.35	173.51
33	V	61.27	34.05	173.59
34	I	59.99	38.51	169.61
35	M	56.44	32.85	170.98
36	A	49.66	21.88	174.31
37	C	57.59	32.01	172.76
38	R	58.5	29.68	175.15
39	N	51.93	38.33	174.81
40	L	58.07	41.82	178.09
41	E	59.68	29.52	178.01
42	K	59.64	32.65	177.38
43	A	55.5	18.75	177.6
44	E	60.2	29.68	178.19
45	Q	60.01	28.86	178.62
46	A	54.73	18.74	178.35

Residue Number	Residue	Cα	Cβ	Co
47	A	55.13	18.84	177.7
48	K	60.01	32.39	177.63
49	N	56.51	38.08	176.68
50	L	57.8	41.23	177.88
51	Q	58.22	28.69	178.61
52	I	63.7	37.83	177.05
53	P	62.91	32.21	176.94
54	P	63.64	31.67	174.42
55	E	54.62	32.34	175.42
56	A	50.45	19.85	175.68
57	Y	56.88	38.62	175.49
58	T	61.27	71.34	172.91
59	I	61.19	41.19	174.19
60	L	53.84	45.38	175.28
61	H	56.88	31.03	173.39
62	L	54.79	43.98	174.6
63	D	52.68	43.74	174.93
64	L	55.59	41.63	176.8
65	S	58.87	63.74	172.8
66	S	56.2	64.64	173.49
67	L	57.96	41.78	177.21
68	A	55.18	19.11	177.91
69	S	61.42	62.98	174.47
70	V	66.49	31.65	178.21
71	R	59.65	29.93	177.7
72	G	46.82		173.81
73	F	61.21	39	175.32
74	V	66.84	31.55	178.77
75	E	60.44	29.76	177.77
76	S	62.38	63.23	174.56
77	F	57.6	37.3	177.32
78	R	60.65	30.44	178.33
79	A	55.17	18.77	177.71
80	L	56.27	42.5	178.02
81	N	54.64	41.21	176.69
82	R	57.59	26.85	174.9
83	P	65.32	31.82	175.18
84	L	54.19	45.12	175.18
85	R	57.77	31.1	170.43
86	A	50.69	21.88	175.09
87	L	52.83	45.44	173.97
88	V	60.82	33.31	170.97

Residue Number	Residue	Cα	Cβ	Co
89	C	57.93	26.42	173.71
90	N	53.71	39.74	174.1
91	A	53.38	19.13	176.38
92	A	51.61	21.28	174.87
93	V	60.21	35.56	173.8
94	Y	57.09	42.24	173.32
95	Y	58.02	39.37	175.82
96	P	63.54	32.2	177.7
97	L	55.82	42.83	177.61
98	L	55.84	41.94	177.37
99	K	55.2	32.59	175.75
100	E	54.68	30.66	175.25
101	P	63.24	31.69	176.2
102	I	61.32	35.96	176.08
103	Y	59.33	38.55	176.11
104	S	62.59	63.15	174.93
105	V	65.2	31.35	177.98
106	D	57.27	40.85	177.63
107	G	46.69		173.46
108	Y	61.28	38.63	176.09
109	E	59.87	29.7	177.1
110	I	65.6	38.32	178.41
111	T	65.91	68.96	173.52
112	V	66.44	31.69	178.78
113	A	55.89	18.83	177.59
114	T	64.61	69.36	173.16
119	H	59.6	30.04	174.61
120	F	59.56	39.41	175.68
121	L	58.2	41.15	176.78
122	L	57.89	41.48	177.47
123	I	65.13	35.76	178.59
124	N	56.88	39.16	175.46
125	L	57.96	41.72	177.49
126	L	58.39	41.87	177.82
127	L	57.9	41.57	176.96
128	E	60.37	29.98	178.58
130	L	55.81	42.23	177.65
131	K	59.45	32.53	178.43
132	N	55.39	37.77	176.16
133	S	58.68	64.22	173.95

Residue Number	Residue	Cα	Cβ	Co
134	P	62.87	31.97	175.79
135	E	54.8	30.84	173.18
136	S	59.41	64.02	173.78
137	D	55.69	39.94	174.98
138	K	56.19	33.82	171.97
139	R	54.67	33.16	174.02
140	L	53.94	45.78	174.9
141	V	60.69	34.22	173.42
142	I	58.73	37.35	170.76
143	L	54.57	41.39	176.11
144	G	44.9		172.39
145	T	60.31	69.34	173.7
146	V	65.04	31.63	178.09
147	T	65.64	68.03	174.18
148	A	55.23	18.79	178.09
149	N	56.03	37.63	176.29
150	R	59.11	28.58	178.18
151	K	58.76	31.83	177.79
152	E	60.11	30.17	179.17
153	L	57.03	42.29	178.26
154	G	44.96		173.37
155	G	44.83		137.04
156	K	58.44	32.59	175.97
157	I	61.62	38.17	174.98
158	P	62.73	32.04	175.86
159	I	60.54	38.9	175.43
160	P	62.67	31.48	174.67
161	A	52.79	18.87	176.68
162	P	60.58	32.39	175.06
163	P	63.41	32.85	178.87
164	D	57.74	40.75	178.44
165	L	57.51	41.88	178.77
166	G	46.72		173.87
167	N	56.33	37.93	175.22
168	L	58.36	41.77	177.86
169	E	60.16	30.16	177.44
170	G	46.08		174.1
171	F	61.6	39.07	176.59
172	E	60.08	29.79	176.85
168	L	58.36	41.77	177.86

Residue Number	Residue	Cα	Cβ	Co
169	E	60.16	30.16	177.44
170	G	46.08		174.1
171	F	61.6	39.07	176.59
172	E	60.08	29.79	176.85
173	K	59.57	32.53	177.58
174	G	46.57		173.92
175	F	57.61	37.32	176.88
176	K	60.27	32.5	177.8
177	K	58.96	31.46	178.23
178	P	65.98	31.62	179.01
179	I	64.69	38	178.79
180	A	52.36	19.35	176.34
181	M	57.32	30.03	176.16
182	I	60.34	39.66	175.81
183	N	57.08	39.06	176.88
184	G	46.42		175.12
185	K	59.13	32.61	178.27
186	P	63.35	31.89	177.37
187	F	59.99	36.77	174.92
188	K	54.89	32.86	175.95
189	S	59.64	65.05	173.22
190	G	45.88		173.49
191	K	55	33.2	176.51
192	A	55.58	18.86	177.8
193	Y	60.61	38.86	177.19
194	K	60.13	32.62	178.1
195	D	57.54	41.54	177.72
196	S	61.7	63.1	174.49
197	K	58	30.6	177.83
198	L	58.67	41.55	177.86
199	C	63.68	27.39	173.42
200	N	56.79	40.26	175.89
201	M	59.36	32.53	178.42
202	L	57.75	41.49	176.67
203	T	67.11	68.89	173.95
204	A	55.83	19.41	177.46
205	R	60.08	29.94	179.09
206	E	60.21	29.15	179.21
207	L	57.66	40.29	177.54
208	H	60.57	29.92	175.33
209	R	59.92	29.83	178.13
210	R	59.14	30.24	178.66

Residue Number	Residue	Cα	Cβ	Co
211	F	57.79	37.95	176.83
212	H	57.14	29.91	174.51
213	E	58.84	30.17	176.57
214	S	58.04	65.36	173.46
215	T	65.07	69.9	172.99
216	G	45.04		171.69
217	I	61.99	38.08	173.95
218	V	60.88	34.39	173.85
219	F	57.06	42.65	171.84
220	N	52.46	43.16	172.75
221	S	56	65.65	173.04
222	L	54.31	44.64	175.12
223	Y	56.93	41.34	174.67
224	P	62.19	31.56	176.64
225	G	45.24		172.41
226	C	57.29	27.22	171.94
227	V	61.33	32.9	175.12
228	A	52.24	18.97	176.3
229	D	55.22	40.55	175.14
230	T	60.89	71.67	173.27
231	P	64.65	31.69	178.82
232	L	57.88	41.79	178.27
233	F	59.54	40.47	174.78
234	R	57.4	30.7	176.11
235	H	56.63	31.11	175.01
253	G	45.2		171.54
254	G	45.17		172.72
255	Y	60.76	39.48	174.34
256	V	59.92	33.35	170.44
257	S	57.19	64.78	173.51
258	Q	60.38	27.83	178.76
259	E	60.47	28.63	178.1
260	L	57.55	41.61	180.45
261	A	54.93	18.6	177.67
262	G	45.95		173.39
263	E	60.74	30.1	178.13
264	R	59.13	28.28	178.17
265	V	66.36	31.63	177.59
266	A	54.84	18.73	176.92
267	M	57.95	32.08	177.84
268	V	65.14	31.53	176.36
269	V	64.97	31.79	174.52

Residue Number	Residue	Cα	Cβ	Co
270	A	52.28	19.07	176.37
271	D	55	42.19	174.98
272	P	62.04	32.8	176.75
273	E	58.05	30.46	175.55
274	F	56.78	42.38	174.07
275	R	58.8	30.86	175.52
276	Q	55.72	30.3	175.41
277	S	57.57	65.47	173.42
278	G	46.84		172.84
279	V	60.52	33.15	170.95

Residue Number	Residue	Cα	Cβ	Co
280	H	55.74	31.91	173.62
281	W	56.48	32.29	174.16
282	S	57.74	65.34	173.74
283	W	58.04	26.28	175.8
284	G	45.78		172.53
285	N	52.55	41.88	173.83
286	R	55.3	30.34	175.24
287	Q	54.06	30.16	175.86
288	K	56.84	33.37	176.49
289	E	58.12	30.53	176.94

Table 5.6: Table of predicted chemical shifts generated using the Phyre homology model and Shift X2.

5.5 Simulated data sets

The 264 generated chemical shifts were used to create another spin system list, where the pCα, pCβ and pCo of one residue necessarily matched the Cα, Cβ and Co respectively of the previous residue. This new spin system list, along with an adjusted sequence string was fed into Asstools (detailed 2.22) and the assignment programme was run with 30 iterations, starting off with a new random seed each time. The spin systems which were analysed by Asstools had been numbered according to their position in the POR sequence. Since no information was given to Asstools to inform the programme of this, it would assign them in an unbiased way yet make the analyses of the simulated data sets easier. Scripts were written using 'nawk' to analyse how many spin systems had been assigned to their correct position in the sequence in all 30 iterations. Although 264 spin systems were inputted into Asstools, 18 of the spin systems correspond to proline residues which do not have an amide group so will therefore not be assigned in the HSQC spectrum. Therefore when analysing the data, any dataset which contains 246 correct assignments has fully assigned the sequence.

Following completion of the analysis by Asstools, the simulated data set assigned perfectly with every spin system being placed in the correct position every time. This model data set was then used to test how experimental deficiencies and errors would be expected to affect the assignments. As a first test, random amounts of noise (detailed Table 5.7) were

incorporated into the pCα, pCβ and pCo chemical shift values of each residue (detailed 5.4.2).

	Noise added (ppm)			
	pCα	**pCβ**	**pCo**	**Correct Assignments**
Noise 1	± 0.05	± 0.05	± 0.2	246
Noise 2	± 0.2	± 0.2	± 0.4	246
Noise 3	± 0.4	± 0.4	± 1.0	212
Noise 4	± 0.6	± 0.6	± 1.5	0

Table 5.7: Table detailing the level of noise randomly assigned to the fake spin system list, along with the number of correct assignments which Asstools made using that data.

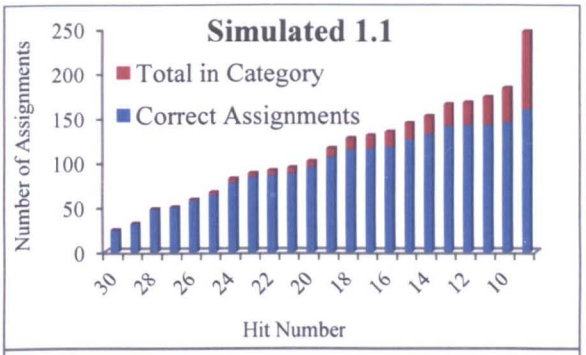
As expected as more noise is incorporated into the dataset the number of correct assignments made starts to decrease, to the point that no correct assignments are made whilst assigning the noise 4 dataset. The level of noise present in the real POR dataset ranges from the level of noise present in noise 1 to noise 3 (Table 5.7), so provided that a full dataset of chemical shift values exists it should be possible to assign the POR spectrum. This is not the case however. As detailed in Table 5.3 there are a significant number of chemical shifts missing from the POR spin system list. It was therefore deemed worthwhile to simulate datasets which had chemical shifts randomly removed in the proportions found experimentally (detailed Table 5.3) with a random amount of noise present (detailed Table 5.8). Scripts were written using 'nawk' to fulfil the criteria stipulated in Table 5.3, and using the noise detailed in Table 5.8, 10 datasets were created for each simulation, using a different random starting seed for each dataset.

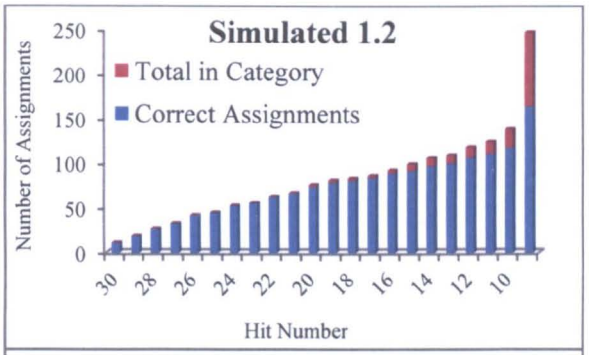
	Noise added (ppm)		
	pCα	pCβ	pCo
Simulated 1	± 0.1	± 0.1	± 0.4
Simulated 2	± 0.15	± 0.15	± 0.6
Simulated 3	± 0.2	± 0.2	± 0.8

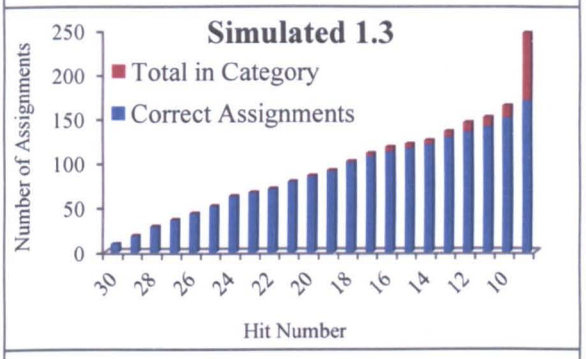
Table 5.8: Table detailing the level of noise randomly assigned to the three different simulated data sets.

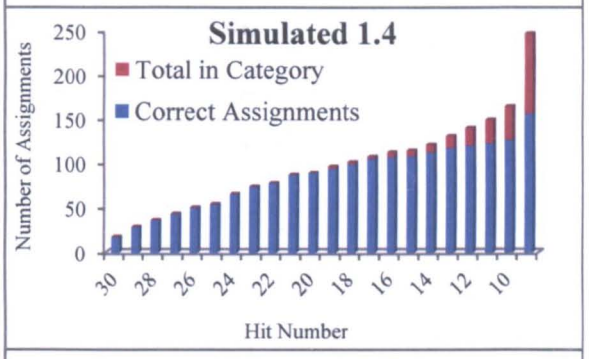
Each dataset was analysed in Asstools (detailed 2.22) using 30 iterations, with a new random seed each time. As detailed above the number of correct assignments made was analysed using scripts written in nawk and the number of correct and incorrect assignments made in each dataset was plotted on a histogram (Figure 5.4). Upon doing this it was evident that although the spin systems assigned in each run were different, every spin system which had been placed in the same position 30 times had been assigned correctly. It was therefore decided to investigate the number of spin systems which had been placed in the same position 29 of the 30 times and so on right the way down to 10 of the 30 times. Along with the number of correct assignments made, the number of spin systems that had been incorrectly assigned was also plotted on the same histogram (Figure 5.4).

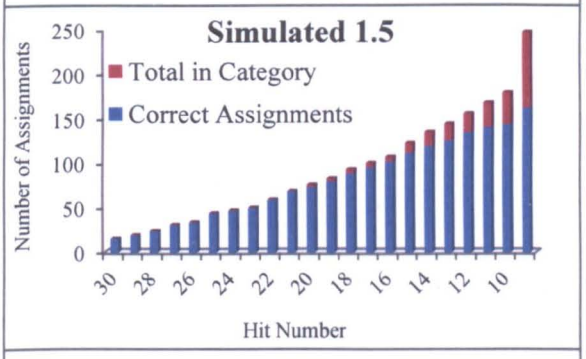
Following analysis of the simulated datasets, it was evident that for all datasets analysed, spin systems that had been placed in the same position for 29 of the 30 iterations had also been assigned correctly. In fact for 25 of the 30 datasets analysed, no significant error in the assignment of spin systems was noticed until you look at the spin systems which had been assigned to the same place in 26 of the 30 iterations run. It was therefore decided to fix the spin systems in the real POR dataset, which had been assigned to the same position in 27 or more of the 30 iterations run. Asstools was run again, fixing any spin system which had previously been assigned to the same position 27 times or more; this process was repeated with each output until no more assignments were made. Following this the number of assignments deemed correct and fixed had increased from 20 to 46 (Table 5.9).

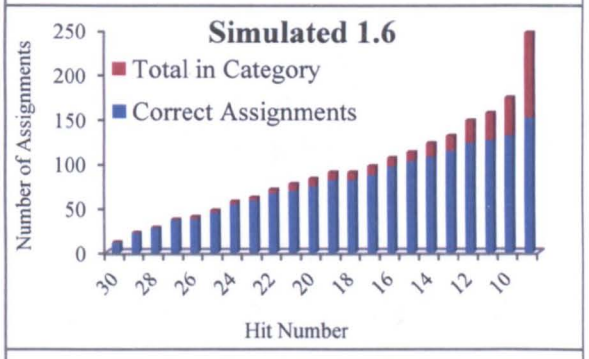

Simulated 1.1
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

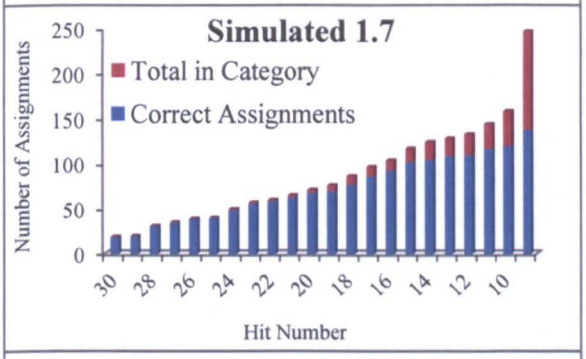

Simulated 1.2
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

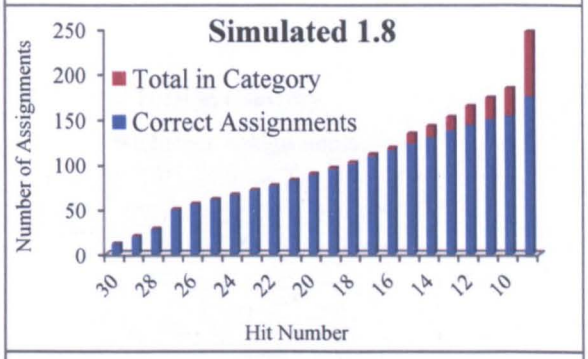

Simulated 1.3
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

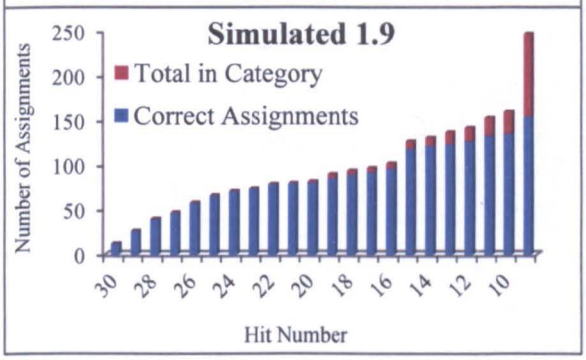

Simulated 1.4
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

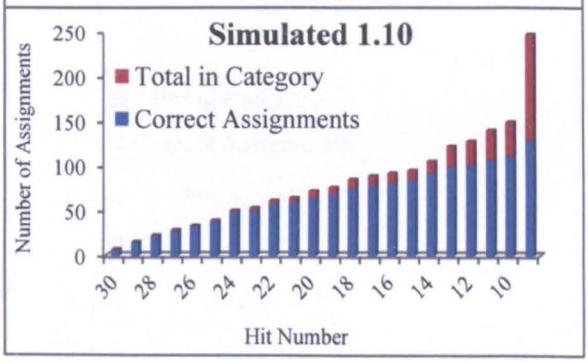

Simulated 1.5
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

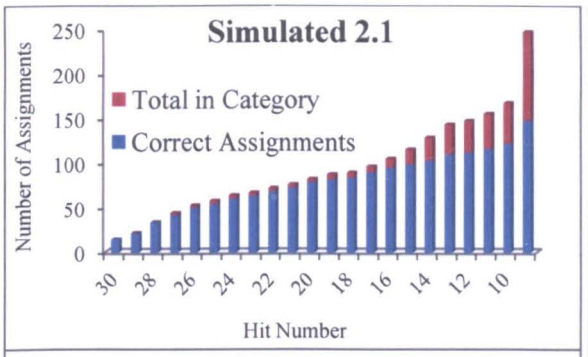

Simulated 1.6
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

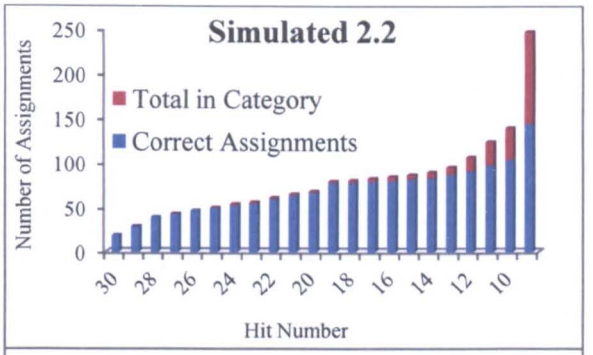

Simulated 1.7
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

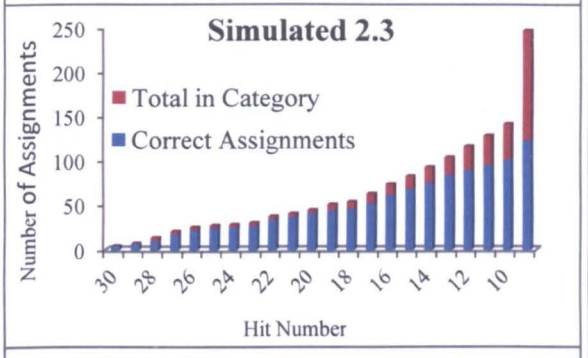

Simulated 1.8
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

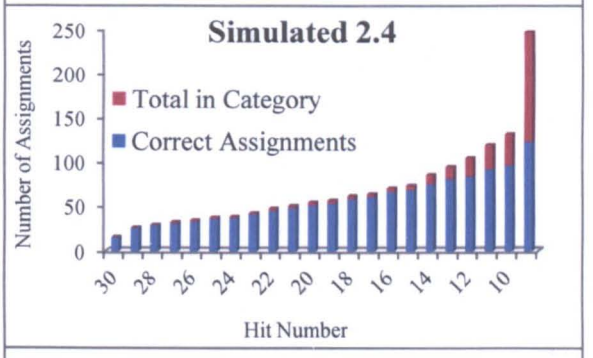

Simulated 1.9
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

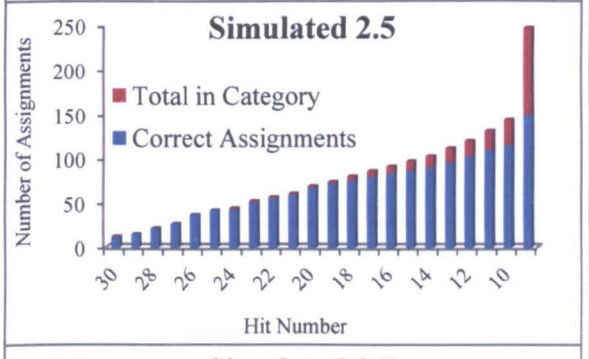

Simulated 1.10
Total in Category
Correct Assignments
Number of Assignments
250
200
150
100
50
0
30 28 26 24 22 20 18 16 14 12 10
Hit Number

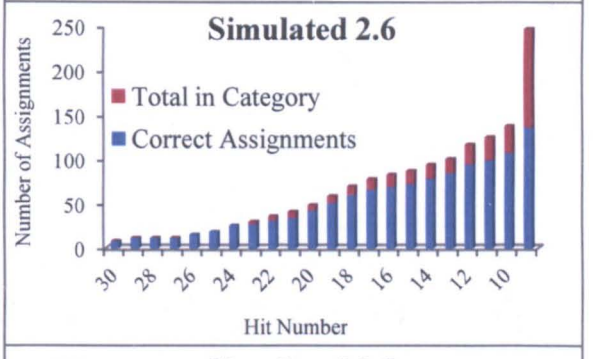

Simulated 2.1
Total in Category
Correct Assignments
Number of Assignments
Hit Number

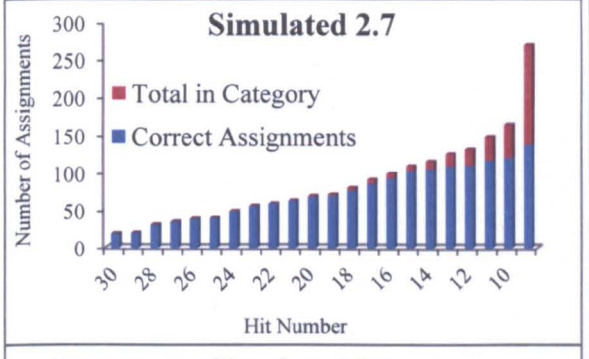

Simulated 2.2
Total in Category
Correct Assignments
Number of Assignments
Hit Number

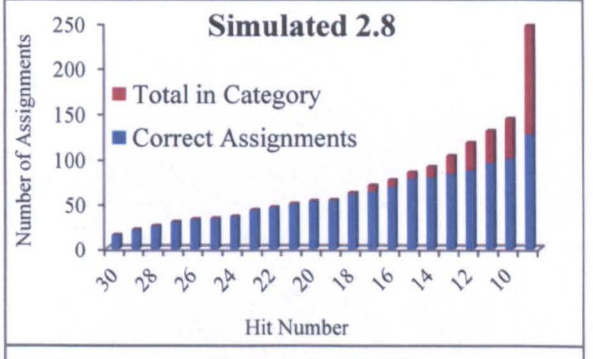

Simulated 2.3
Total in Category
Correct Assignments
Number of Assignments
Hit Number

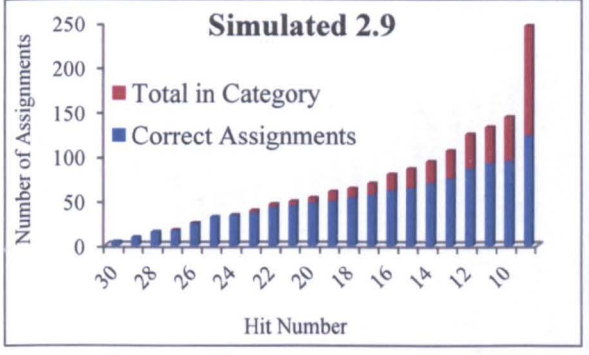

Simulated 2.4
Total in Category
Correct Assignments
Number of Assignments
Hit Number

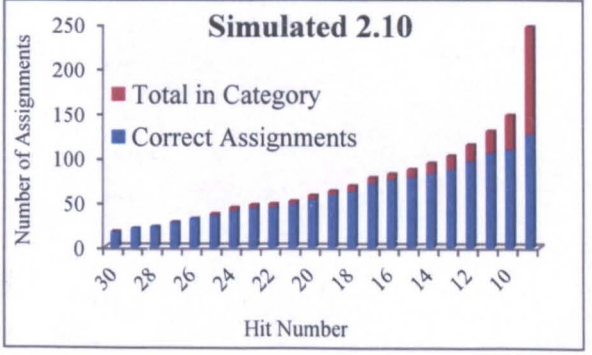

Simulated 2.5
Total in Category
Correct Assignments
Number of Assignments
Hit Number


Simulated 2.6
Total in Category
Correct Assignments
Number of Assignments
Hit Number


Simulated 2.7
Total in Category
Correct Assignments
Number of Assignments
Hit Number


Simulated 2.8
Total in Category
Correct Assignments
Number of Assignments
Hit Number


Simulated 2.9
Total in Category
Correct Assignments
Number of Assignments
Hit Number


Simulated 2.10
Total in Category
Correct Assignments
Number of Assignments
Hit Number

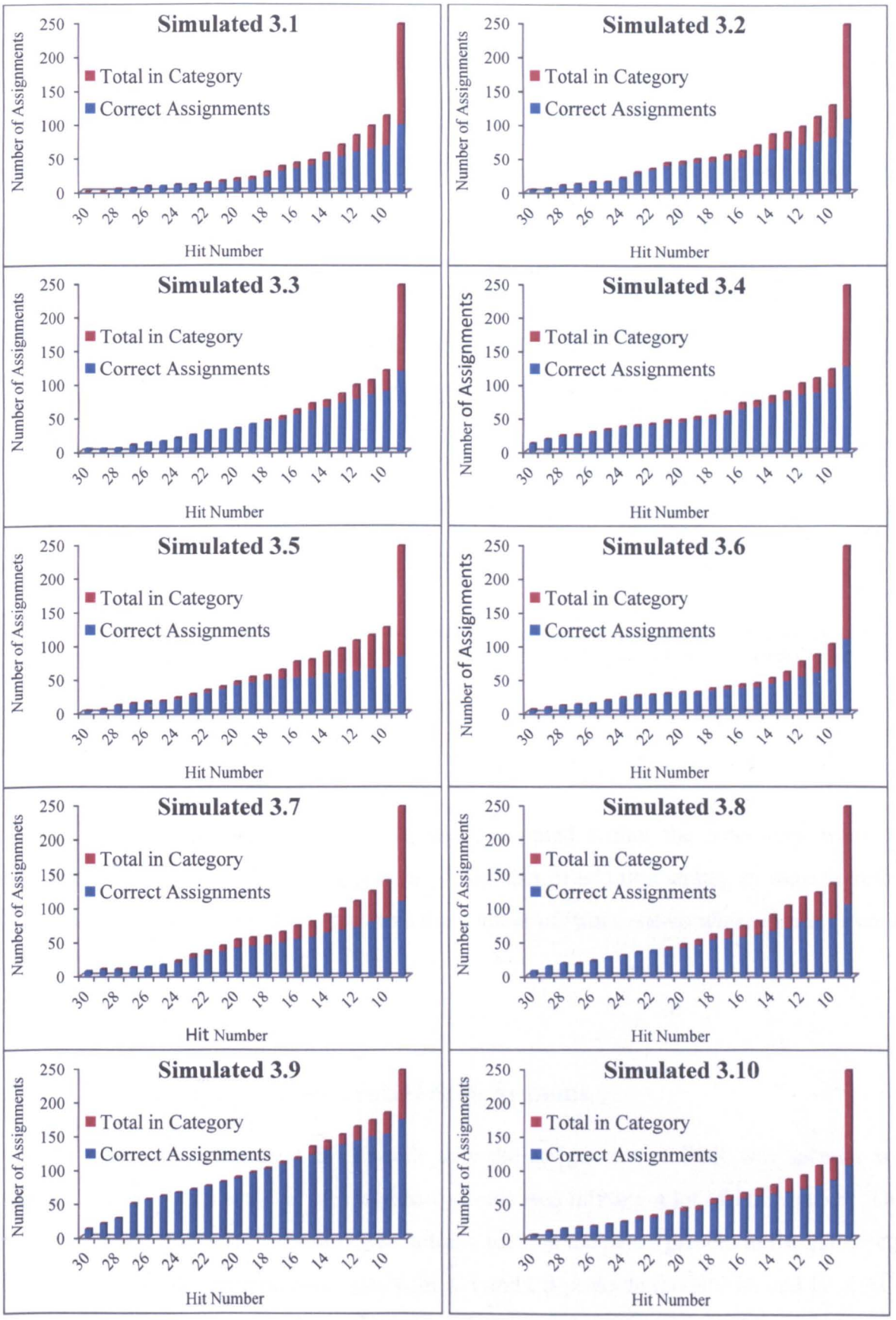


Figure 5.4: Histograms showing the number of correct assignments (blue) along with the total number of incorrect assignments made (red) for simulated datasets 1, 2 and 3.

Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System
10	G	36	93	N	84	255	L	414
49	C	198	94	R	147	256	F	409
50	R	395	163	K	209	258	L	336
54	K	315	164	E	195	259	F	162
55	A	207	165	L	381	260	Q	384
56	E	150	166	G	20	264	T	385
57	Q	134	167	G	17	265	G	48
58	A	333	168	K	188	266	G	310
59	A	373	169	I	191	301	E	189
60	K	314	196	G	6	302	G	30
67	E	101	226	S	41	303	R	309
68	A	204	227	T	13	304	K	208
69	Y	85	228	G	71	305	A	258
70	T	127	242	T	92	333	L	177
71	I	297	253	Q	157	334	A	288
			254	K	159			

Table 5.9: Table detailing the spin systems which have been confidently assigned to their corresponding residue using the new parameters. Entries in Red indicate spin systems which had previously been assigned.

5.6 Novel Strategies

As detailed above, the normal strategies adopted within the laboratory were not sufficient to fully assign the POR spectrum. A number of addition strategies were therefore developed and utilised, in order to improve the number of spin systems which Asstools could confidently assign.

5.6.1 Comparison of Sequential Spin Systems

It was clear that one of the reasons why the assignment of POR was so poor was because the list of spin systems provided to Asstools was missing a lot of information. One of the reasons for this was because whilst initially picking the peaks present in the 3D spectra acquired, some peaks present, particularly the Cα and Cβ peaks in the HNCA and HNCACB spectra respectively, were barely above the level of the noise. As a result they could not be picked with any confidence. A comparison macro, .comp, was written which loaded HN(CO)CA, HNCA, CBCA(CO)NH and HNCACB strips in Felix from two spin systems.

Since sequential residues share the same Cα and pCα, along with Cβ and pCβ frequencies, it should be possible to compare stretches of spin systems and thus pick weaker peaks with confidence, when previously they had been overlooked.

The assignment output distribution file acquired from running Asstools for the first time, provided a starting point as to which spin systems to compare. Possible sequential spin systems for example Asp-195 and Gly-196 (Figure 5.5) were loaded into Felix and using the pCα and pCβ information from the spin systems linked to Gly-196 as a guide, any weak Cα and Cβ peaks present in the HNCA and HNCACB spectra respectively could be picked for Asp-195. Likewise the same process could be used to pick pCα and pCβ using the Cα and Cβ frequencies of the previous residue as a guide.

In the case of Figure 5.5 a peak was present in the HNCACB spectrum corresponding to spin system 115; however there was no Cβ information present in any of the other spectra acquired. Although one would expect the Cβ peak present in the HNCACB spectrum to be more intense than the pCβ peak there are cases where this is not true. As a result peaks present which could not be confidently assigned were noted but not picked. After running Asstools, spin system 115 was assigned to Asp-195 in 23 of the 30 runs, whilst spin system 6 was assigned to Gly-196 in 28 of the 30 runs. Since one would expect no Cβ information to be present for a glycine residue, the peak present in the HN(CA)CB spectrum corresponding to spin system 6 must be a pCβ peak. The frequency of this pCβ peak aligns exactly with the frequency of the peak in the HN(CA)CB spectrum belonging to spin systen 115, therefore the peak present must be a Cβ peak. This was enough to ensure that in all following Asstools assignments, spin systems 115 and 6 were assigned to residues 195 and 196 with a much higher degree of confidence.

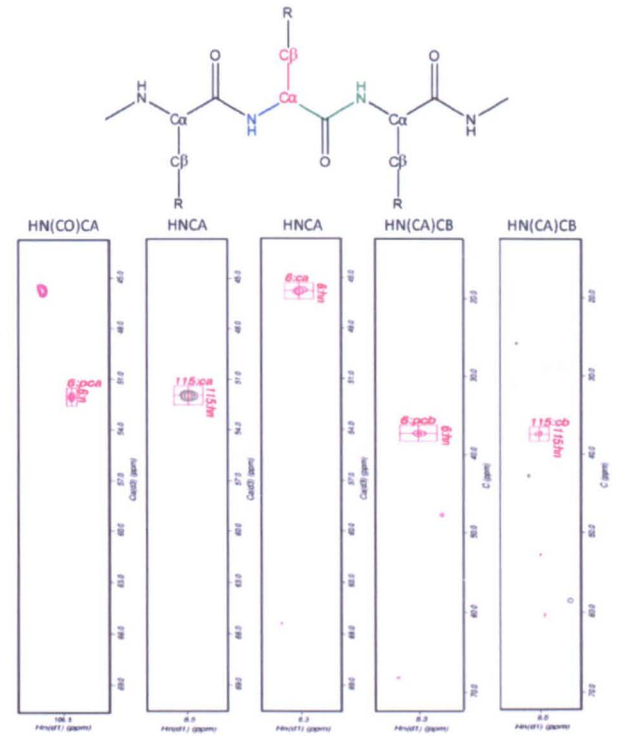


Figure 5.5: Proton carbon planes from all spectra used by the programme find_me to assign the single peak in the HN(CA)CB spectrum of spin system 115 as the intraresidue Cβ.

Having gone through and compared all possible sequential spin systems and picked all the additional peaks present, all spin systems were dumped out of Felix again and a new spin system list was produced (as described 5.3) This new spin system list was used for all future assignment runs of POR using Asstools.

5.6.2 Structure Based prediction

The programme 'find_me' (detailed Apendix B) was written using nawk, to compare the chemical shift values of the spin systems picked, along with the predicted chemical shift vales calculated on the basis of the homology model (detailed 5.4.2). The programme calls on three data files; these are a list of residues along with their predicted chemical shift values, a spin system list (detailed 5.3) and the most recent Asstools assignment output distribution file. The programme was developed using the spin systems which had already been assigned to residues.

To assign Lys-163 the command find_me 163 was typed into a linux shell, which was working in the directory where all the above files were located. Upon running this command, find_me searches through the spin system list for all spin systems which have chemical shifts within the specified range of the predicted chemical shifts for that residue, along with spin systems that have previous chemical shifts within the specified range of the predicted chemical shifts of the previous residue. The error which was applied to the predicted chemical shift values was specified in the scripts "current" and "previous" prior to running the programme. All spin systems which fall into these criteria are outputted into two lists which are then compared to each other; all spin systems which are present in both of these lists are then outputted to one final file. Following the formation of this file, a list is generated using the most recent Asstools assignment output distribution of all spin systems which were used to assign that residue in all 30 iterations. The spin systems which appear in this list are then compared to the spin systems present in generated file and any spin systems which appear in both are outputted to the screen. Providing the model is accurate, which we believe it to be in the conserved secondary structure, any output generated corresponds to plausible assignments for that residue.

Find_me was run for all residues four times, incorporating increasing levels of error (detailed Table 5.10), each time outputting spin systems which have chemical shifts within the specified range of the predicted chemical shifts.

Run Number	Incorporated Error (ppm)
1	± 1.5
2	± 2.0
3	± 3.0
4	± 4.0

Table 5.10: Table detailing the level of error incorporated into the predicted chemical shifts during the four iterations of the programme find_me.

Interpretation of the results outputted from the four runs was done with care. Any spin systems which were identified during the first run, were generally accepted to be correct assignments, providing that the spin system identified had been assigned to the corresponding

amino acid, more than 20 times during the latest Asstools assignment. In order for a spin system to be assigned during the latter runs of find_me, additional criteria needed to be fulfilled prior to selection. Spin systems were only assigned from the second run, providing a sequence of at least three consecutive residues were identified to have the same error associated with the predicted value, thus suggesting a systematic error had occurred in the positioning of the sequence, in the model used to generate the simulated data set. The spin systems outputted from runs three and four of find_me appeared to have a significant proportion of error associated with them, therefore to avoid the incorporation of incorrect assignments, none of the spin systems outputted in the third and fourth runs were used in the assignment.

5.6.3 Final Assignment of POR

Having fixed all spin systems previously assigned using Asstools and find_me, Asstools was run for the final time (detailed 2.22) using the new spin system list fixing all spin systems which were assigned to the same position 27 times or more. Table 5.11 details the residues which have been assigned in POR and Figure 5.6 details where on the model these residues are situated. This final set of assignments is somewhat conservative; on the other hand, it is clear that using the data available we will never be able to assign the protein completely, and there seems little point in broadening the criteria to make a relatively small number of potentially incorrect assignments, because we will not be able to trust them. These assignments therefore represent all the backbone assignments that we are able to make with confidence.

Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System	Residue Number	Residue	Assigned Spin System
10	G	36	74	L	289	228	G	71
25	G	67	93	N	84	242	T	92
26	A	131	94	R	147	253	Q	157
27	S	9	163	K	209	254	K	159
48	A	306	164	E	195	255	L	414
49	C	198	165	L	381	256	F	409
50	R	395	166	G	20	258	L	336
54	K	315	167	G	17	259	F	162
55	A	207	168	K	188	260	Q	384
56	E	150	169	I	191	263	I	190
57	Q	134	171	I	202	264	T	385
58	A	333	173	A	250	265	G	48
59	A	373	176	D	341	266	G	310
60	K	314	177	L	158	289	S	165
61	N	234	178	G	7	290	G	8
62	L	231	179	N	137	291	V	139
67	E	101	191	I	305	301	E	189
68	A	204	192	A	299	302	G	30
69	Y	85	193	M	75	303	R	309
70	T	127	195	N	115	304	K	208
71	I	297	196	G	6	305	A	258
72	L	279	226	S	41	333	L	177
73	H	185	227	T	13	334	A	288

Table 5.11: Table detailing all spin systems which have been confidently assigned to their corresponding residue.

Figure 5.6: I-Tasser model of POR detailing the positions of the assignments made (Red).

As shown in Figure 5.6 the majority of assignments made are on the outside of the protein, in mainly alpha helical and unstructured random coil regions. This seems very reasonable because such regions could well be rather more mobile than the structural core. Interestingly, many of the residues assigned are located in the 33 residue insert present in POR. In the model, these residues appear to be in the unstructured part of the loop however by looking at these residues it may be possible to determine if the loop is a rigid structure or if it has some mobility.

5.6.4 Correlation Time of POR

^{15}N T_1 and T_2 relaxation experiments were conducted in a 600 MHz spectrometer which had been pre-heated to 50 °C, using a 0.3 mM ^{15}N POR sample in 50 mM sodium phosphate pH 5.5, 100 mM sodium chloride, 20 mM DTT and complete protease inhibitor. The relaxation delays in the experiments run were as detailed in Table 5.12. All spectra were processed, using in-house macros, and the intensities of the assigned peaks in the five spectra acquired were plotted against time (Figure 5.7 & Figure 5.8). All plots were fitted to an exponential decay using a Levenberg-Marquardt least-squares algorithm and the respective T_1 and T_2 times of all assigned residues were calculated. The calculated T_1 and T_2 value for each assigned residue were plotted against each other and used to determine a correlation time of POR (Figure 5.9).

Experiment Number	T_1 Relaxation Delay (ms)	T_2 Relaxation Delay (ms)
1	40	16.144
2	180	40.360
3	360	80.720
4	640	121.08
5	1200	201.80

Table 5.12: Table detailing, in ms, the relaxation delay applied to each experiment run to determine the correlation time of POR.

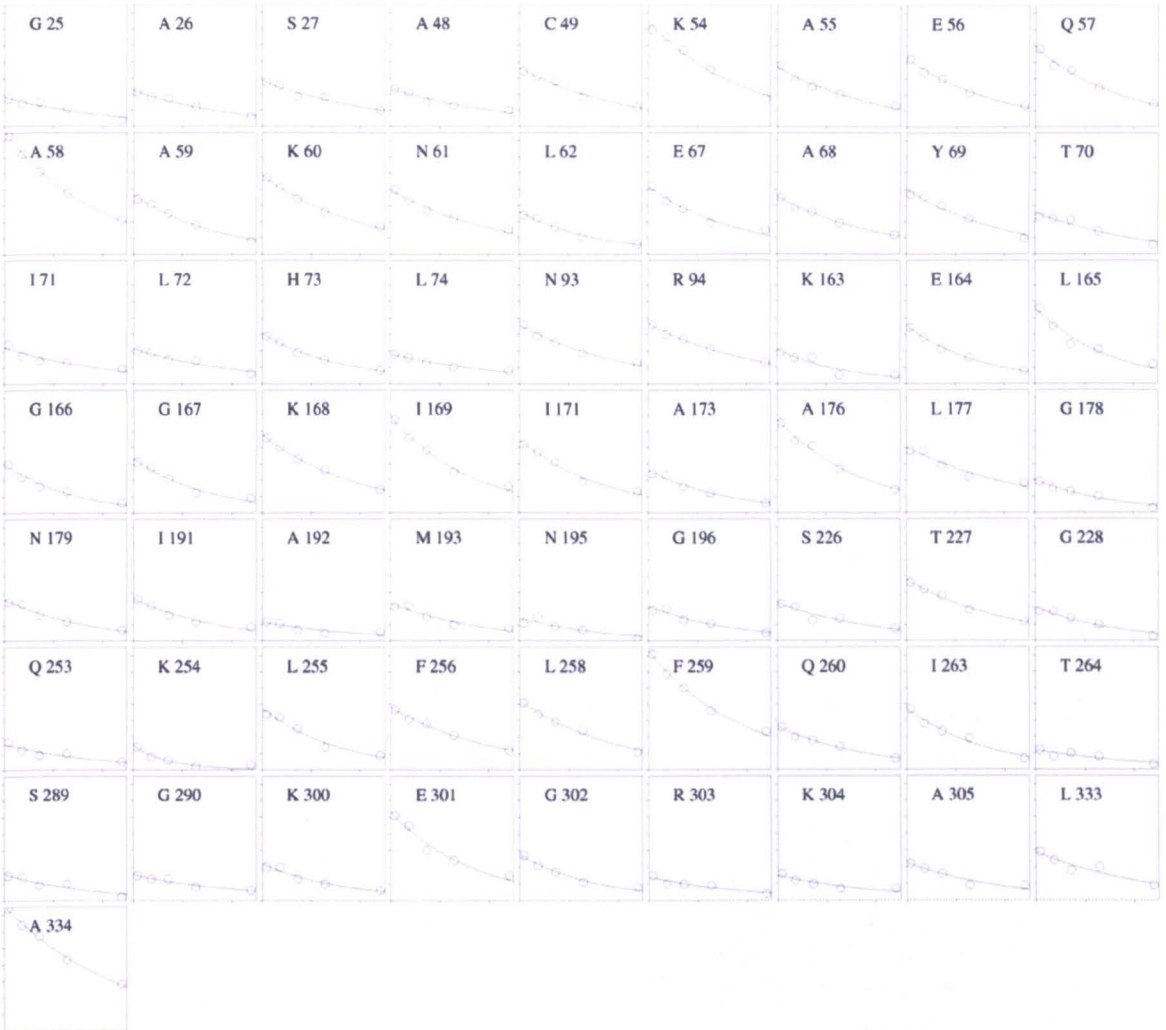


Figure 5.7: Fitted T_1 plots of all assigned peaks, with the relaxation delay being plotted along the x-axis and the remaining intensity of the peaks being plotted on the y-axis. The range on the x-axis is from 40 to 1200 ms as detailed in Table 5.12.

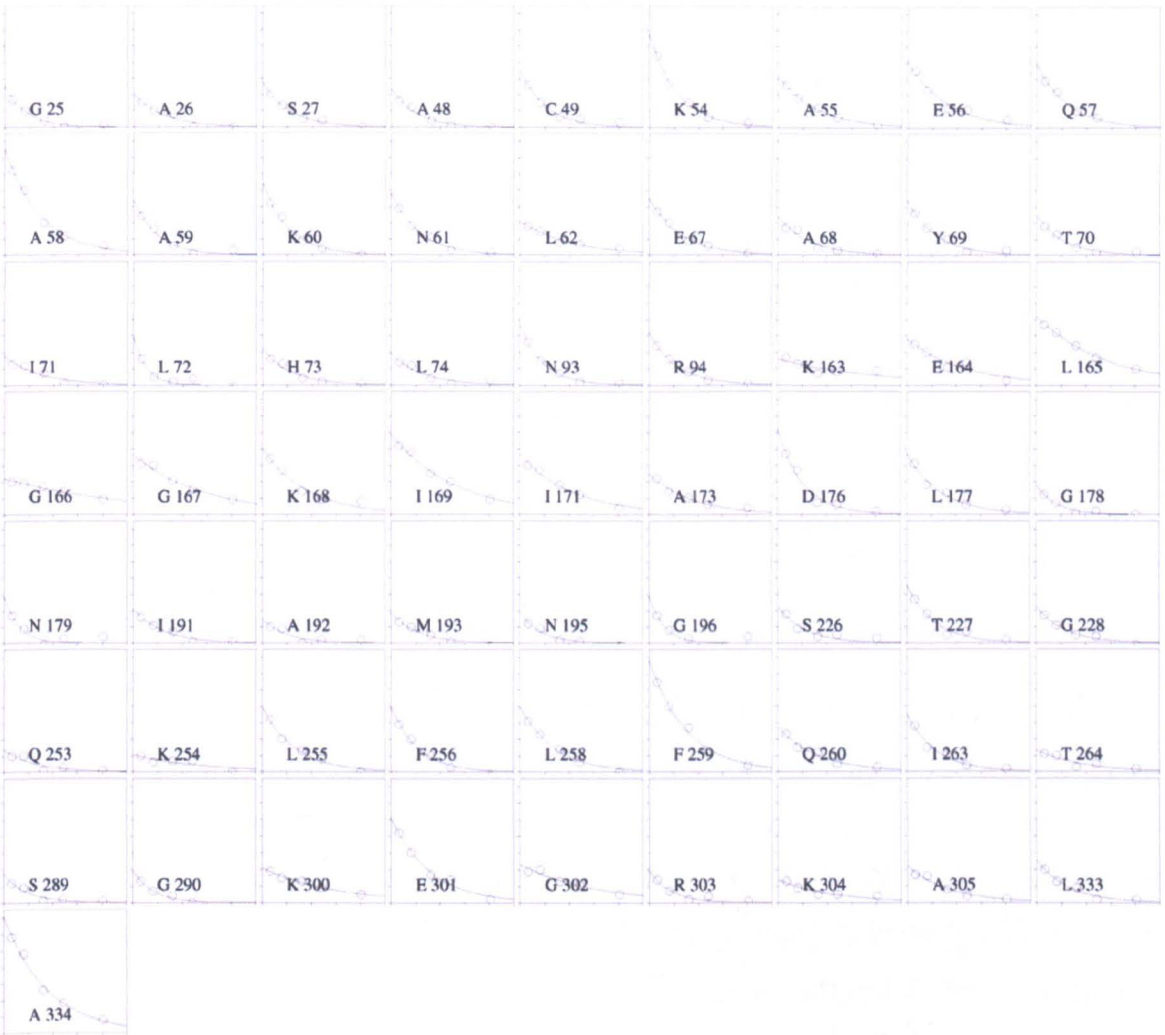


Figure 5.8: Fitted T_2 plots of all assigned peaks, with the relaxation delay being plotted along the x-axis and the remaining intensity of the peaks being plotted on the y-axis. The range on the x-axis is from 16.144 to 201.80 ms as detailed in Table 5.12.

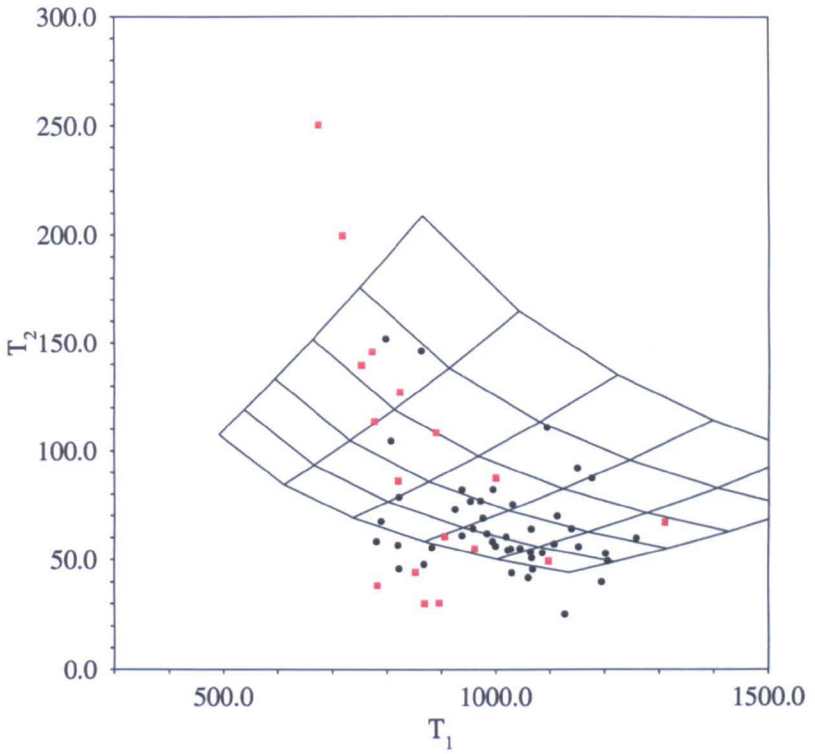


Figure 5.9: Graph showing the fitted T_1 value for each assigned peak plotted against its corresponding T_2 value. Points corresponding to peaks predicted to be present in the mobile loop are plotted as red squares. The lines on the grid represent order parameters (S^2) from 1 to 0.5 from bottom to top, and correlation times from 0.8 to 1.8 μs from left to right.

It is apparent from Figure 5.9 there is a very wide range of correlation times present for all the assigned residues, however they do seem to cluster into two groups. In plots like these, a rigid globular protein with no internal mobility would be expected to have points grouped together somewhere near the bottom of the grid. Points below the bottom of the grid can only be due to residues with slow internal motion (with reduced T_2 values because of chemical exchange), while mobile regions such as chain termini would be expected to have longer T_2 times and therefore lie at smaller S^2 values. All residues which have been assigned to belong in the 33 residue insertion present in POR have been marked in red whilst the residues which belong to the main structure of the protein have been marked in black. At 50 °C the residues assigned in the main body of POR appear to have a correlation time ranging somewhere between 13 to 16 ns. Using Equation 4.1, it was possible to determine that this is what you would expect from a monomer of POR at that temperature. Thus, under our conditions, apo POR appears to be monomeric. The residues assigned to the loop however have in general a significantly faster correlation time ranging from 9 ns upwards, thus indicating that the loop region of POR in solution has a higher mobility than the rest of

the protein. This result seems very reasonable in view of the fact that the loop is a variable length region not found in many other homologous proteins and is therefore likely to be on the outside of the protein (as it is in the new model). This may also help to explain why crystal trials of POR have failed to date. If the 33 residue insertion loop present in POR is moving around independently of the rest of the protein whilst in solution then this would disrupt any crystal packing.

Several residues in the loop also have anomalously short T_2 values (Figure 5.9). Short T_2 values are normally indicative of motions that are slow on the chemical shift timescale, i.e. 10^4 s^{-1} or slower. Such motions are often of interest because this timescale is often similar to the timescale for release of bound substrate, possibly implying that movement in the loop may be linked to substrate binding or release. This idea is explored further in Chapter 6.

5.7 Discussion

A disappointingly small fraction of the residues of POR have been assigned so far. This can be ascribed at least in part to problems in producing deuterated POR, described in Chapter 4. A clear task for the future is to produce deuterated POR. The spectra obtained at 50 °C are nevertheless of reasonable quality, and show that the protein is monomeric. The number of spin systems identified here is similar to the expected number; the reason for the limited assignment is the low intensity of signals in the 3D spectra rather than missing spin systems. It therefore looks as if the protein is reasonably well-behaved, and there are not extensive regions that will not be assignable. If we are able to obtain spectra of slightly better signal-to-noise (probably by use of deuterated protein), the simulations reported above suggest that a much more complete assignment should be possible.

The assignment is based on the standard heteronuclear experiments. Some use has been made of shifts predicted from the model structure, but this is relatively limited in its scope because the model is not likely to be fully reliable especially in loop regions. It is therefore a reasonable procedure to use the assignments to see what can be said about the structure.

The program TALOS uses assigned shifts (Cα, Cβ, Co, N, H, although the largest influences come from Cα and Cβ) and uses them to predict the local structure, based on matches to databases (Cornilescu *et al.* 1999). It is not 100% reliable since there will always

be occasions in which the most likely structure that matches the shifts does not correspond to the actual (unusual) structure. However, TALOS is used widely to make structural predictions. TALOS was run on the assignments obtained here. Only 11 residues resulted in confident predictions Table 5.13. Where these residues are in regular secondary structures in the i-TASSER model, the TALOS prediction almost always generates backbone dihedral angles that agree with the model. Success rates in the loops are less good, as expected. Thus whilst the assignments made do not permit any kind of structural prediction, they do provide a strong indication that the i-TASSER model is basically correct. It is interesting to note that the TALOS predictions agree reasonably well with the predicted secondary structure in the loop, suggesting that the model for the loop could be correct at least in outline.

Amino Acid	Predicted 2° Structure	Actual Angles		Predicted Angles	
		Phi	Psi	Phi	Psi
R 50	Random Coil	-111 ± 32	147 ± 23	-74.41	-33.53
K 54	α-helix	-66 ± 7	-36 ± 14	-67.99	-36.47
I 64	Random Coil	-95 ± 33	118 ± 14	-66.24	-19.44
F 231	β-sheet	-119 ± 22	143 ± 16	-137	127.88
L 244	Random Coil	-72 ± 14	-25 ± 22	-54.82	-50.32
F 249	α-helix	-117 ± 29	122 ± 21		
L 251	α-helix	-65 ± 6	-40 ± 11		
M 279	α-helix	-65 ± 10	-42 ± 9	-51.62	-48.01
F 286	Random Coil	-128 ± 28	140± 29	-149.55	173.36
V 291	β-sheet	-113 ± 24	143 ± 26	-101.26	166.33
H 292	β-sheet	-123 ± 24	141 ± 24	-102.29	120.44

Table 5.13: Table detailing the assigned amino acids which have successfully had a structure assigned by Talos, along with the Phi and Psi angles assigned by TALOS and those predicted by the programme phipsi3.f (Williamson personal communication) using the iTasser model.

The Chemical Shift Index (CSI) program was also run (Wishart *et al.* 1992). This program predicts secondary structure, based on patterns of chemical shifts in at least 4 contiguous residues. There was no useful prediction, in large part because there were very few continuous assignments long enough to make CSI possible.

It so happens that a string of residues have been assigned at the start of the insert. Measurements of ^{15}N relaxation on POR show that some of these residues have strikingly longer T_2 times than the average (Figure 5.9). This implies that the loop is very likely to be more mobile than most of the rest of the protein. Moreover, some residues in the loop have short T_2 values, implying slower motions. Overall, the results therefore highlight the loop as being an interesting part of the protein, possibly involved in some way in substrate binding and release, although structurally it does not form part of the predicted substrate binding sites.

CHAPTER 6

Formation and Analysis of a POR-Pchlide-$NADP^+$ Ternary Complex

6.1 Introduction

The original aim of the project was to elucidate structures of POR with and without substrates and products bound. The apo POR protein was initially picked as, following preliminary experiments, it was thought that this gave the best spectra to work with, however as detailed above, we have been unable to assign the spectrum of apo POR alone. There are a number of reasons why this could be: the lack of improvement in spectral quality following deuteration being the significant factor, however the mobility of the 33-residue loop could play a factor in the relaxation of signals around that region. As detailed previously, the loop has many proposed functions, one of which being its suggested function in the binding of substrates (Wilks *et al.* 1995). If this were to be the case then conducting NMR on the ternary complex may well result in better quality spectra as the mobility of the loop may be reduced upon binding substrates.

It has been well documented that since POR is light activated the ternary complex can be made in the absence of light and if held in these conditions, is completely inactive (Heyes *et al.* 2005). The standard method developed for the delivery of Pchlide in the Professor Hunter lab involves the use of a detergent at concentrations above the detergent CMC. The Pchlide pigment is proposed to bind to the detergent molecules and get delivered into the enzyme active site (Heyes *et al.* 2002b). Whilst this has proved to be an effective method for steady state assays and kinetic analysis, the presence of detergent, in the concentrations required, would affect the acquisition of the NMR spectrum, as the presence of detergent molecules will produce noise and will increase the viscosity of the sample, reducing the T_2 times and thus degrading the signal to noise. It was therefore decided to produce a ternary complex sample which was free of detergent and thus analysable by NMR.

6.2 Purification of Protochlorophyllide (Pchlide)

Pchlide was purified from the bacterium *Rhodobacter capsulatus* ZY5 essentially as described in Heyes *et al.* 2003, although a number of amendments were made to the published protocol. *R. capsulatus* ZY5 was streaked out onto a VN agar plate, containing 25 μg/μl rifampicin, from a glycerol stock and incubated at 34 °C for three days. Single colonies were picked and used to inoculate semi-anaerobic VN starter cultures (detailed 2.14.1). Starter cultures were incubated at 34 °C, with agitation, until they were green in appearance; at this stage one starter culture was used to inoculate each 2 l VN semi-anaerobic growth. Six autoclaved polyurethane foam bungs were placed in each 2 l growth and incubated at 34 °C with agitation (detailed 2.14.2). *R. capsulatus* ZY5 has a single point mutation in one of the three subunits required by the bacterium for the light independent reduction of Pchlide (Heyes *et al.* 2003b). The mutation results in the accumulation of the pigment which is excreted by the bacterium into the medium and is then adsorbed onto the polyurethane bungs. After 24 hours the polyurethane bungs were removed and soaked in 100 % methanol to re-solubilise the pigment. Freshly autoclaved bungs were placed in the medium and the process was repeated until the bacteria stopped excreting pigment (detailed 2.14.3).

The extracted pigment was placed in a rotary evaporator and the methanol was boiled off under vacuum. The process was repeated until all the pigment had been dried down to the bottom of the sample flask. The sample was left under vacuum pressure overnight to ensure all solvent was evaporated off the sample. 50 ml of methanol was placed in the sample flask and was mixed until it was saturated with pigment. The methanol was removed and made up to 1 l using pure acetone before being filtered through a Whatman filter to remove insoluble impurities (detailed 2.14.4). The solubilised pigment was loaded onto a CM sepharose column and the Pchlide was eluted from the column and purified as described in Heyes, Ruben *et al.* (2003). Following elution from the column, the purified Pchlide was concentrated to 3.2 mM, before being dried down and stored in black eppendorfs at -20 °C.

1D NMR analysis of the purified Pchlide, in fully deuterated methanol, (Figure 6.1) indicated that there was a 2:1 ratio of monovinyl to divinyl Pchlide isoforms present. The NMR analysis also indicated that there were a significant number of contaminants present in the sample, likely to be carotenoids which have co-purified with the pigment. However since

these contaminants do not absorb at 630 or 670 nm then it was decided that the current purity of pigment was sufficient for the purposes which we needed it for.

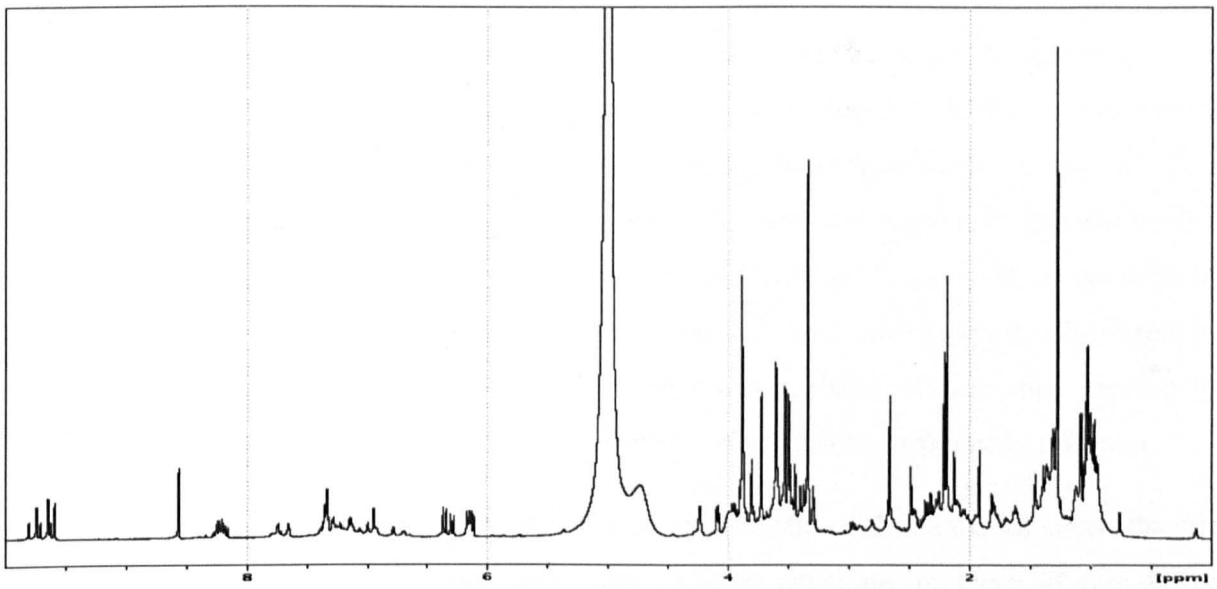


Figure 6.1: 1D proton NMR spectrum of Pchlide in CD_4 methanol acquired at 25 °C using a 600 MHz spectrometer. Peaks present between 6.0 and 6.5 along with peaks around 9.5 ppm correspond to Pchlide whilst other peaks between 7 and 8 ppm belong to contaminants, presumably carotenoids.

6.3 Optimising Conditions for the Production of Ternary Complex

As detailed in 6.2 the produced Pchlide was stored as a dry powder at -20 °C and when required for use, it was re-suspended in methanol. This was fine for standard enzyme assays, however whilst re-suspending the sample in methanol a maximum solubility was achieved of approximately 2 mM. The normal solvent of choice is methanol; however a number of other solvents were investigated, along with methanol, to establish if a greater maximum solubility could be achieved. Three solvents were investigated; these were methanol, ethanol and di-ethyl ether. However Pchlide was only sparingly soluble in the latter so the tolerance of POR for di-ethyl ether was not investigated. Although the Pchlide requires a solvent to be present to re-solubilise it, the presence of a detergent is also required in order to deliver the Pchlide to the enzyme active site (Heyes *et al.* 2002b). As standard, a concentration of 0.1 % Triton X-100 has been used in all assays (Heyes *et al.* 2000) however it was also decided to try a number of different detergents along with detergent concentrations to see what effect this had on enzyme activity.

6.3.1 Solvent Concentration

Two eppendorfs of dried down Pchlide were taken and one was re-suspended in 500 µl of methanol whilst the other was re-suspended in 500 µl of ethanol. The eppendorfs were vortexed to ensure maximum re-suspension before being centrifuged at 13,000 rpm in a bench-top centrifuge. The solubilised Pchlide was removed from the eppendorfs and a further 250 µl of solvent was added to each eppendorf. The process was repeated to ensure maximum re-solubilisation and the solubilised pigment was once again removed. The concentrations of pigment present in all four solutions were investigated by measuring the absorbance at 630 nm using a Cary 50 spectrophotometer (detailed 2.15.1). When solubilised in methanol the maximum solubility of Pchlide was 2.2 mM whilst when solubilised in ethanol the maximum solubility of Pchlide achieved was 1.0 mM. It was then decided to investigate the tolerance of POR to work at different concentrations of the two solvents.

Assays were conducted with increasing concentrations of methanol and ethanol. Increasing concentrations of solvent were added to the reactions, in place of the buffer (detailed

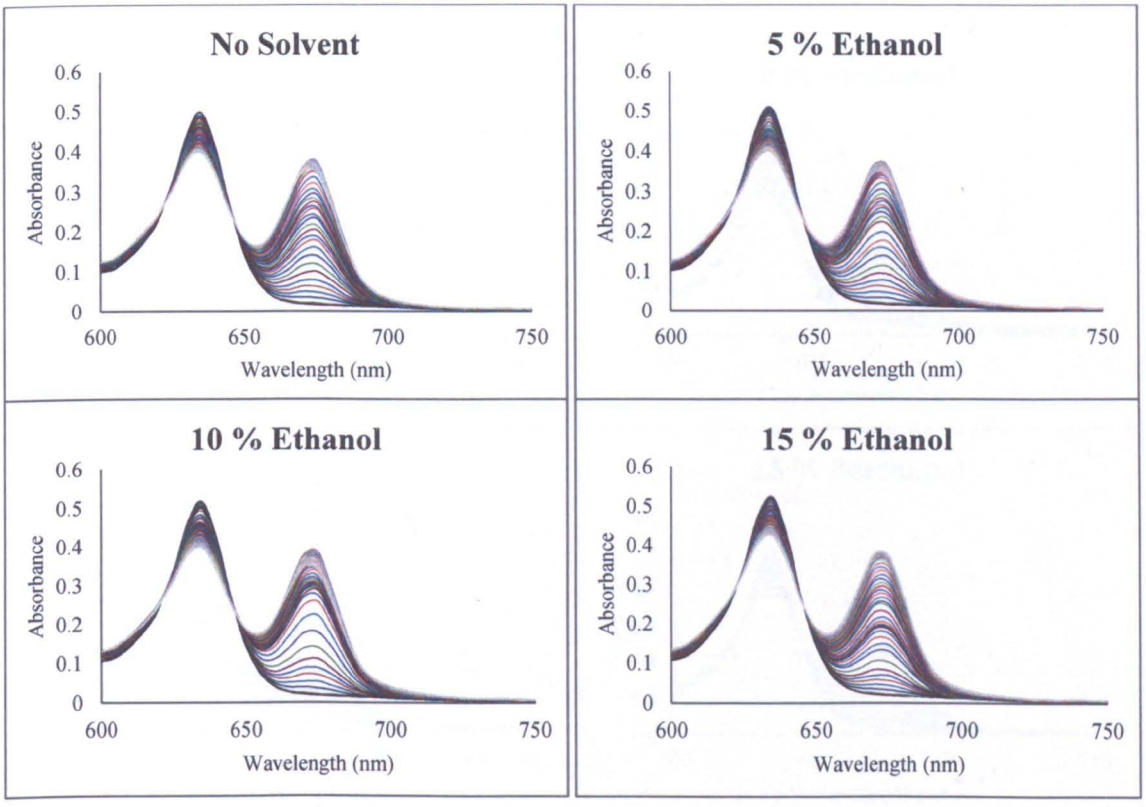
Table 6.1) thus ensuring that the total volume of the assay and the concentrations of all the constituents remained constant at all times.

Solvent	Methanol	Ethanol
Concentration (%)	0	0
	5	5
	10	10
	15	15
	20	20
	25	25
	30	30
	35	35
	40	
	45	
	50	

Table 6.1: Concentrations of the different solvents to be used in the POR assay.

The reactions were set up and assayed at 55 °C in a Cary 50 spectrophotometer with illumination being provided from a fibre optic light source with a 400-500 nm filter attached (detailed 2.15). Figure 6.2 and Figure 6.3 show the absorbance assays acquired over a period of 90 seconds, with increasing ethanol and methanol concentrations respectively.

As the concentration of ethanol increases, the rate of reaction and total Chlorophyllide (Chlide) production decreases. Since the detergent, reactants and enzyme concentration were kept constant, along with the length of illumination, the maximal amount of Chlide produced in each reaction could be used as an indication of how solvent concentration inhibited enzyme activity. The maximum Chlide absorbance at 670 nm was plotted out for the ethanol (Figure 6.4) and methanol (Figure 6.5) concentrations. These plots indicated that POR could withstand a maximum concentration of 15 % ethanol before the enzyme reaction was inhibited, whilst in the presence of methanol, a maximum concentration of 20 % could be achieved before inhibition occurs. The extinction co-efficient of Chlide is estimated to be approximately three times that of Pchlide, thus explaining why the absorbance increase at 670 nm appears to be far more significant than the decrease in 635 nm.



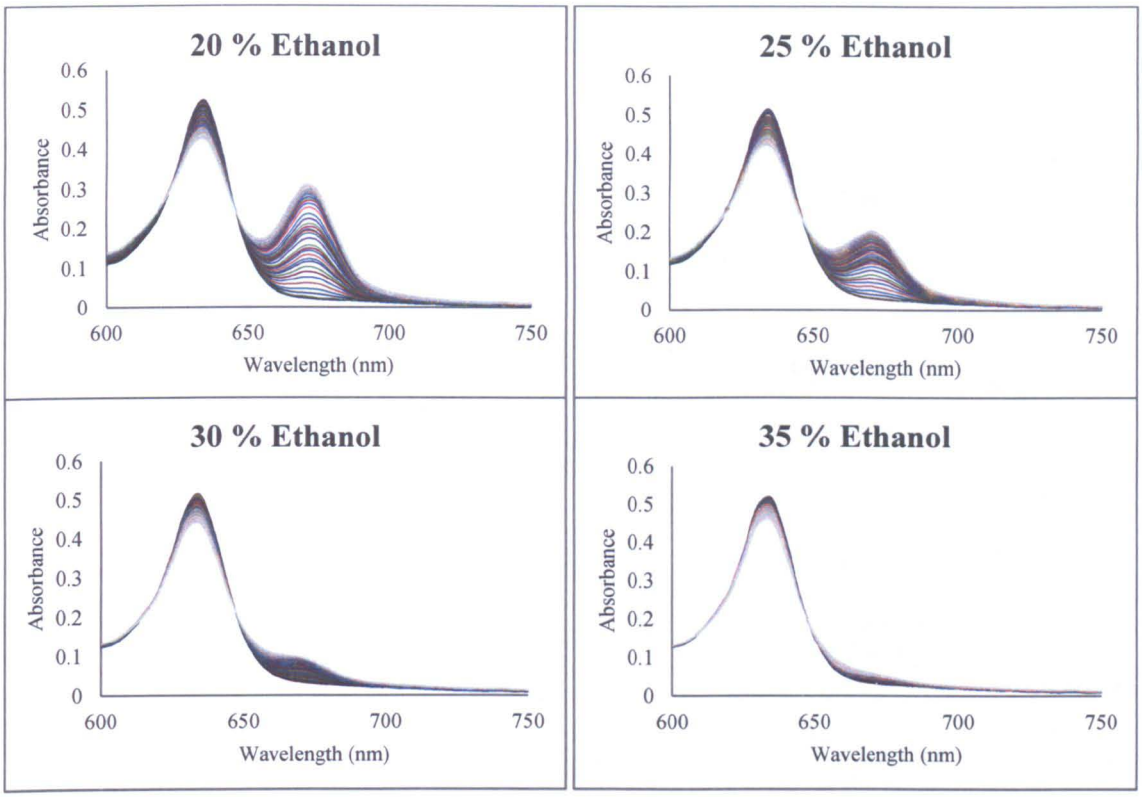
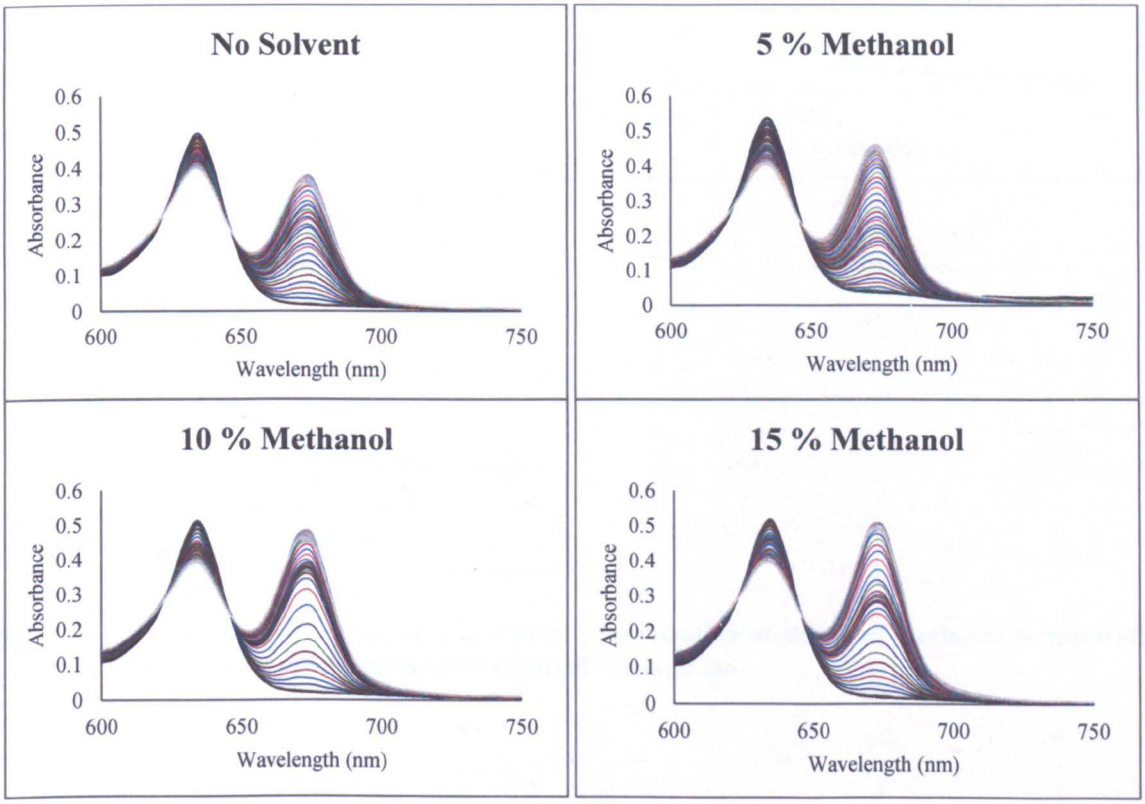


Figure 6.2: Spectra showing the increase in Chlide concentration at different ethanol concentrations over a period of 90 seconds with constant illumination.



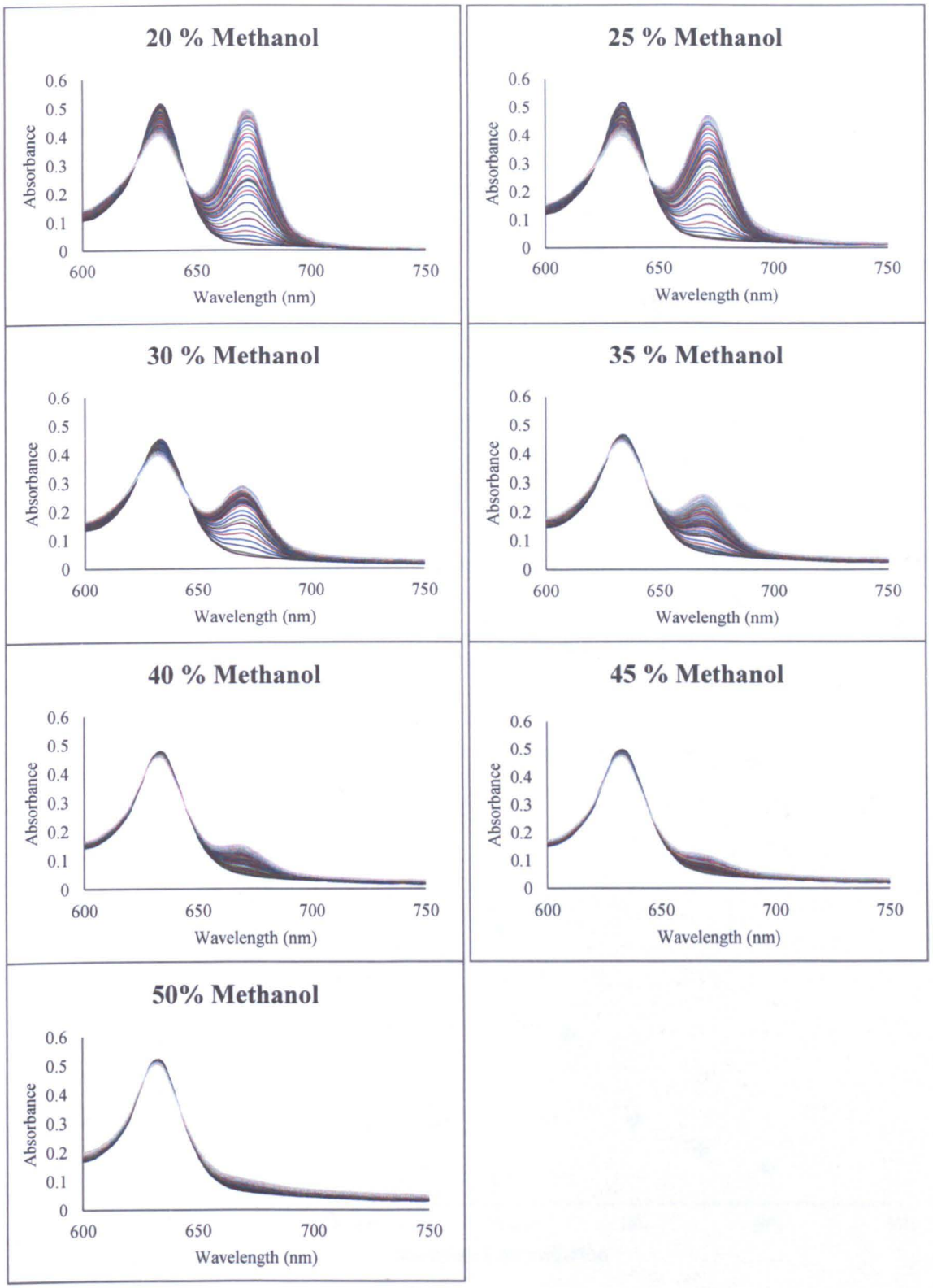


Figure 6.3: Spectra showing the increase in Chlide concentration at different methanol concentrations over a period of 90 seconds with constant illumination.

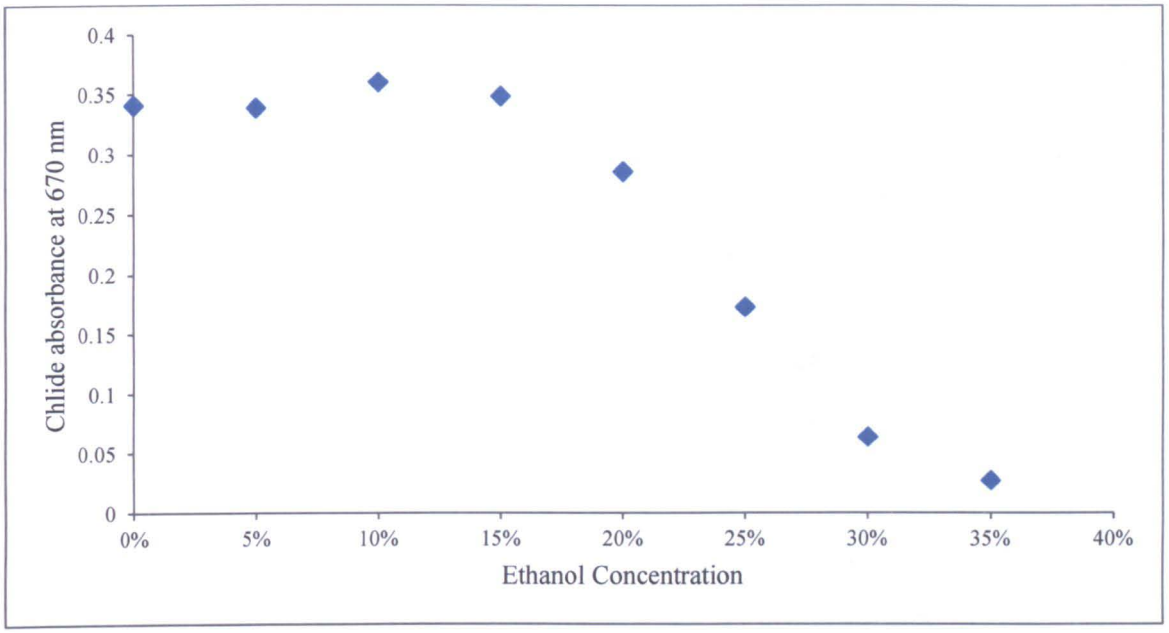


Figure 6.4: Graph showing the maximum amount of Chlide produced following 90 seconds illumination at increasing concentrations of ethanol.

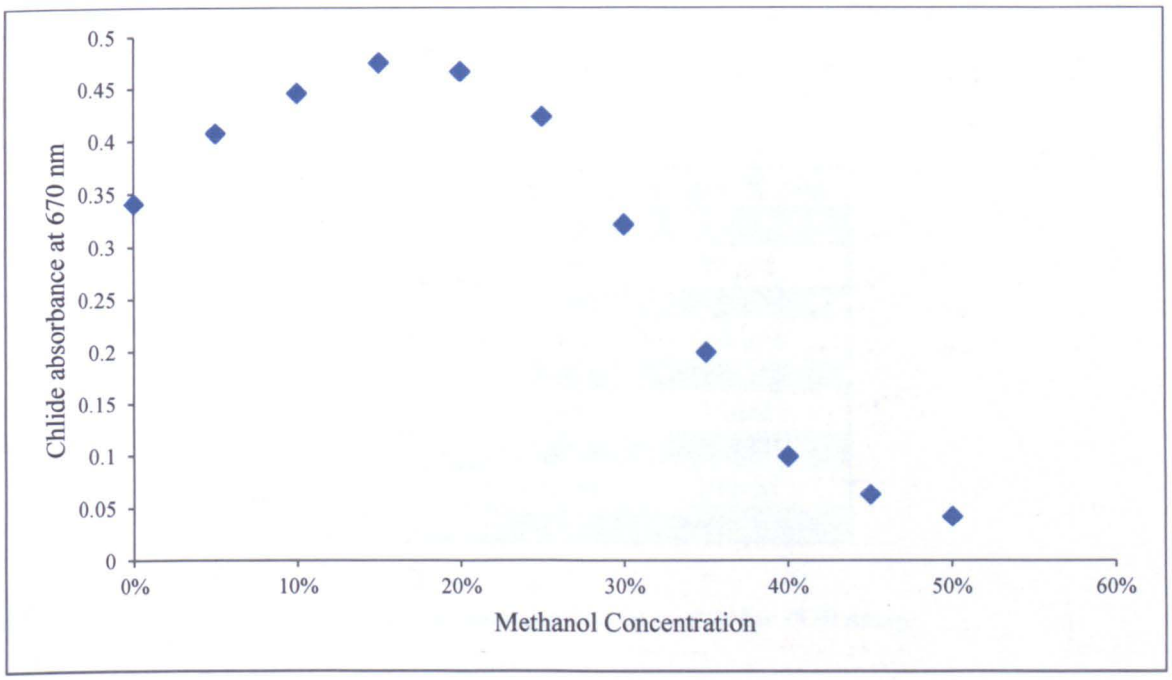


Figure 6.5: Graph showing the maximum amount of Chlide produced following 90 seconds illumination at increasing concentrations of methanol.

The combination of the reaction withstanding a higher concentration of methanol than ethanol, along with the higher solubility of Pchlide in methanol indicated that methanol would be the best solvent to deliver the Pchlide to the reaction. Also whilst forming the ternary complex, the amount of Pchlide present in each reaction will vary depending upon the concentration of POR present, however the concentration of methanol present in the reaction should not exceed 20 %.

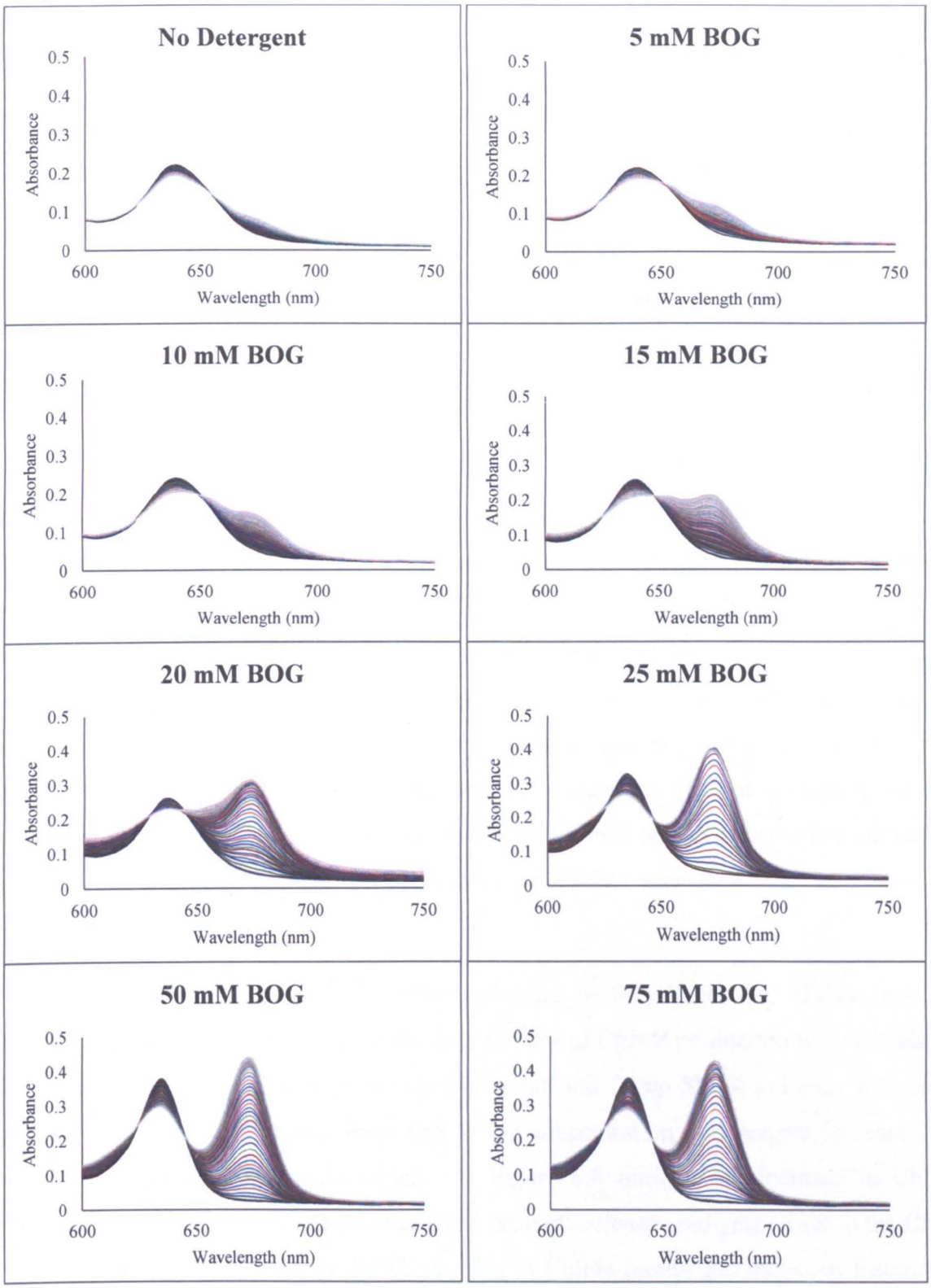
6.3.2 Detergent Concentration

In order to establish the optimum concentration of detergent needed to optimise ternary complex formation, assays were conducted as previously described in 6.3.1 with varying concentrations of detergent used instead of solvent (detailed Table 6.2). To date, all published work with this enzyme has used Triton X-100 as the detergent in the activity buffer. The detergent dependence has never been investigated and thus the decision was made to try a different detergent; n-octyl-β-D-glucoside (BOG) was the detergent chosen as it has very similar characteristics to Triton X-100, but has a larger CMC value of 23 mM, instead of 0.2 mM found with Triton X-100.

Detergent	BOG	Triton X-100
Concentration	0	0
	5 mM	50 µM
	10 mM	200 µM
	15 mM	350 µM
	20 mM	500 µM
	25 mM	1 mM
	50 mM	5 mM
	75 mM	10 mM
	100 mM	

Table 6.2: Concentrations of the different detergents to be used in the POR assay.

The reactions were set up and assayed at 55 °C in a Cary 50 spectrophotometer and assayed for a period of 90 seconds with constant illumination as detailed above (detailed 2.15). Figure 6.6 and Figure 6.7 show the absorbance assays acquired over a period of 90 seconds, with increasing BOG and Triton X-100 concentrations respectively.


No Detergent
5 mM BOG
10 mM BOG
15 mM BOG
20 mM BOG
25 mM BOG
50 mM BOG
75 mM BOG
Absorbance
Wavelength (nm)
0
0.1
0.2
0.3
0.4
0.5
600
650
700
750

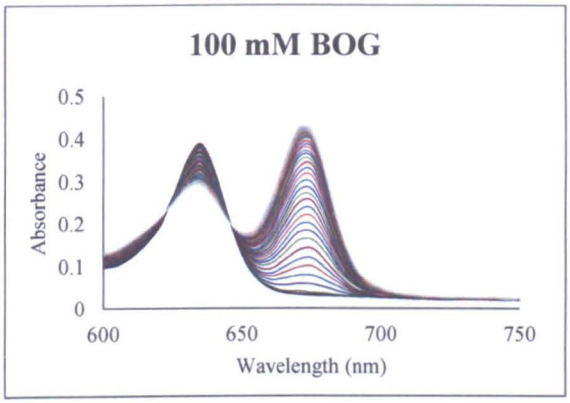


Figure 6.6: Spectra showing the increase in Chlide concentration at different BOG concentrations over a period of 90 seconds with constant illumination.

Figure 6.6 and Figure 6.7 clearly show that as the detergent concentration increases, the amount of Chlide produced also increases. This was as expected since it has been documented that a detergent is required to deliver the Pchlide to the enzyme active site. It is also apparent from Figure 6.6 and Figure 6.7 that as the concentration of detergent increases, the absorbance of Pchlide also increases. Since each assay was taken from the same stock of activity buffer, the concentration of Pchlide present should be the same in each assay, however the addition of detergent above the CMC value is sufficient enough to further solubilise more Pchlide thus increasing the absorbance at 630 nm. The same is observed in Figure 6.2 and Figure 6.3 where the absorbance of Pchlide increases with the increasing solvent concentrations.

The rate of Chlide production calculated using the first 30 seconds of data acquired from each spectrum was plotted separately and the rates of Chlide production were calculated and plotted against concentration of BOG in Figure 6.8 and Triton X-100 in Figure 6.9. Both Figure 6.8 and Figure 6.9 clearly show that as the concentration of detergent increases the rate of Chlide production also increases. In Figure 6.8 initially the increase in Chlide production is slow, but as concentration of BOG present increases and gets closer to the CMC value of the detergent, 23 mM for BOG, the rate of Chlide production increases faster. At detergent concentrations above the CMC value the increase in the rate of Chlide production appears to be approximately linear. This implies that the detergent present must be in a micellar form before it can be effectively used to deliver the Pchlide to the enzyme active site.

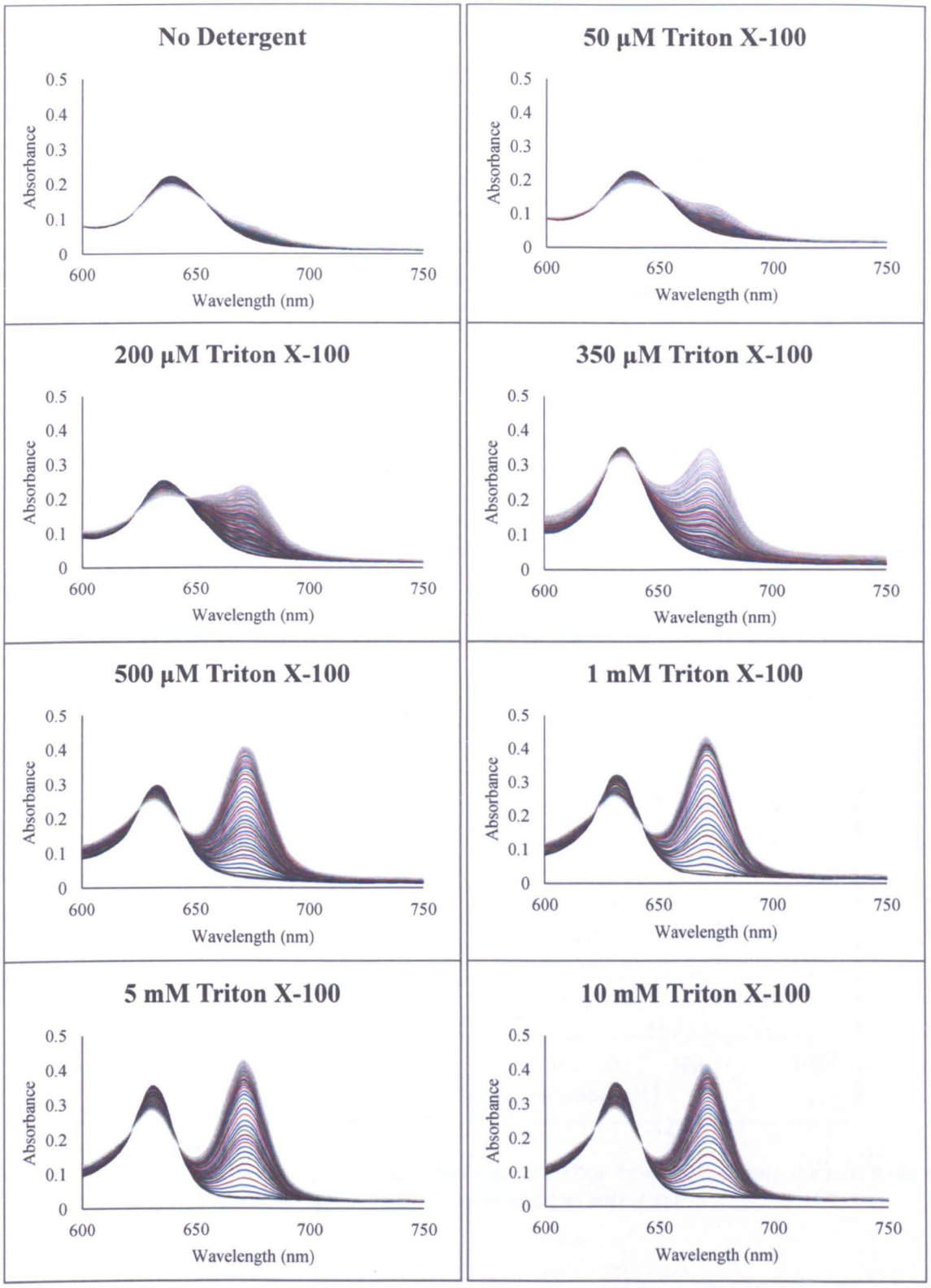


Figure 6.7: Spectra showing the increase in Chlide concentration at different Triton X-100 concentrations over a period of 90 seconds with constant illumination.

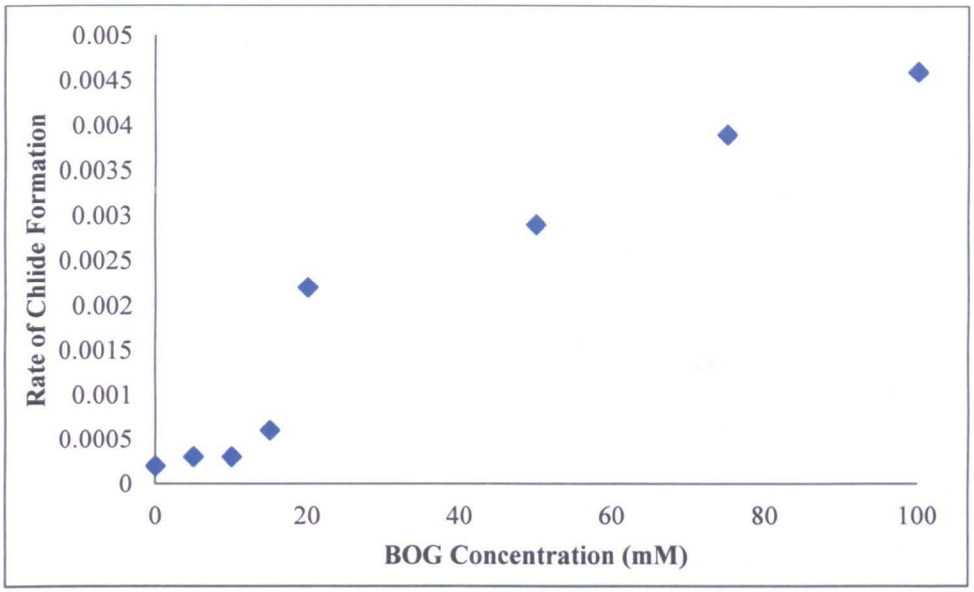


Figure 6.8: Graph showing the increase in the initial rate of Chlide formation, during the first 9 seconds of an assay following illumination, at increasing concentrations of BOG.

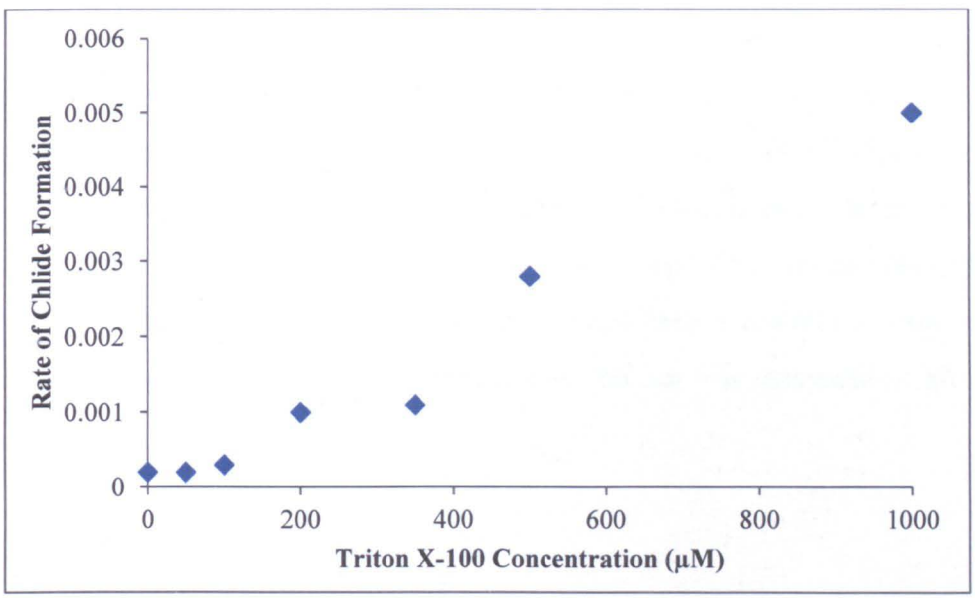


Figure 6.9: Graph showing the increase in the initial rate of Chlide formation, during the first 9 seconds of an assay following illumination, at increasing concentrations of Triton X-100.

Activity is still observed however at concentrations below the CMC value presumably because some of the detergent present is present in micelles whilst the majority of detergent molecules remain as monomers. The same is observed with the data acquired whilst using Triton X-100 which has a documented CMC value of 0.2 mM.

The data acquired in Figure 6.8 and Figure 6.9 imply that the concentration of detergent required in the sample needs to be at concentrations above the detergent CMC value in order to act as an effective deliverer of Pchlide to the enzyme active site. This therefore implies that *in vivo* some form of carrier molecule exists to transport the Pchlide produced by Mg-protoporphyrin IX monomethyl ester cyclase and deliver it to POR. It has also been suggested that POR and Mg-protoporphyrin IX monomethyl ester cyclase can form a complex, thus ensuring the effective delivery of Pchlide from one complex to the other. However to date experiments to co-purify such a complex have failed.

6.4 Production of the Ternary Complex

A 2 ml solution containing 50 mM Tris pH 7.5, 25 mM NaCl, 50 μM POR, 5 mM $NADP^+$, 100 μM Pchlide solubilised in 100 % methanol, 50 mM BOG, 1 mM DTT and Roche complete protease inhibitor was incubated in a 42 °C water bath for 15 minutes, to allow the substrates to bind to POR. The sample was diluted 5 fold, to reduce the detergent concentration well below its CMC value, before being centrifuged for 15 minutes at 13,000 rpm in a bench top centrifuge to remove all Pchlide aggregates which had formed. The sample was loaded onto an SP sepharose column, which had been pre-equilibrated in low salt buffer, at a rate of 0.5 ml/minute using a peristaltic pump. The column was connected to an Äkta Prime purification machine and the protein eluted from it at a rate of 1 ml / min using a 100ml 2 M NaCl gradient. The UV absorbance at 280 nm was measured of all the eluant (Figure 6.10)

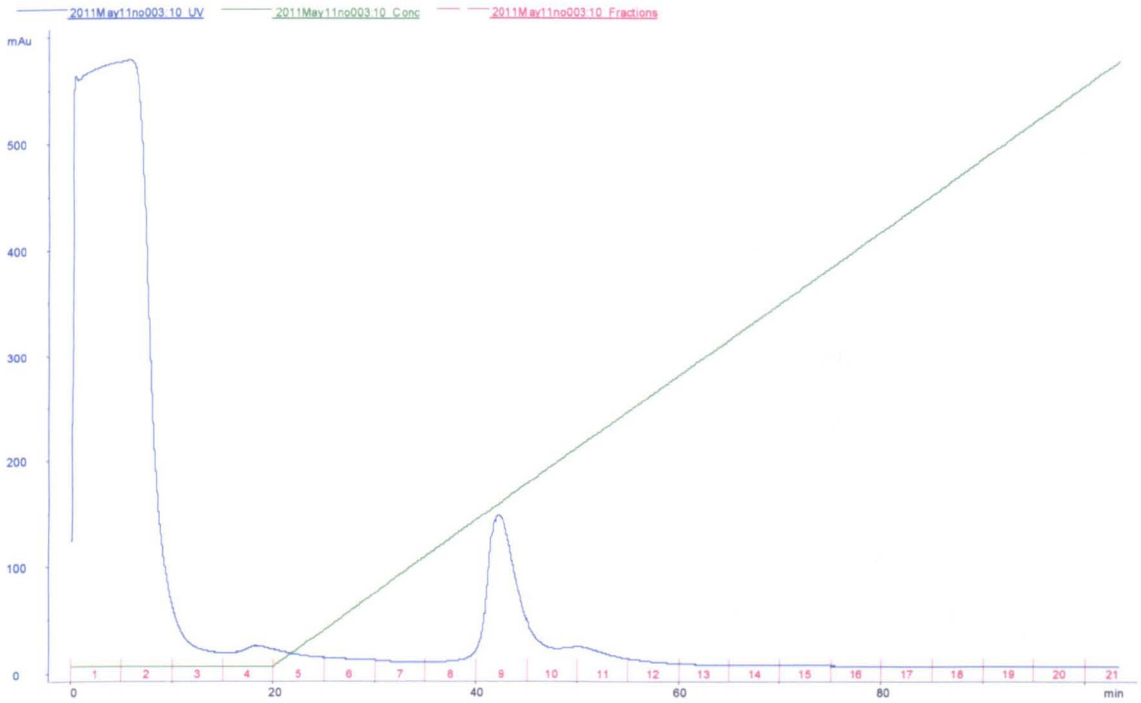


Figure 6.10: Elution profile showing the change in absorbance at 280 nm of the eluant from the SP sepharose column.

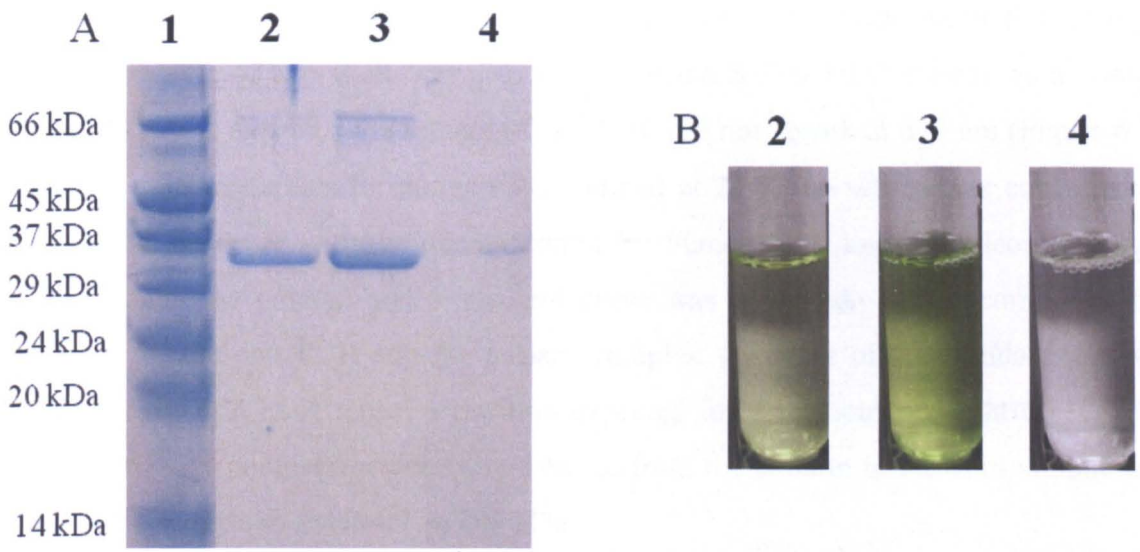


Figure 6.11: A) SDS PAGE gel of the protein present in and B) the colour of fractions 2) 8, 3) 9, 4) 11.

6.4.1 Analysis of the Ternary Complex

The fractions eluted from the column which gave an absorbance at 280 nm also had a green colour to them (Figure 6.10 and Figure 6.11 B); these fractions were therefore assumed to contain both POR and Pchlide. SDS PAGE analysis of these fractions confirmed the presence of a 37 kDa protein (Figure 6.11 A) and an absorbance spectrum taken between 260 and 700 nm confirmed the presence of both protein and Pchlide. Also upon the addition of NADPH and illumination with white light, the conversion of Pchlide to Chlide was observed in the spectrophotometer (Figure 6.15). Although all the results described above indicated both the presence of POR and Pchlide, there was no evidence to confirm that enzyme and pigment were associated with each other. To determine this HPLC analysis was conducted on the ternary complex, with absorbance spectra being taken at both 280 and 630 nm.

6.4.2 HPLC of the Ternary Complex

50 µl of produced ternary complex was loaded onto an S2000 HPLC gel filtration column and the column was run (described 2.17). An elution peak was observed at 11.5 minutes which had an absorbance both at 280 nm and 630 nm (Figure 6.12 A, Figure 6.12 B). This confirmed that the Pchlide present was associated with the POR present in solution. A sample of apo POR was also run down the S2000 HLC column as a control (Figure 6.12 C). As expected the sample of apo POR did not absorb at 630 nm (Figure 6.12 D), however quite unexpectedly the apo POR eluted at 22.5 min which was considerably different to that observed with the ternary complex. Standards of known molecular weight were run through the column and a standard curve was produced. When comparing the elution times of the apo POR and the ternary complex with that of the standards, it was observed that the apo POR ran at a position expected for a molecular weight of 37 kDa, however the ternary complex evidently was eluted from the column in the void volume and has a molecular weight of greater than 200 kDa.

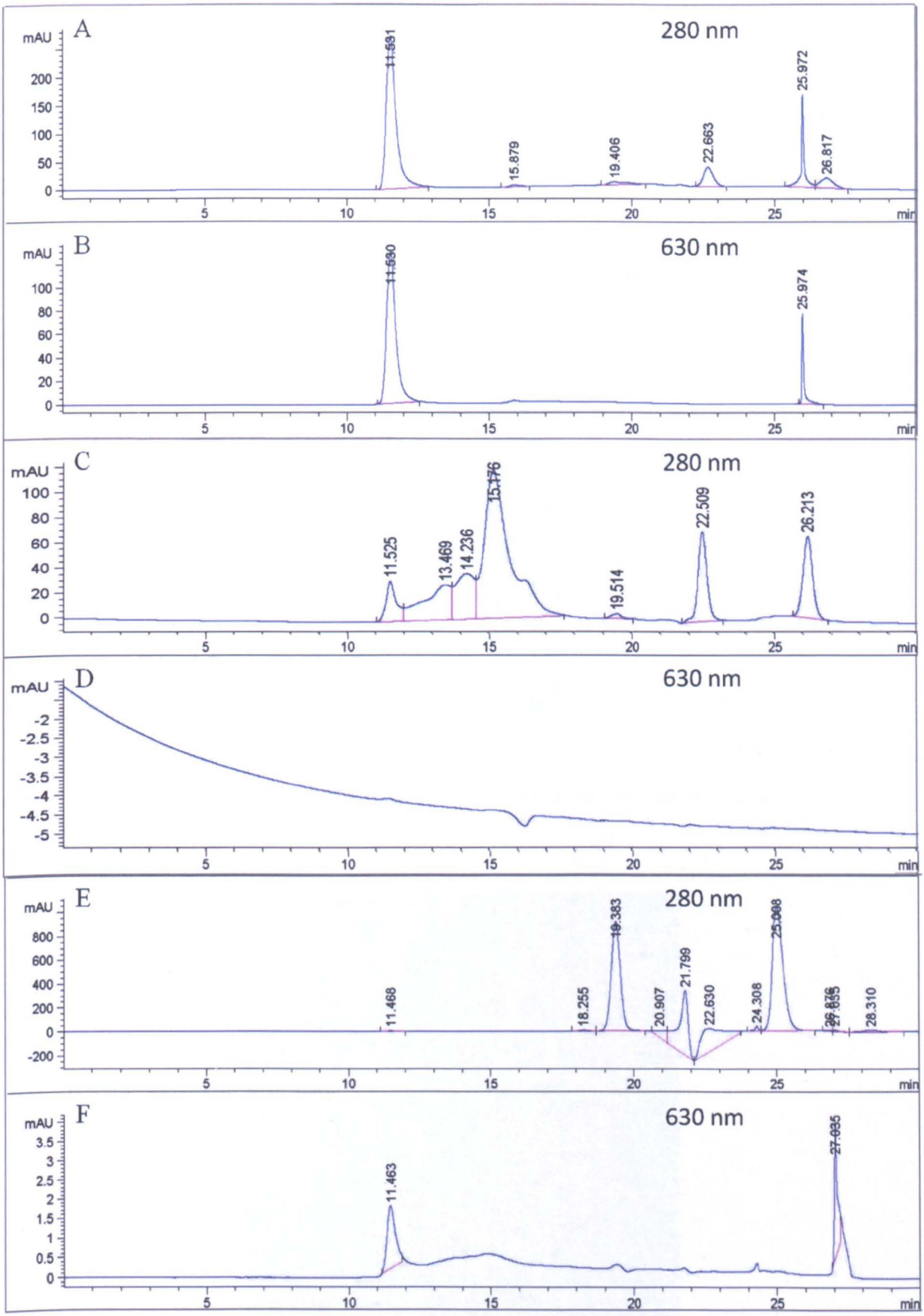


Figure 6.12: HPLC elution profiles of A & B) purified ternary complex, C & D) apo POR, E & F) illuminated ternary complex at 280 and 630 nm.

Another sample of ternary complex was analysed using HPLC and fractions of the ternary complex eluted at 11.5 minutes were collected and given to Dr Pu Qian for analysis under the Electron Microscope (Figure 6.14) as well as analysis by native gel electrophoresis (Figure 6.13). Images taken indicated that POR was present as aggregates, which could be interpreted as a pentamer of dimers. This is supported by the native gel (detailed 2.20), which indicates that the complex runs at a molecular weight roughly corresponding to a decamer of POR.

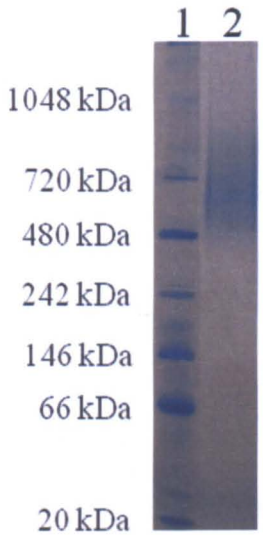


Figure 6.13: Native gel of the sample of ternary complex analysed by Electron Microscopy.

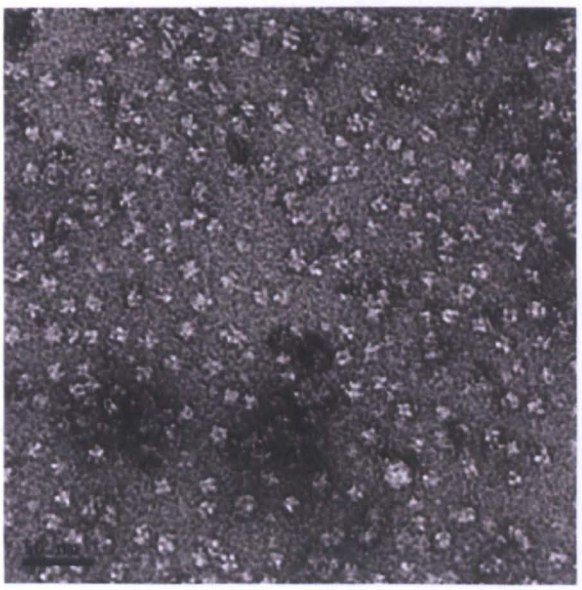


Figure 6.14: Electron Microscopy images of the POR ternary complex following elution from the S2000 HPLC gel filtration column.

It is well known that in dark grown etiolated plants, POR Pchlide complexes associate together in an ordered manner forming prolamellar bodies. In 2003 Reinbothe *et al* published work detailing the observation of the production of a light harvesting POR-Pchlide complex (LHPP) within light-grown plants. Previous results published by the same group indicated that five POR A-Pchlide b NADPH complexes and one POR B-Pchlide a NADPH complex were required to interact and form the aforementioned LHPP complex (Reinbothe *et al.* 1999; Reinbothe *et al.* 2003a; Reinbothe *et al.* 2003b). As mentioned above the work conducted by Reinbothe *et al.* has been heavily criticised, with critics claiming that the formation of such complexes is specific to monocotyledonous plants (Armstrong *et al.* 2000; Masuda *et al.* 2004). Whilst the experiments conducted by Reinbothe *et al.* indicate that the LHPP consists of six POR-Pchlide complexes, the complex produced *in vitro* with *T. elongatus* POR consists of more. Whilst the complex produced here is different to that reported by Reinbothe *et al.*, the data acquired here could be the first evidence that such LHPP complexes also exist in organisms which only contain one isoform of Pchlide and POR.

Following the illumination of the ternary complex another sample was analysed using HPLC as detailed above. The results obtained indicate that upon illumination the complex disassociates and returns to a monomeric form of POR which no longer has a Pchlide associated with it (Figure 6.12 E Figure 6.12 F). This is also consistent with the results published by Reinbothe *et al.* with the plant LHPP complex formed *in vitro* (Reinbothe *et al.* 1999; Reinbothe *et al.* 2003a).

The proposal that POR does not exist natively as a monomer is supported by the identification that all other identified members of the SDR family exist as dimers or higher order oligomers (Oppermann *et al.* 2003). In homologous enzymes, which form oligomers, the oligomerisation interface for POR is along the hydrophobic surface of α-helices D and F, which are normally protected from solvent exposure by the loop. Upon binding of the substrates I hypothesise that the loop lifts up, thus exposing the hydrophobic surface underneath and allowing the oligomerisation to occur.

Large antenna-like arrays of complexes are not uncommon in the photosynthetic pathway, with phycobilisomes and chlorosomes found in cyanobacteria and all other photosynthetic bacteria respectively, being arranged in antenna complexes to feed light energy into photosystem II (Zouni *et al.* 2001). That way if a photon hits the antenna complex anywhere, then the energy can be transferred to the reaction centre and a reaction

can be stimulated. Likewise POR-Pchlide array have only been reported to exist in dark grown etiolated plants (Benli *et al.* 1991). It is therefore a perfectly plausible idea that these arrays of complexes have evolved to exist thus allowing the formation of Chlide to be an efficient process at levels of low light intensity i.e. at the beginning of every day.

6.4.3 Assays with the Purified Ternary Complex

1 ml of purified ternary complex was placed in a 1.5 ml cuvette along with 100 μM NADPH and incubated at 55 °C for 2 minutes. The NADPH K_d for *T. elongatus* POR is reported to be approximately 21 nM (Menon *et al.* 2010), which is significantly less than the reported K_d for $NADP^+$ (Heyes et al. 2002b). By incorporating a significant excess of NADPH into the sample it was hoped that the NADPH present would replace the bound $NADP^+$ and prevent re-binding of the $NADP^+$, thus converting the complex produced into a photo-active ternary complex. The sample was placed in a Cary 50 and assayed with constant illumination (detailed 2.15) (Figure 6.15). Turnover occurred however the rate of Chlide formation and release was very slow. The same reaction was conducted as detailed above with the inclusion of 50 mM BOG as well (Figure 6.15) and the time taken for the reaction to come to completion is significantly decreased. Whilst the concentrations of Pchlide and Chlide present in each reaction should be the same, the incorporation of a detergent will further solubilise the two pigments, explaining the increase in absorbance observed in Figure 6.15 B compared to Figure 6.15 A (detailed 6.3.2).

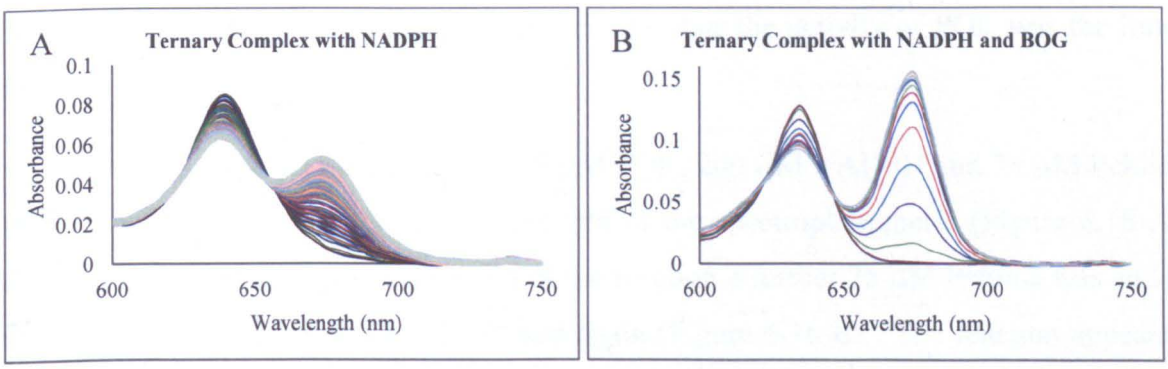


Figure 6.15: Spectra showing the increase in Chlide concentration over a period of 90 seconds with constant illumination in the A) absence of detergent B) presence of 50 mM BOG.

It is apparent from Figure 6.15 B that as the reaction progresses a 5 nm blue shift is observed in the Pchlide absorbance, whilst a 5 nm red shift is observed in the Chlide absorbance. The peak maximum observed for the Pchlide-POR-NADPH ternary complex is at 640 nm both in the presence and absence of detergent, which is within error of the published absorbance value of 642 nm (Heyes *et al.* 2004). In the absence of detergent (Figure 6.15 A) the absorbance of Chlide is at and remains at 675 nm, however in the presence of detergent (Figure 6.15 B) the absorbance of Chlide red shifts to an absorbance maximum of 680 nm, implying the presence of two different species. Since the state which absorbs at 675 nm must be present in both samples and is formed on the way to the dominating Chlide state which absorbs at 680 nm, then these two species could correspond to the Chlide-POR-NADPH and free Chlide species respectively. It can therefore be inferred from these experiments that the detergent is required not only to deliver the Pchlide to the enzyme, but also plays a role in either the removal of Chlide from the active site and or the solubilisation of free Chlide.

6.5 Limiting Factors in the POR Assay

As detailed above, and in all assays conducted, the reaction appears to come to completion before all the Pchlide is used up. In all assays conducted the concentration of NADPH present is in a large excess, compared to the concentration of Pchlide present in the reaction. Therefore the concentration of Pchlide should not be a limiting factor in the reaction. Likewise the concentration of detergent present is well above the CMC and remains constant so that should not limit the reaction either. The assumption made was that during catalysis the enzyme was becoming damaged and thus the activity of POR was the limiting factor.

Assays were set up containing 1.5 μM POR, 200 μM NADPH and 75 μM Pchlide in buffer containing 50 mM BOG and analysed in the spectrophotometer (Figure 6.16 A) as detailed in 2.15. Following completion of the reaction a further 75 μM Pchlide was added to the reaction and the assay was illuminated again (Figure 6.16 B). The reaction appeared to start again and further Chlide formation was observed. This was repeated with a further 75 μM Pchlide being added to the sample. Following illumination the formation of more Chlide was observed (Figure 6.16 C). At this stage an additional 200 μM NADPH was added to the sample however this did not result in any further Chlide formation (Figure 6.16 D).

Likewise the addition of a further 1.5 μM POR to the sample did not yield any further conversion of the Pchlide present (Figure 6.16 E). The addition of a further 75 μM Pchlide at this stage however resulted in catalysis resuming and further Pchlide being converted to Chlide (Figure 6.16 F).

The assays conducted would therefore imply that the Pchlide is the limiting factor in the assays conducted. This is very hard to understand as the absorbance maximum at 630 nm must correspond to Pchlide so Pchlide must be present in the reaction; these results therefore imply that it is present in a form which is non-catalytically active.

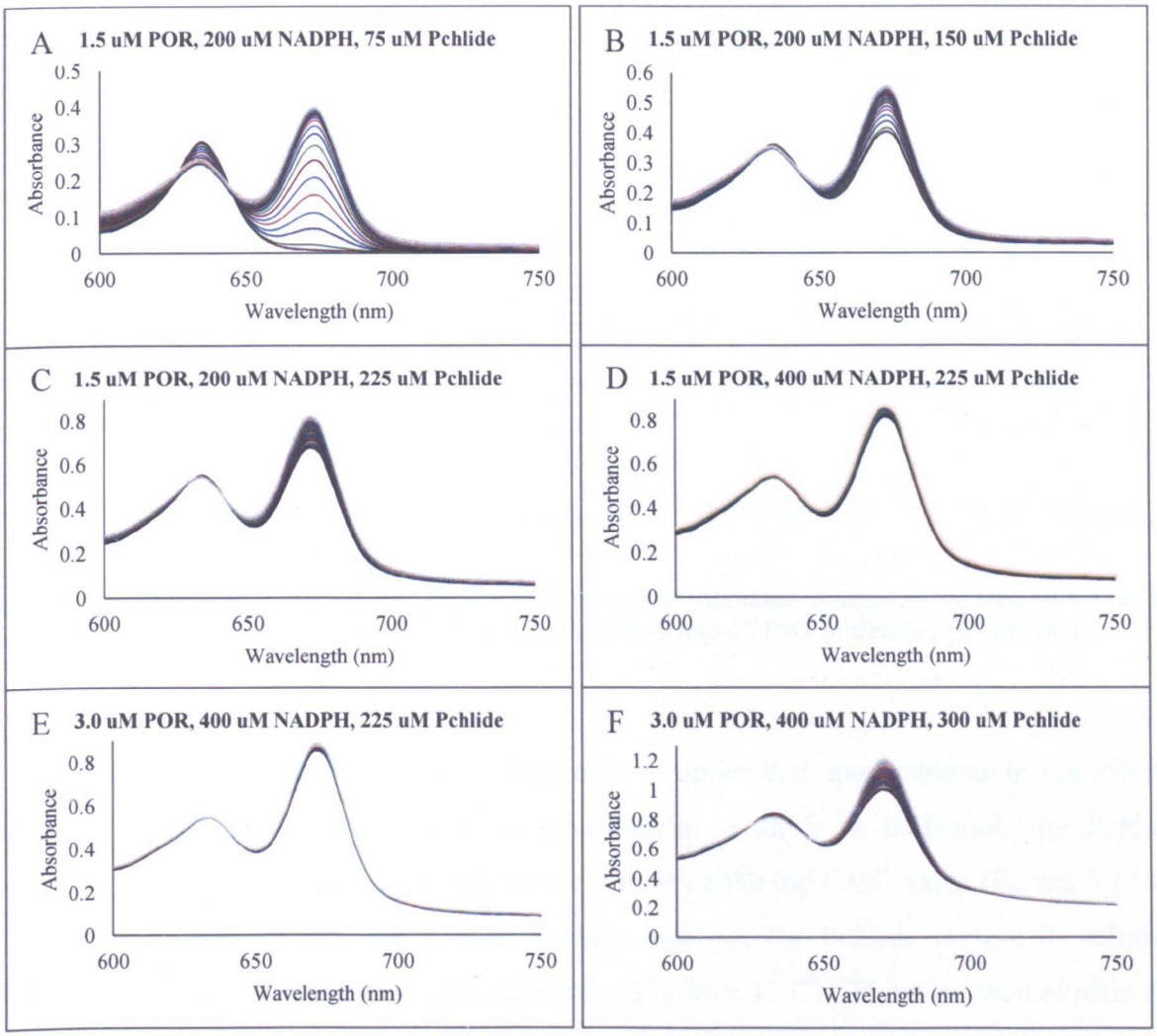


Figure 6.16: Spectra showing the increase in Chlide concentration over a period of 90 seconds with constant illumination under A) normal conditions and following the addition of B & C) 75 μM Pchlide, D) 200 μM NADPH, E) 1.5 μM POR and F) 75 μM Pchlide.

1D proton NMR analysis of Pchlide re-suspended in fully deuterated methanol (Figure 6.1) shows all peaks present to be sharp and discrete, implying that all Pchlide present is monomeric. Figure 6.17 A shows a section of the proton 1D spectrum focussing on the peaks corresponding to the meso protons. Upon addition of the solubilised Pchlide to D_2O the peaks corresponding to the meso protons, along with all other protons corresponding to Pchlide, disappeared (Figure 6.17 B). Following the incorporation of 50 mM BOG into the sample, three very broad peaks re-appeared where the discrete meso protons had previously been (Figure 6.17 C).

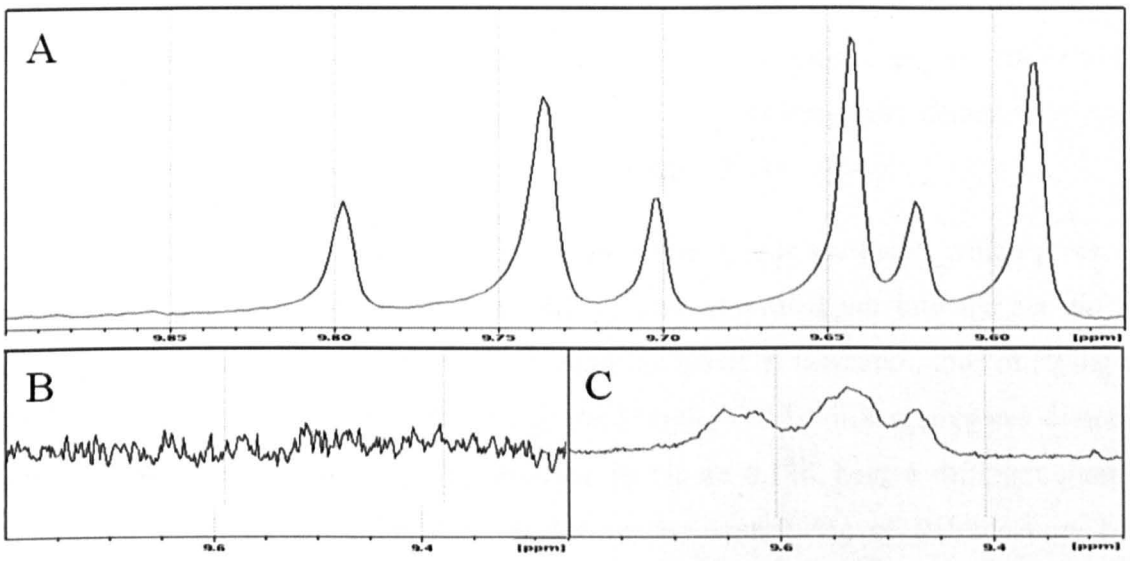


Figure 6.17: 1D proton NMR spectra showing the signals of the meso protons associated with mono a divinyl Pchlide when in A) D_4 methanol, B) D_2O and C) D_2O containing 50 mM BOG.

The lack of peaks in Figure 6.17 B therefore implies that upon removal of the solvent and placing the Pchlide in an aqueous environment, with 2 % methanol, the Pchlide aggregates. The incorporation of a detergent at levels above the CMC value (Figure 6.17 C) is sufficient to disaggregate the Pchlide slightly however the Pchlide present in solution whilst assaying POR is clearly not all monomeric (Figure 6.17 C). This may well explain the presence of the peak at 630 nm following completion of the reaction. As all the monomeric Pchlide is used up by the enzyme this leaves only Pchlide aggregates remaining in solution. If these Pchlide aggregates can not be converted into Chlide by POR then they would remain in solution and explain the remaining absorbance at 630 nm. The presence of the absorbance

at 630 nm in Figure 6.12 F does however imply that POR is capable of binding Pchlide aggregates, but just not capable of converting these aggregates into Chlide.

6.5.1 Disaggregation of Pchlide aggregates

Upon searching the literature it was apparent that previous 1D NMR studies conducted with Pchlide were done in the presence of deuterated Pyridine (Helfrich *et al.* 1999). Pyridine is one of the strongest coordinating solvents due to the nitrogen atom, which replaces the carboxylate as a ligand for the central Mg ion (Fiedor *et al.* 2008). Thus the incorporation of pyridine into the sample should be sufficient to form monomeric Pchlide in an aqueous environment. The 1D proton NMR experiments conducted above with Pchlide in D_2O containing 50 mM BOG were repeated, this time incorporating fully deuterated pyridine in 1 % increments to a final concentration of 10 % (Figure 6.18).

As the concentration of pyridine increases in the sample the peaks which previously appeared broad and non-discrete (Figure 6.18 B) have separated out into the six discrete, sharp peaks previously observed when conducting the NMR in methanol, thus implying that as the concentration of pyridine increases in the sample, the Pchlide aggregates dissociate back into monomers. The meso peaks observed in Figure 6.18K have a different chemical shift to those observed in Figure 6.18 A due to the central Mg of Pchlide now being complexed to pyridine and not water. This is sufficient to significantly alter the electron distribution resulting in the chemical shifts associated with the meso protons. Naturally having found conditions where the Pchlide present in solution is monomeric, it was of great interest to see if POR could convert all the Pchlide present in solution to Chlide in the presence of pyridine. Assays were set up containing increasing concentrations of Pyridine between 0 and 10 %. The assays were placed in a Cary 50 spectrophotometer and assayed at 55 °C (described 2.15) (Figure 6.19).

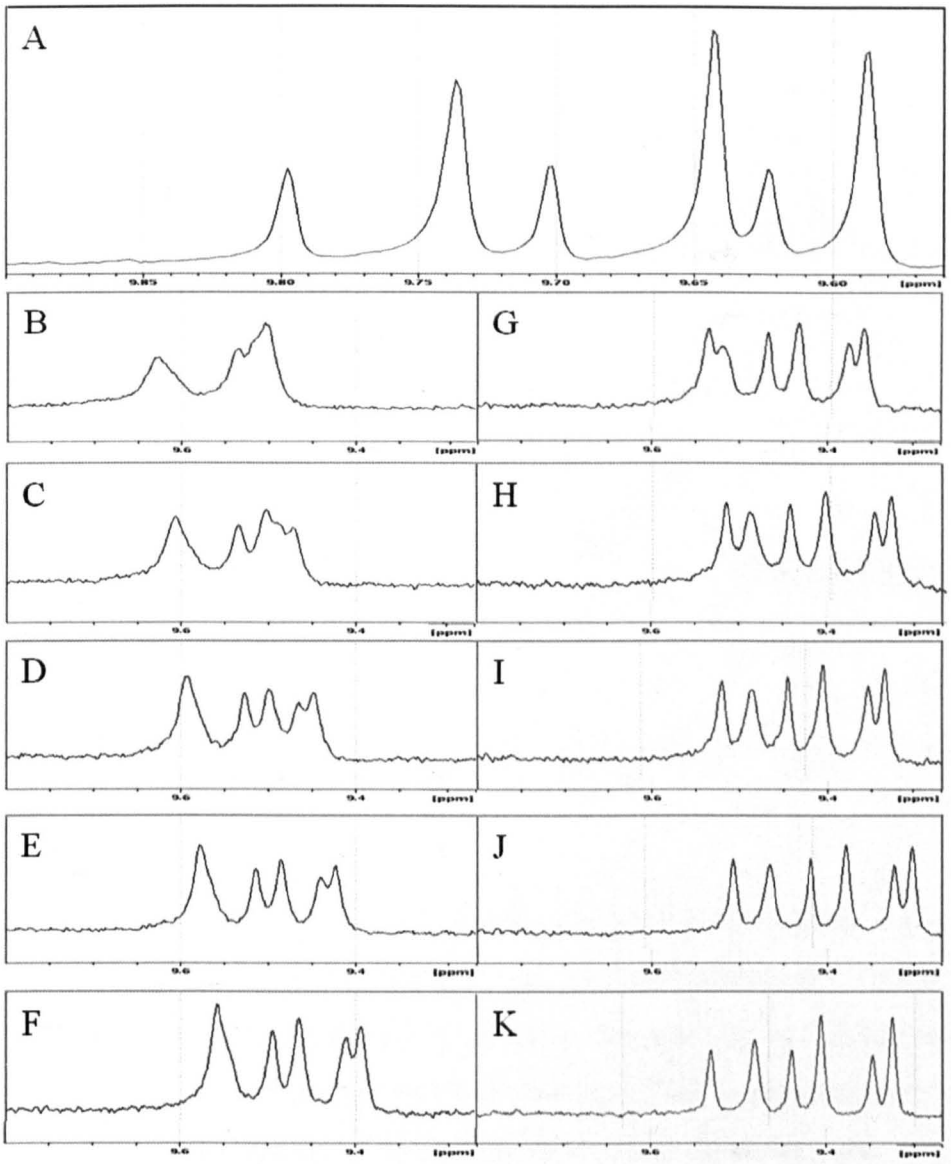


Figure 6.18: 1D proton NMR spectrum of Pchlide in A) D_4 methanol and in D_2O containing 50 mM BOG with B-K) 1% increases in pyridine to a final concentration of 10 %.

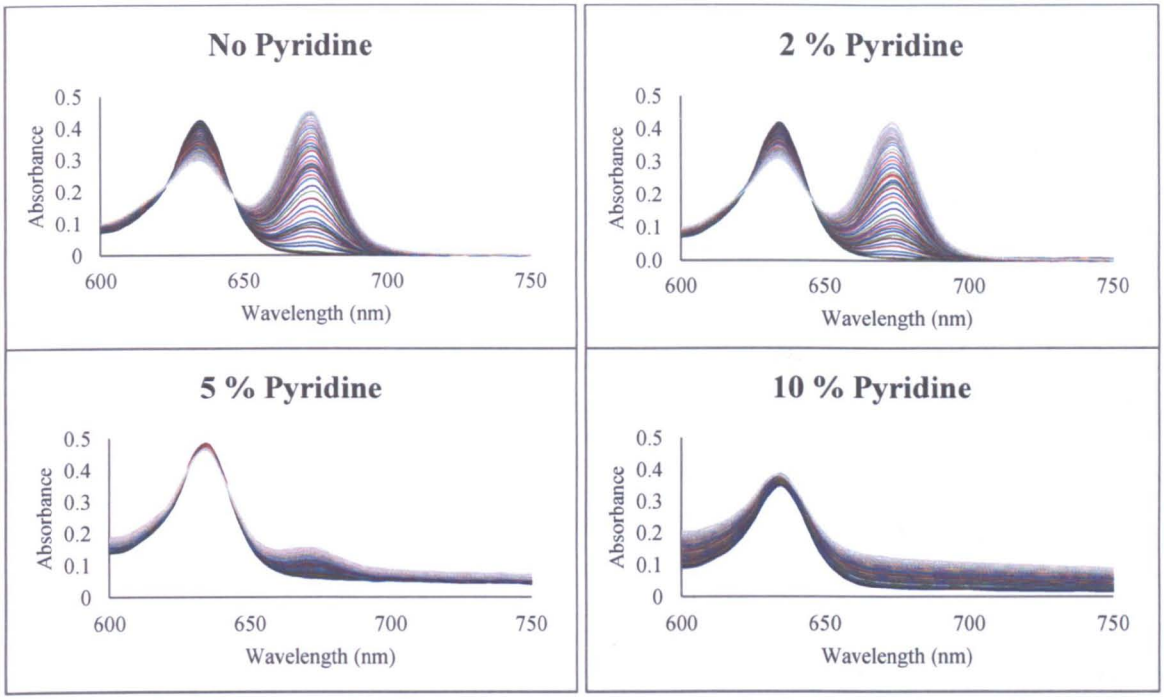


Figure 6.19: Spectra showing the increase in Chlide concentration at different pyridine concentrations over a period of 90 seconds with constant illumination.

A concentration of 2 % pyridine is sufficient to start inhibiting the activity of POR and by a concentration of 5 % pyridine the enzyme is essentially dead. This is not a vast amount of help as at concentrations of 2 % pyridine, the majority of Pchlide remaining in solution is still present as aggregates. At concentrations of 10 % pyridine there is no POR activity however there was significant movement in the baseline of the spectrum. This was due to the high concentration of pyridine etching away at the clear face of the cuvette when placed in the spectrophotometer thus distorting the light as it entered and left the solution.

The structures of imidazole and pyridine are very similar and have the same functioning group responsible for the dissociation of the Pchlide aggregates. However since imidazole is chemically similar to histidine, which is the normal physiological axial ligand for the central Mg ion of the pigment, it was wondered if high concentrations of imidazole could be used to the same effect. Since POR is eluted from the Ni His-60 column on a 1 M imidazole gradient (detailed 2.9.2), it was thought that POR would be able to withstand high concentrations of imidazole. Assays were set up containing increasing concentrations of imidazole, between 0 and 1 M. The assays were placed in a Cary 50 spectrophotometer and assayed at 55 °C (described 2.15) (Figure 6.20).

Similar to the assays conducted with increasing concentrations of pyridine (Figure 6.19), as the concentration of imidazole increases the activity of POR decreases. At a concentration of 200 mM imidazole however there does appear to be a slight increase in the concentration of Chlide produced. The increase in Chlide produced here however were nowhere near the increased levels initially expected. It is important to find a set of conditions in which the Pchlide present in solution is monomeric whilst maintaining full enzyme activity. Work to date has been published detailing the K_d for Pchlide to be 5.6 ± 0.6 μM (Menon *et al.* 2010) however if the Pchlide present in solution is aggregated and not monomeric then it is difficult to have confidence in these figures; with that in mind however the actual K_d for Pchlide is not going to be any larger than that currently published.

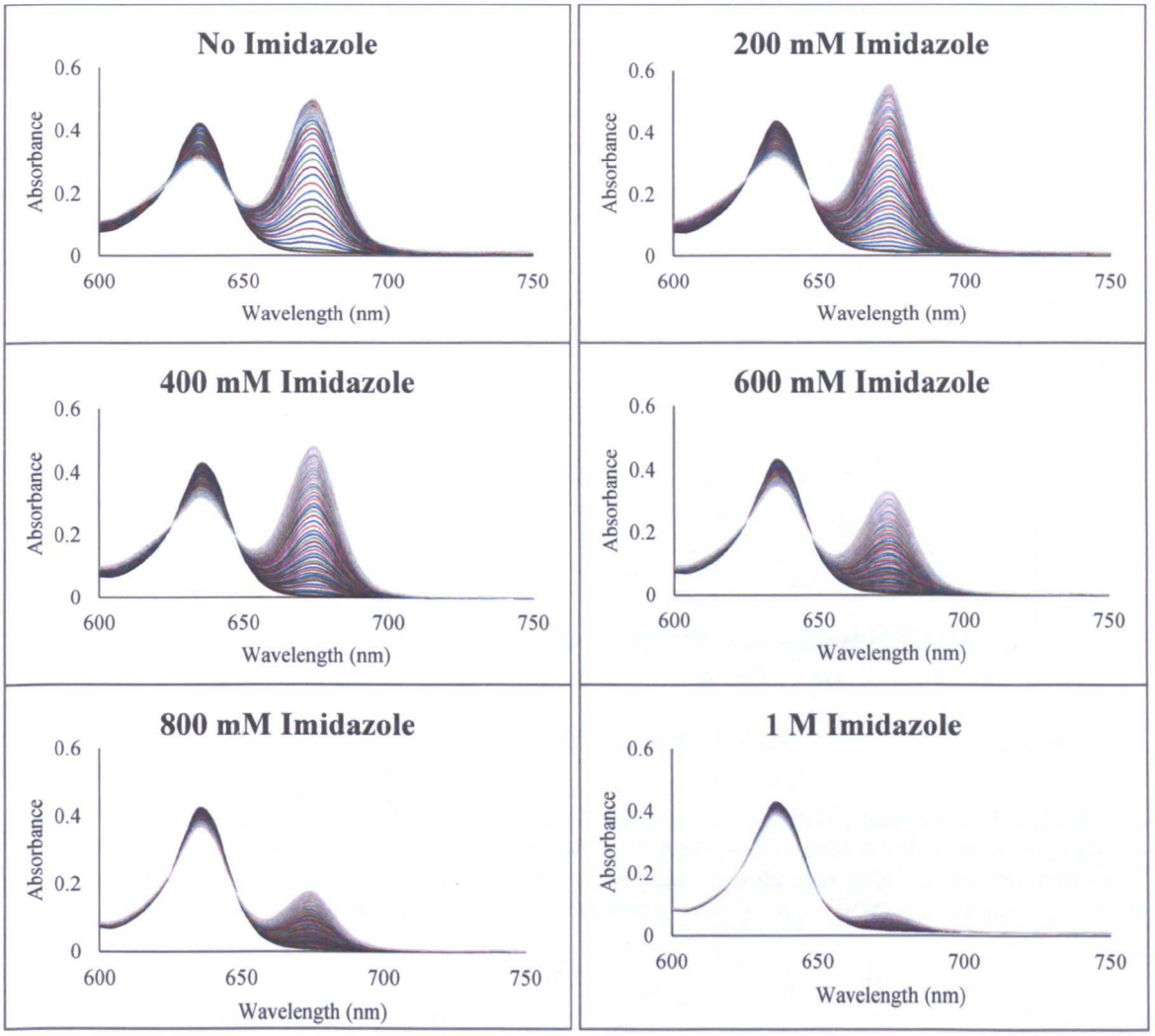


Figure 6.20: Spectra showing the increase in Chlide concentration at different imidazole concentrations over a period of 90 seconds with constant illumination.

6.5.2 Analytical Ultracentrifugation of the Ternary Complex

Five samples were produced (detailed Table 6.3) and sent to Professor Steve Harding at the University of Nottingham for analysis by analytical ultracentrifugation (AUC). Analysis of the samples was conducted at 300 nm for protein, 630 nm for Pchlide and 670 nm for Chlide. The results for all samples were combined and plotted in Figure 6.21. Whilst there is no exact relationship between sedimentation coefficient and aggregation state, the sedimentation coefficients of peaks 1 to 5 are consistent with the analysis that they correspond to free pigment in solution, be it Pchlide or Chlide, POR monomer, dimer, pentamer and decamer respectively.

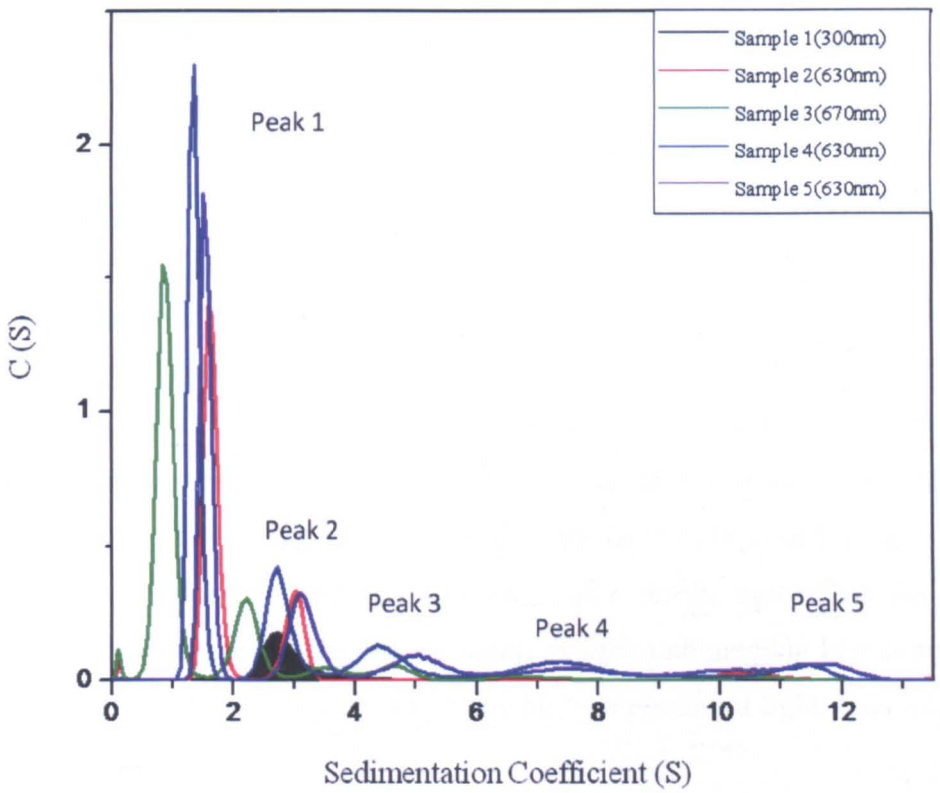


Figure 6.21: Analytical Ultra Centrifugation data of Sample 1) Apo POR, Sample 2) Purified ternary complex, Sample 3) Ternary complex with 50 mM BOG, 1 mM NADPH and saturated with light for 10 minutes, Sample 4) Ternary complex with 50 mM BOG, 1 mM NADPH and no light exposure and Sample 5) Ternary complex with 50 mM BOG and saturated with light for 10 minutes.

Sample	Contents
1	Apo POR
2	Purified Ternary Complex
3	Ternary Complex with 50 mM BOG, 1 mM NADPH and saturated with light for 10 minutes
4	Ternary Complex with 50 mM BOG, 1 mM NADPH with no light exposure
5	Ternary Complex with 50 mM BOG and saturated with light for 10 minutes

Table 6.3: Details of the samples sent for analysis by Analytical Ultra Centrifugation.

From the AUC data it is fair to say that apo POR has a sedimentation coefficient of approximately 3.3 S; this is consistent with what would be expected for a monomeric protein of molecular weight 37 kDa. Following formation of the ternary complex, a number of aggregates are formed which are predicted to include monomer, dimer, pentamer and decamer. Upon addition of detergent and NADPH to the sample and following illumination, all the higher order aggregates disappear leaving monomeric and dimeric POR which have Chlide bound in the enzyme active site (Table 6.4). By contrast, shining light on a ternary complex sample which has not had NADPH added and adding NADPH to a sample which has been kept in the dark, results in a distribution of species similar to that observed with the purified ternary complex from the column (Table 6.5). It can therefore be said that upon the addition of NADPH and detergent, exposure of the sample to light is sufficient to reduce the aggregates back to their monomeric constituents. These results agree with the hypothesis proposed above which suggested the ternary complex forms an antenna like complex which forms to increase the chance of the complex being hit by a photon of light, thus increasing the chance of Chlide formation.

	Percentage Proportion of Species Present in Peaks				
Sample	**2**	**3**	**4**	**5**	**6**
2	34	16	15	32	3
3	65	35	0	0	0

Table 6.4: Table detailing the proportion of species present in 2) a sample of purified wild type POR-Pchlide-NADP$^+$ ternary complex and 3) an illuminated sample of purified ternary complex supplemented with NADPH.

	Percentage Proportion of Species Present in Peaks				
Sample	**2**	**3**	**4**	**5**	**6**
2	34	16	15	32	3
5	32	19	15	22	12

Table 6.5: Table detailing the proportion of species present in 2) a sample of purified wild type POR-Pchlide-NADP$^+$ ternary complex and 3) an illuminated sample of POR-Pchlide-NADP+ ternary complex.

The high concentration of unbound Pchlide present in the sample of purified ternary complex implies that a significant proportion of Pchlide in the sample is non-specifically associated with the POR sample. This could result from the way the ternary complex was produced and purified. POR has a number of hydrophobic patches on it which may well bind free or aggregated Pchlide following dilution of the detergent. It may be useful therefore to investigate whether the ternary complex sample can be washed with low concentrations of detergent to remove this non-specifically bound pigment prior to elution from the SP sepharose column.

6.5.3 NMR of the Ternary Complex

A 0.2 mM sample of POR-Pchlide-NADP$^+$ ternary complex (detailed 6.4) in Tris buffer pH 7.5, 100 mM NaCl, 1 mM DTT and Roche complete protease inhibitor was placed in an standard NMR tube and analysed in a 600 MHz spectrometer which had been pre-equilibrated at 50 °C (detailed 2.21). Sharp discrete protein peaks, which had previously

been present in the amide region between 6 and 11 ppm, were broad and in many cases not present. Likewise proton peaks corresponding to folded protein, found between 0 and -1.0 ppm, were no longer present. The implication of this result was that NMR analysis of the ternary complex was not possible. This was to be expected, as the correlation time of a complex consisting of 5 to 10 molecules of POR will be significantly increased, even at 50 °C.

6.6 Mutagenesis of the 33 Residue Insertion Loop

As documented above the initial aim of creating and purifying a ternary complex with no detergent or solvent present was to assist with the NMR studies of POR in the substrate bound state. It was proposed that the mobility of the 33-residue insertion loop could play a role in broadening the signals in that region and it was therefore hoped that binding the substrates to POR would significantly reduce the loop mobility. However the discovery that POR associates into high molecular weight aggregates of approximately 370 kDa upon binding of the substrates to the enzyme makes NMR analysis of the ternary complex impossible (detailed 6.5.3).

The POR sequences from a number of POR isoforms were placed into the Clustal: Multiple Sequence Alignment tool (ClustalX) and the sequences were aligned. Figure 6.22 shows the alignment for three POR sequences found in *Gloeobacter violaceus, Synechococcus elongatus* and *Thermosynechococcus elongatus.*

```
G. violaceus ---MAEQTVI ITGASSGVGL HAADSLAQSG RWHVVMACRD KAKAQNAAAS
S. elongatus MSETQQPTVI ITGASSGVGL YGAKALAARG -WHVVMACRN LQKAAEAAKS
T. elongatus MSDQPRPTVI ITGASSGVGL YATKALANRG -WHVIMACRN LEKAEQAAKN

             LGMRPQSYTI VHLDLASLAS VRRFVQDFRA LGQPLDALVC NAAVYTPTAR
             LGINPENYSL MEIDLGSLAS VRRFVDQFRA TGRSLDALVC NAAVYLPRLK
             LQIPPEAYTI LHLDLSSLAS VRGFVESFRA LNRPLRALVC NAAVYYPLLK

             EPRYTADGFE LSVGTNHLGH FLLCNLLLED LQNSPAAEPR LVILGTVTHN
             EPQRSPEGYE ISVATNHFGH FLLCNLLLDD LKRSPAPEKR LVILGTVTAN
             EPIYSVDGYE ITVATNHLGH FLLINLLLED LKNSPESDKR LVILGTVTAN

             PKELGGSIPP --RPDLGDLR GLEQGFKAPH TMIDGKAYNP VKAYKDSKVC
             SKELGGKIPI PAPADLGNLE GLEAGFKAPI AMIDGKKFKP GKAYKDSKLC
             RKELGGKIPI PAPPDLGNLE GFEKGFKKPI AMINGKPFKS GKAYKDSKLC

             NLLTMRELHR RFHTSHKITF SALYPGCVAT SGLFRESPRL FQILFPVFQK
             NMITTRELHR RFHDSTGIVF GSLYPGCVAD TPLFRNTPKL FQKIFPWFQK
             NMLTARELHR RFHESTGIVF NSLYPGCVAD TPLFRHHFPL FQKLFPLFQK

             YVTGGFVSEA EAGGRVAALV DDPAYSRSGV YWSWGNRQKK DGKSFIQDVS
             NITGGYFTQE LAGERVAQVV ADPEFKTSGV HWSWGNRQQK DRQSFVQELS
             KITGGYVSQE LAGERVAMVV ADPEFRQSGV HWSWGNRQKE GRKAFVQELS

             TEASDEDKAR RLWDLSAGLV GLA
             DKASDDRTAQ RLWDLSAKLV GL-
             AEASDEQKAR RLWELSEKLV GLA
```

Figure 6.22: Sequence homology stacks of POR from the organisms *Gloeobacter violaceus, Synechococcus elongatus* and *Thermosynechococcus elongatus*. The loop is coloured in blue with the conserved glycine residues being coloured red.

Significant sequence homology was observed throughout the three sequences, but specifically in the loop region. Notably six glycine residues were observed to be conserved in the loop region with Gly-154 and Gly-155 being present at one end of the proposed loop position whilst Gly-190, which is situated just out side of the predicted loop, is conserved across two of the species. Gly-144 is also conserved across all sequences investigated however the position of this residue is some 10 places away from where the loop is predicted to be placed (Figure 6.23). It was thought that the conserved glycine residues present at each end of the loop could act as hinges resulting in the loop's mobility. By mutating the conserved glycine residues to alanine residues it was hoped that this would reduce the mobility of the loop and thus improve the quality of the NMR spectrum acquired.

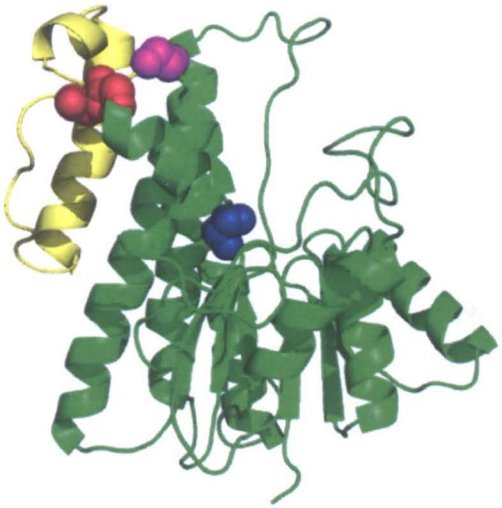


Figure 6.23: Model of POR detailing the position of the 33 residue insertion loop (yellow) along with the position of Gly 144 (blue), Gly 154,155 (red) and Gly 190 (pink).

6.6.1 NMR of Glycine Mutants

Mutagenesis of Gly-154 and Gly-155 to alanine was done in one PCR reaction with Gly-190 being mutated to alanine in another. Both of these reactions were conducted by third year biochemistry undergraduates Alexandra Mellors and Keiran Buckler. The plasmid DNA created was sequenced to confirm the presence of the mutations and the G154/5A, G190A and G154/5/90A forms of POR were over-expressed in ^{15}N labelled M9 media (detailed2.6) and purified (detailed in 2.9), before being concentrated (detailed 2.11) and buffer exchanged into NMR buffer (detailed 2.10). All versions of POR could be purified in the conventional way however significant reductions in yield were observed with all POR forms containing the G190A mutation. Yields of protein were estimated using the Bradford Assay and are summarised in Table 6.6. All samples were concentrated down to a volume of 500 µl and placed in a shigemi tube for analysis by NMR. The sample was placed in a 600 MHz spectrometer which had been pre-equilibrated to 45 °C before 1D and HSQC spectra were acquired (Figure 6.24).

Type of Protein Produced	Yield of Protein Achieved / Litre of Medium (mg)
Wild Type	30
Gly 154/5 Ala	25
Gly 190 Ala	1.15
Gly 154/5/90 Ala	0.9

Table 6.6: Table detailing the amount of wild type and mutant protein produced from 1 litre of LB growth medium.

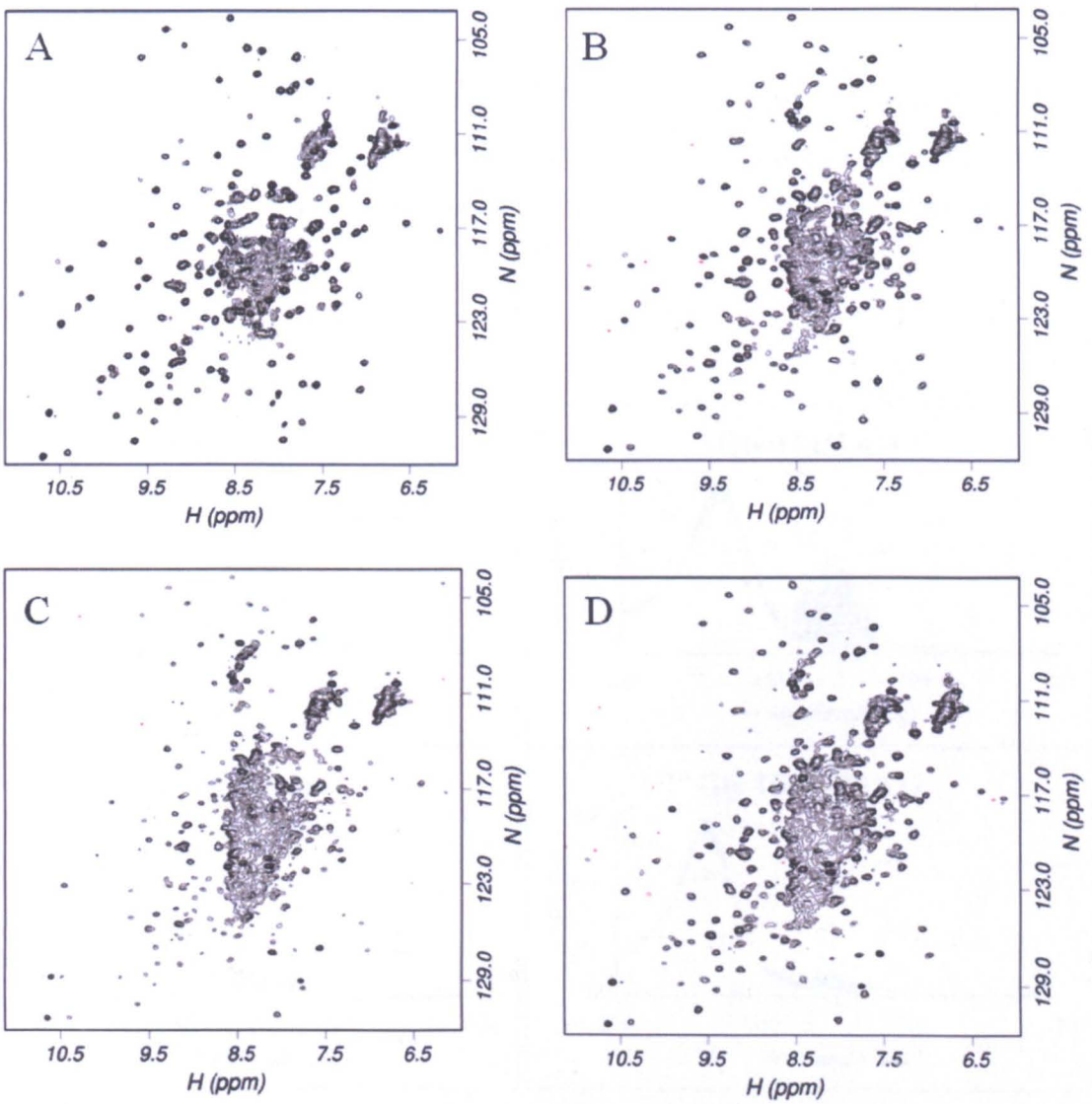


Figure 6.24: HSQC spectra acquired using A) apo POR, B) G154/5A POR, C) G190A POR and D) G154/5/90A POR.

NMR analysis of the glycine mutants showed no significant improvement in spectral quality, with the G154/5A mutant looking essentially identical to wild type POR. There was no detectable difference in the glycine region of the HSQC spectrum (towards the top of the spectrum). Figure 6.24 C and Figure 6.24 D which correspond to the G190A and G154/5/90A forms of POR respectively do show slightly improved spectral quality in the central heavily overlapped region, however the improvements observed are nowhere near significant enough to justify the significant extra cost required to produce a 500 µl sample which is 0.3 mM in concentration and double labelled.

6.6.2 Activity Studies

Following over-expression and purification, the glycine mutant POR proteins were assayed to see if the mutations had affected the enzyme activity. Reactions were set up and placed in a Cary 50 spectrophotometer and assayed at 55 °C under constant illumination (described 2.15) (Figure 6.25).

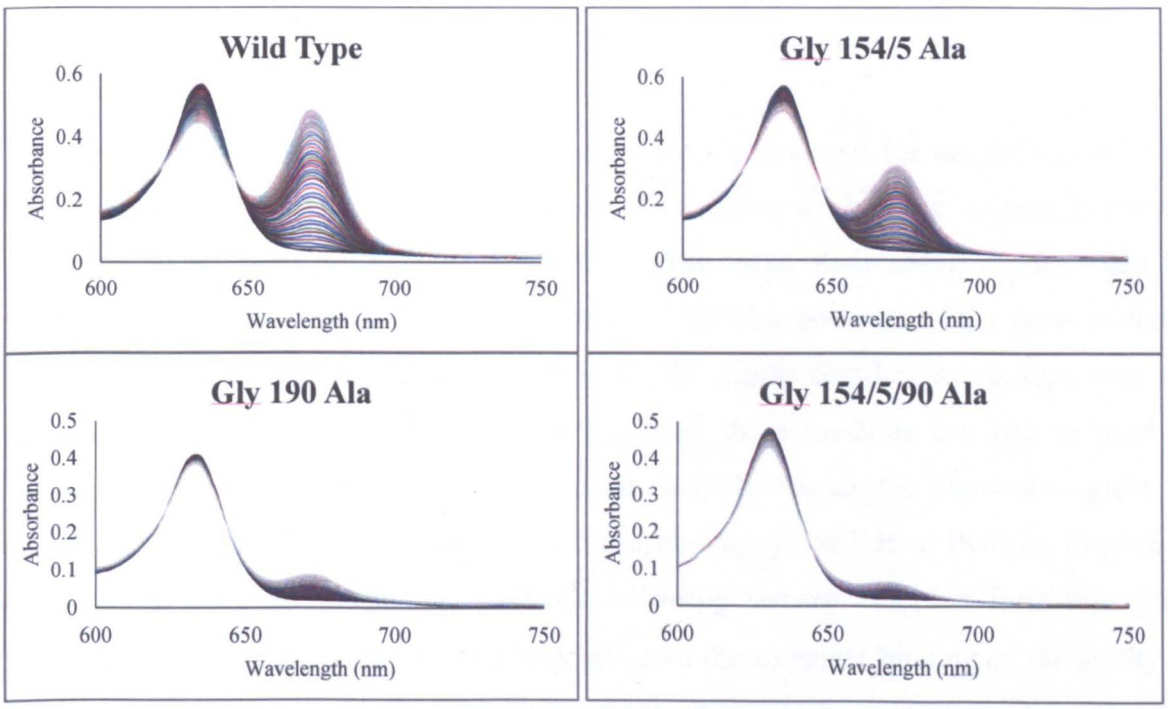


Figure 6.25: Spectra showing the increase in Chlide concentration, using the mutant forms of POR, over a period of 90 seconds with constant illumination.

The concentration of enzyme and reactants was kept constant across all experiments so the amount of Chlide produced in each reaction can be used as a direct measure of how the mutations have affected the enzyme activity. All mutations have resulted in a reduction of enzyme activity however all forms of POR containing the G190A mutation appear to have been affected the most. A summary of remaining activity is detailed in Table 6.7.

Mutation	μM Chlide Increase/Minute/mg POR	Relative Activity Remaining (%)
Wild Type	145.5	100
G154/5A	96.5	66
G190A	28.51	20
G154/5/90A	3.56	2.5

Table 6.7: Table detailing the levels of enzyme activity present in wild type POR along with that remaining with the glycine mutant protein.

6.6.3 Substrate Binding Studies

It is apparent that the 33-residue insertion loop is essential for the activity of POR although it remains unclear whether it has a role in substrate binding or protein-protein association. The ability to measure substrate binding using fluorescence is described by Heyes *et al* in (Heyes *et al.* 2000) and (Heyes *et al.* 2002b), however whilst these methods will give accurate information on the NADPH K_d, the results detailed above, regarding the aggregation of free Pchlide in solution, explain why these methods can not be used to generate accurate data on the binding of Pchlide to POR. Since the pigment is green in colour however it should be possible to assess the binding of Pchlide to POR by inspecting the colour of the POR containing fractions following ternary complex formation. By establishing if the mutations to the loop have affected the substrate binding or the ability of POR to form its higher molecular weight structures, it should be possible to determine the function of the loop in the activity of POR.

6.6.3.1 NADPH Binding Studies

The amount of NADPH bound to POR was analysed using the changes in NADPH fluorescence at 454 nm (detailed 2.18.2). Heyes *et al* previously reported that upon binding of NADPH to POR an enhancement of the signal at 454 nm is observed when compared to the signal observed from free NADPH alone (Heyes *et al.* 2000). Figure 6.26 A shows the fluorescence spectrum acquired from Milli Q water along with that observed from fluorescence buffer (detailed A.3.12), with and without NADPH.

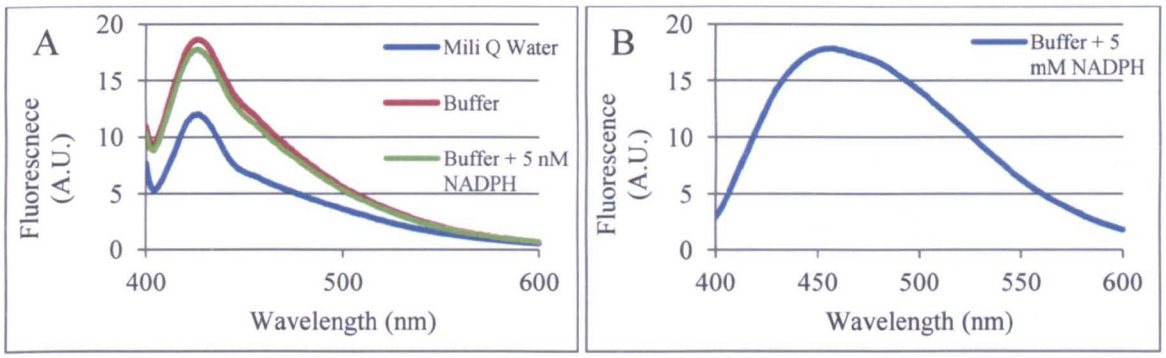


Figure 6.26: Emission fluorescence spectra acquired at 25 °C of A) Milli Q water, fluorescence buffer and fluorescence buffer containing 5 nM NADPH and B) Fluorescence buffer containing 5 mM NADPH.

It is apparent that a significant fluorescence signal is observed at 454 nm, which is present in Milli Q water. This signal is enhanced when a sample of the buffer is analysed and upon the addition of nM concentrations of NADPH to the sample, an initial quenching of the signal at 454 nm is observed. As previously documented, once the concentration of NADPH present in the sample is of the mM concentration, the fluorescence of water is obliterated and the characteristic increase in fluorescence at 454 nm is observed Figure 6.26 B. Heyes *et al* documented that upon binding of NADPH to POR, an enhancement of the signal at 454 nm is observed (Heyes *et al.* 2000). Whilst this is not the case when measuring nM concentrations of NADPH, upon binding of NADPH to POR, an increase in the level of quenching observed at 454 nm, compared to that observed with NADPH alone, can be used to monitor the binding of NADPH to POR (Figure 6.27).

The fluorescence spectrum of a sample containing 20 nM POR was analysed, following the sequential addition of between 0 and 40 nM NADPH (Figure 6.27 A). Figure 6.27 A shows the raw data acquired for free NADPH along with that of the wild type

and G154/5/90A mutant, whilst Figure 6.27 B shows the reciprocal of the wild type and mutant values fitted to Equation 2.2, following the subtraction of the NADPH control.

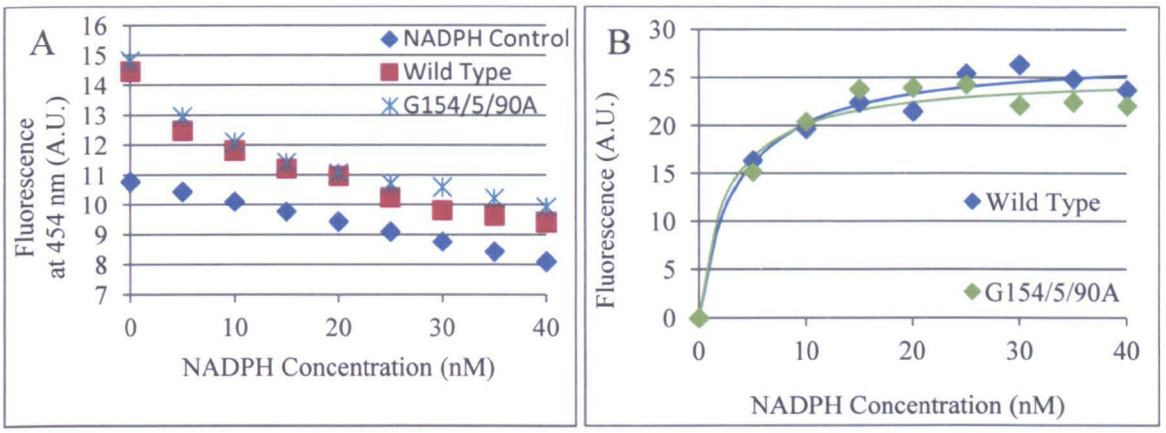


Figure 6.27: A) Graph showing the changes in fluorescence, measured at 454 nm, in the presence and absence of POR, with increasing concentrations of NADPH and B) Graph showing the binding of NADPH to wild type and G154/5/90A POR fitted to Equation 2.2.

The two enzymes appeared to bind NADPH with the same affinity with the K_d for wild type and G154/5/90A POR being calculated as 3.5 and 2.6 nM respectively. These values calculated are within error of those previously calculated for the *T. elongatus* POR enzyme (Menon *et al.* 2010) and implied that the mutations made had not affected the enzyme's ability to bind NADPH. Since the G154/5/90A mutant showed a similar K_d for NADPH as that observed for the wild type, it was decided not to investigate the binding of NADPH to any of the other mutants, as the reduction in enzyme activity was likely to be a result of another factor.

6.6.3.2 Formation of Glycine Mutant POR – Pchlide Ternary Complexes

POR, $NADP^+$ and Pchlide Ternary complexes were made with wild type and the three glycine mutant enzymes and purified from an SP sepharose column (detailed 2.16). The UV absorbance at 280 nm was measured for the eluant from all four complexes (Figure 6.28).

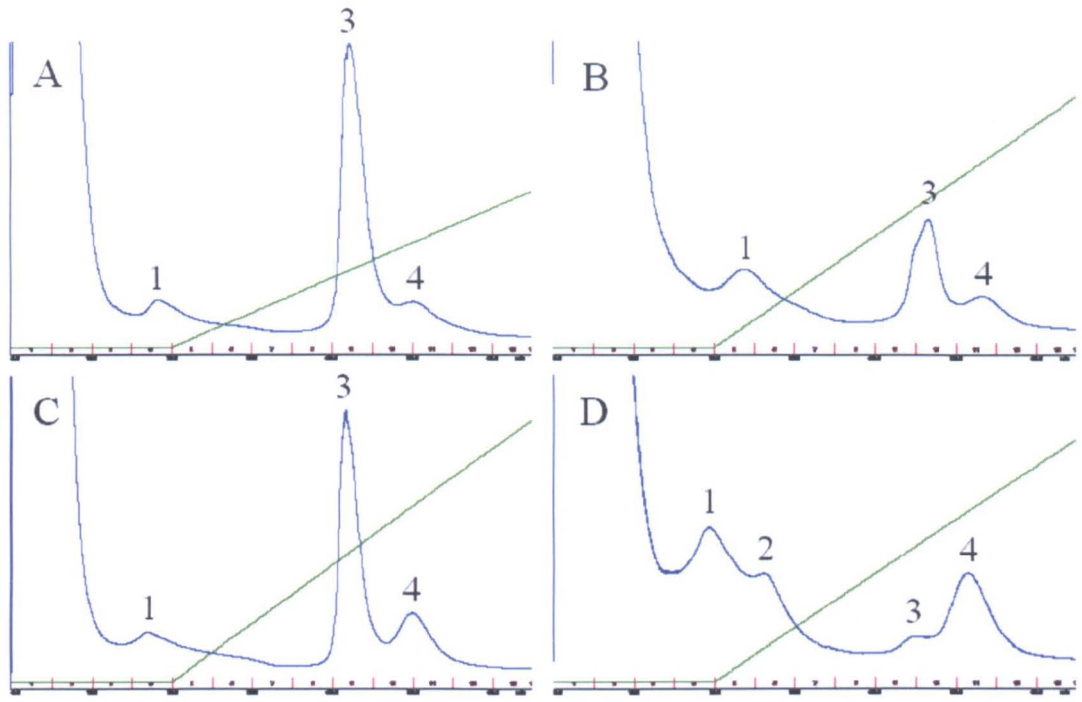


Figure 6.28: Elution profile showing the change in absorbance at 280 nm of the eluant from the SP sepharose column during the purification of A) Wild Type, B) G154/5A, C) G190A, D) G154/5/90A ternary complex.

The purification of the G154/5A and G190A ternary complexes was the same as that previously observed with wild type ternary complex (detailed 6.4), with the proportion of protein present in peaks 1, 3 and 4 differing between all preparations. SDS PAGE analysis of peaks 1, 3 and 4 showed that all peaks contain POR, with peaks 1 and 3 being green in colour and peak 4 being colourless. The elution profile of the G154/5/90A ternary complex was different to that observed with the other three complexes, with the appearance of a fourth peak. The distribution of POR between the four fractions was also considerably different, with the majority of POR being eluted from the column in peaks 1 and 2. As with all other preparations, peaks 1, 2 and 3 were green in colour and peak 4 was colourless. The interpretation of these results was that peaks 1 and 3 corresponded to different oligomeric forms of the POR-Pchlide-$NADP^+$ ternary complex, whilst peak 4 corresponded to a POR enzyme which was incapable of binding Pchlide. If this is true, the increasing proportion of enzyme found in Peak 4 could contribute to the loss of enzyme activity observed in Figure 6.25 and Table 6.7.

The major POR-Pchlide-$NADP^+$ ternary complex-containing peaks were concentrated in a viva spin containing a 10 kDa PES membrane. The protein concentration of the four

samples was determined, using the Bradford assay (detailed 2.11) and an absorbance spectrum taken between 260 and 700 indicated that the ratio of pigment to enzyme was 3:1. Assays were conducted with wild type and mutant POR enzymes, which already had Pchlide bound in the active site, to give a more accurate idea of how the mutations had affected the enzyme activity. Reactions were set up containing an excess of NADPH and placed in a Cary 50 spectrophotometer and assayed at 55 °C under constant illumination (described 2.15) (Figure 6.29). As detailed in 6.6.2 the concentration of enzyme and reactants was kept constant across all experiments, therefore allowing the amount of Chlide produced to act as a direct measure of how the mutations have affected the enzyme activity. A summary of remaining activity is detailed in Table 6.8.

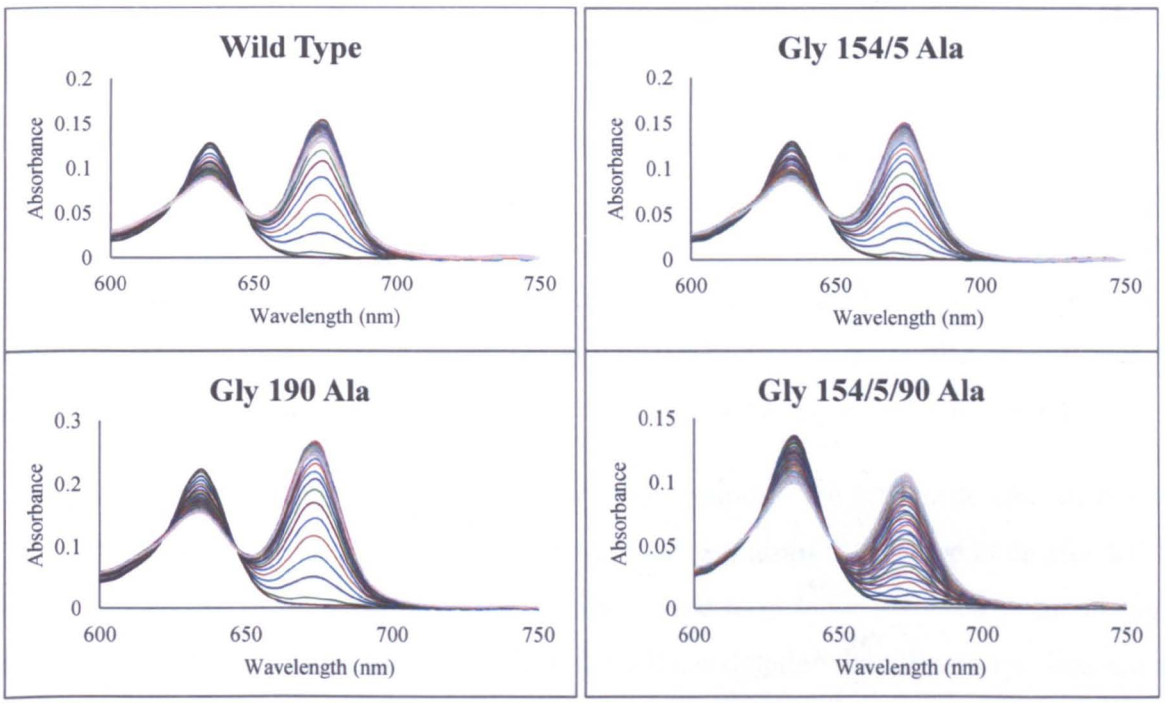


Figure 6.29: Spectra showing the increase in Chlide concentration, using the glycine mutant POR ternary complex, over a period of 90 seconds with constant illumination.

In all spectra acquired, only one third of the Pchlide present is photo-converted into Chlide. This implies that although there is three times the concentration of Pchlide to enzyme, only the pigment which is bound to the enzyme is photo-convertible.

Mutation	μM Chlide Increase/Minute/mg POR	Relative Activity Remaining (%)
Wild Type	9211	100
G154/5A	9211	100
G190A	4665	51
G154/5/90A	598	6

Table 6.8: Table detailing the levels of enzyme activity present in wild type POR ternary complex along with that remaining with the glycine mutant ternary complexes.

The rate of Chlide formation in the wild type and G154/5A complexes, which already had the pigment bound to the active site, were identical. Although this is likely to be a coincidence, it illustrates the point that the mutations made to Gly-154 and Gly-155 are not catalytically relevant. The rates of Chlide formation in the G190A and G154/5/90A complexes are significantly reduced however, compared to the rate found in the wild type enzyme. Since the rate of Chlide formation in the G154/5/90A mutant is further reduced, compared to that found in the G190A and G154/5A mutants, this implies that the different glycine mutations have a co-operative effect in reducing the enzyme activity.

As the purified enzyme already has a Pchlide bound to the active site and the binding of NADPH is not affected, these results imply that the mutations to the loop have affected the catalytic properties of the enzyme, thus implicating the loop to be involved in the catalytic reaction of POR. As detailed above, published work has detailed that protein motions are not required for the initial light-driven catalytic step (Heyes *et al.* 2002b; Heyes *et al.* 2003b). There remain several explanations for the observation made above: It is of course quite possible that movement of the loop is necessary to allow Pchlide to sit properly into the active site. So whilst Pchlide is bound, this does not prove that it is bound in the correct conformation. Also whilst the mutations to the loop are unlikely to have directly affected the enzymes ability to reduce the $C_{17} - C_{18}$ double bond of Pchlide, the mutations to the loop may have affected the enzymes ability to form an aggregate, which may be a pre-requisite for POR activity. To test this, a native gel was run (detailed 2.20) of all purified complexes Figure 6.30, which had previously been analysed Figure 6.29.

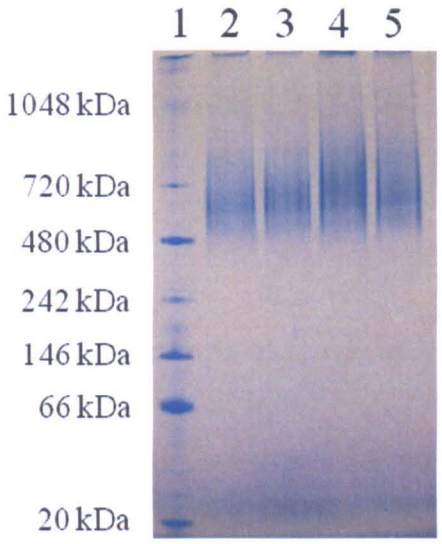


Figure 6.30: Native gel of 2) Wild Type 3) G154/5A 4) G190A, 5) G154/5/90A ternary complex.

Whilst the bands present in the native gel indicate that the mutations made have not affected the enzymes ability to form aggregates, the absence of clear distinct bands is concerning. As a result samples of the four ternary complex samples produced have been sent to Professor Steve Harding, at the University of Nottingham for further analysis by analytical ultra centrifugation. However at the point of publication, we are awaiting the results.

The residues to mutate were chosen purely on where the loop was proposed to be located by Townley *et al.* (Townley *et al.* 2001) along with the information acquired from the new homology models created (detailed 5.4.1). However since completing the mutagenesis, a proportion of the POR spectrum has been assigned and details corresponding to the correlation time of certain residues have been obtained (detailed 5.6.4). This data suggests that Gly-154 and Gly-155 are actually in the middle of a mobile region and not at the end like initially thought. Whilst it may be conceivable that there must be a certain level of flexibility associated with Gly-154 and Gly-155 to allow the loop to open up, it may well be the case that Gly-144, highlighted blue in Figure 6.23, may actually be located at the edge of the mobile loop. It would be interesting to observe if mutation of this residue to alanine would result in a similar loss of activity to that observed with the mutation of Gly-190.

CHAPTER 7

Conclusions and Future Work

7.1 Conclusions

Following the optimisation of the growth of *E. coli* and the expression of His_6-POR a three fold increase in protein yield was achieved from 10 mg / litre to 30 mg / litre. This was done by optimising a number of steps in the over-expression protocol. It was determined that following induction, *E. coli* growth and POR production only continues for a further two hours, also by inducing the *E. coli* at a higher optical density (600 nm) than 0.6, more protein could be over-expressed per ml of medium. It was also shown that inducing the *E. coli* at 25 °C resulted in a larger proportion of the over-expressed protein remaining soluble. The use of a thermophilic enzyme allowed the incorporation of a heating step into the purification protocol. Following sonication, heating the initial *E. coli* extract at 45 °C, resulted in the denaturation of a significant proportion of contaminating proteins. The purification of the resulting cell extract through a nickel His-60 IMAC, and an SP sepharose column resulted in the purification of POR to greater than 95 % purity.

POR has never been studied extensively using NMR so a set of optimal conditions needed to be established, to provide the best quality of spectra possible. Optimal spectra were achieved at lower pH and at higher temperature. Wild type POR denatures at a temperature of 60 °C and undergoes significant thermal denaturation following 60 hours incubation at 55 °C. The protein is stable for an indefinite period of time at 45 °C, which is hot enough to significantly reduce the correlation time of POR to between 13 and 16 ns. All studies to date have shown that when left at room temperature, POR will degrade into a 31 kDa and 6 kDa fragment. The incorporation of a protease inhibitor which included EDTA was sufficient to prevent the degradation of POR and maintain the stability of the enzyme at temperatures up to 45 °C. The optimal conditions determined for running NMR experiments were discovered to be 50 mM sodium phosphate buffer pH 5.5, 100 mM sodium chloride, 20 mM DTT and Roche complete protease inhibitor. The quality of spectra acquired also improves when analysed at higher field strength; at field strengths of 800 MHz and above TROSY versions of experiments provide better quality spectra, whilst at field strengths of 600 MHz and below, the non-TROSY HSQC equivalents are preferential.

Deuterated protein has been produced, however the spectra acquired do not exhibit the improvements in spectral quality which are normally associated with such preparations. Purification of the deuterated protein also resulted in the formation of a brown reduced nickel-EDTA salt implying that the protein had co-eluted with nickel during the IMAC stage of the purification protocol. It is currently unknown as to why the quality of the spectra did not improve, as expected with the deuterated sample, however proposals have been made that a proportion of nickel is still present in the sample, which is relaxing the protein signals explaining the reduction in quality of the spectra acquired. Another hypothesis has been proposed that the acclimatisation of *E. coli* to growth in D_2O may have induced certain chaperones or co-metabolites, which have resulted in the production and purification of a protein which has altered spectral characteristics.

Using a ^{13}C, ^{15}N labelled sample of POR, just fewer than 25 % of the residues were confidently assigned to their corresponding peaks in the POR NMR spectrum, which included approximately one third of the residues present in the 33-residue insertion loop. Analysis of the correlation time of the assigned residues indicated that the loop region of POR had a higher mobility than the rest of the protein. A new homology model of POR has also been created using the online structure prediction programme, iTasser. The model produce by iTasser is different to the model produced by Townley *et al.* specifically in the location of the 33-residue insertion loop. Whilst Townley places the loop at the top of the protein, iTasser places the loop adjacent to the fourth and sixth α-helices. The confident assignments were analysed by the structure assignment software TALOS and indicated that where a secondary structure could be assigned, this was in agreement with the model produced by iTasser.

A method was developed for the production and purification of a POR-Pchlide-$NADP^+$ ternary complex in the absence of detergent and methanol. Upon elution from the ion exchange chromatography column, the fractions detected to contain POR were also green in colour. Analysis of the produced ternary complex by HPLC indicated that the POR and Pchlide were associated with each other and also indicated that the complex produced had a molecular weight of greater than 200 kDa. This was confirmed by native gel and electron microscopy images taken of the sample, which indicated that the complex was composed of approximately 10 units, associated as a pentamer of dimers. Analytical ultracentrifugation indicated that following the substitution of the bound $NADP^+$ for NADPH and the exposure to light, the complexes disperse back to a population of monomers and dimers.

Comparison of POR sequences across a number of different organisms indicated that a number of glycine residues were heavily conserved throughout the 33-residue loop region. Three glycine residues, which were located on the two edges of the proposed hinge, were mutated to alanine residues. The G154/5A and G190A mutants showed a reduction in enzyme activity, with the two mutants exhibiting 34 and 80 % less activity than wild type POR. The G154/5/90A mutant however lost 97.5 % of the activity normally associated with wild type POR. Substrate binding studies indicated that the G154/5/90A mutant binds NADPH with the same affinity as that observed with the wild type enzyme, however production of POR-Pchlide-$NADP^+$ ternary complexes indicated that a significant proportion of the mutant proteins were unable to bind Pchlide. It is worth noting that whilst the loop makes no direct contact with Pchlide, the linker running up to the loop does. It is thus not unreasonable that mobility in the loop may be related to productive binding of Pchlide. Also the purification properties of the ternary complex samples produced were altered, with the wild type, G154/5A and G190A complexes purifying the same way and the G154/5/90A mutant having a different affinity for the ion exchange column. Analysis of the rate of Chlide formation of POR samples which already had a Pchlide molecule bound to the enzyme active site indicated that the G190A and G154/5/90A POR mutants were severely reduced in their capacity to produce Chlide, thus implicating the loop in the catalytic reaction of POR.

7.2 Future Work

Although the original aim of this PhD was to assign the POR spectrum and use this to determine the secondary structure, for the reasons detailed above, this was not achieved. However the work conducted does indicate that solving the structure of POR using NMR is a possibility, providing a better sample can be created. There are a number of samples which could be created which in theory should give better spectra than a ^{13}C, ^{15}N double labelled however, initially work should be concentrated upon producing a deuterated sample which provides all the improvements expected from a deuterated sample.

It was unknown as to why the quality of the spectra of deuterated protein acquired were so poor, however a number of things could be altered in the over-expression and purification of the protein which could yield the improvements expected. Firstly instead of acclimatising the *E. coli* in LB medium containing increasing concentrations of D_2O, a fresh triple labelled *E. coli* starter culture could be produced each time. This could be left to grow

to completion over a period of days, allowing the *E. coli* to become acclimatised this way. Instead of bulking up the cell mass in LB medium before over-expressing POR in the triple labelled medium, the *E. coli* could then be inoculated straight into the triple labelled medium. By growing the *E. coli* this way, the overall number of cells over-expressing POR will be less, however if the protein produced provided significantly better quality spectra then this is a trade off which is worth making. Another worthy alteration to the protocol would be to find a nickel resin which did not leach nickel into the final POR sample. A range of resins are available so it should be possible to find one which binds its divalent metal ion stronger.

Another possibility of improving the spectral quality is to produce a monomeric ternary complex sample. Quite unexpectedly the ternary complex produced *in vitro* is an aggregate of approximately 10 units, which can not be analysed using NMR. Initial analytical ultra centrifugation (AUC) data suggests that a Chlide bound ternary complex is monomeric, however more work is required in order to conclusively prove this. Currently the only way of producing Chlide is through the light induced reaction of Pchlide with POR; this however is problematic as Chlide is sensitive to and degrades upon illumination. A protocol therefore needs to be developed by which Chlide can be over-expressed and purified on the same scale as the way Pchlide is prepared. From the assays conducted above, it has also become apparent that there is a problem with the Pchlide provided to the reaction.

To date all crystallography trials have either been conducted with apo POR, or using samples which have had substrates present in excess. To date no crystal trials have been conducted using the Ternary complex produced. Providing a mono-dispersed sample can be produced, this may well provide the best opportunity to date to successfully crystallise POR. Prior to doing this, it would be a good idea to identify why two thirds of the Pchlide present in the Ternary Complex samples remains un-reacted. Suggestions have been made above indicating as to why the Pchlide provided to all reactions remains un-reacted, however it would be a good idea to alter the preparation of the Pchlide thus improving the purity of it. Preparative HPLC would ensure that the Pchlide produced would have the high degree of purity required, however this may not be practical when producing the quantities of Chlide which would be required.

Whilst the glycine mutants were made with the intention of hindering the mobility of the loop, the G154/5A mutation appeared to have no real effect on the POR enzyme activity. NMR experiments conducted suggest that there is a high degree of mobility extending

beyond the predicted position of the 33 residue insertion loop. Another highly conserved glycine is positioned at residue 140, which could conceivably be the hinge at the other end of the loop. Mutation of this residue, coupled to the mutations already made could remove all activity from POR. The AUC experiments conducted act as a good way of characterising the different complexes which are present in each of the ternary complex samples produced. AUC on the ternary complexes produced with the mutant enzymes, could further elucidate the function of the loop, along with starting to explain why there is a significant loss of activity with the mutant enzymes.

References

Addlesee, H. A., Fiedor, L. and 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." Journal of Bacteriology **182** (11): 3175-3182.

Addlesee, H. A., Gibson, L. C. D., Jensen, P. E. and Hunter, C. N. (1996). "Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, *chlP*, of *Synechocystis* sp PCC 6803." Febs Letters **389** (2): 126-130.

Addlesee, H. A. and Hunter, C. N. (1999). "Physical Mapping and Functional Assignment of the Geranylgeranyl-Bacteriochlorophyll Reductase Gene, *bchP*, of *Rhodobacter sphaeroides*." Journal of Bacteriology **181** (23): 7248-7255.

Adra, A. N. and 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." Photochemistry and Photobiology **68** (6): 852-856.

Akhtar, M. (2003). Coproporphyrinogen III and protoporphyrinogen IX oxidases. In. Kadish KM, Smith KM and Guilard R The iron and cobalt pigments: biosynthesis, structure and degradation. New York, Elsevier. **12:** 75-92.

Al-Karadaghi, S., Franco, R., Hansson, M., Shelnutt, J. A., Isaya, G. and Ferreira, G. C. (2006). "Chelatases: distort to select?" Trends in Biochemical Sciences **31** (3): 135-142.

Al-Karadaghi, S., Hansson, M., Nikonov, S., Jonsson, B. and Hederstedt, L. (1997). "Crystal structure of ferrochelatase: the terminal enzyme in heme biosynthesis." Structure **5** (11): 1501-1510.

Apel, K. (1981). "The Protochlorophyllide Holochrome of Barley (*Hordeum vulgare* L.) Phytochrome-Induced Decrease of Translatable mRNA Coding for the NADPH:Protochlorophyllide Oxidoreductase." European Journal of Biochemistry **120** (1): 89-93.

Apel, K., Santel, H. J., Redlinger, T. E. and Falk, H. (1980). "The Protochlorophyllide Holochrome of Barley (*Hordeum vulgare* L.) Isolation and Characterization of the NADPH:Protochlorophyllide Oxidoreductase." European Journal of Biochemistry **111** (1): 251-258.

Argyroudi-Akoyunoglou, J. H. and Prombona, A. (1996). "Light-independent endogenous circadian rhythm in the capacity for chlorophyll formation." Journal of Photochemistry and Photobiology B-Biology **36** (3): 271-277.

Armstrong, G. A. (1998). "Greening in the dark: light-independent chlorophyll biosynthesis from anoxygenic photosynthetic bacteria to gymnosperms." Journal of Photochemistry and Photobiology B-Biology **43** (2): 87-100.

Armstrong, G. A., Apel, K. and Rudiger, W. H. (2000). "Does a light-harvesting protochlorophyllide *al b*-binding protein complex exist?" Trends in Plant Science **5** (1): 40-44.

Armstrong, G. A., Runge, S., Frick, G., Sperling, U. and Apel, K. (1995). "Identification of NADPH:Protochlorophyllide Oxidoreductases A and B: A Branched Pathway for Light-Dependent Chlorophyll Biosynthesis in *Arabidopsis thaliana*[1]." Plant Physiology **108** (4): 1505-1517.

Aronsson, H., Sohrt, K. and Soll, J. (2000). "NADPH:Protochlorophyllide Oxidoreductase Uses the General Import Route into Chloroplasts." Biological Chemistry **381** (12): 1263-1267.

Aubert, C., Vos, M. H., Mathis, P., Eker, A. P. M. and Brettel, K. (2000). "Intraprotein radical transfer during photoactivation of DNA photolyase." Nature **405** (6786): 586-590.

Avissar, Y. J. and Moberg, P. A. (1995). "The common origins of the pigments of life - early steps of chlorophyll biosynthesis." Photosynthesis Research **44** (3): 221-242.

Axelsson, E., Lundqvist, J., Sawicki, A., Nilsson, S., Schroder, I., Al-Karadaghi, S., Willows, R. D. and Hansson, M. (2006). "Recessiveness and Dominance in Barley Mutants Deficient in Mg-Chelatase Subunit D, an AAA Protein Involved in Chlorophyll Biosynthesis." Plant Cell **18** (12): 3606-3616.

Baker, M. E. (1994). "Protochlorophyllide reductase is homologous to human carbonyl reductase and pig 20β-hydroxysteroid dehydrogenase." Biochemical Journal **300**: 605-607.

Batschauer, A. and Apel, K. (1984). "An inverse control by phytochrome of the expression of two nuclear genes in barley (*Hordeum vulgare* L.)." European Journal of Biochemistry **143** (3): 593-597.

Battersby, A. R., Fookes, C. J. R., Gustafsonpotter, K. E., Matcham, G. W. J. and McDonald, E. (1979a). "Proof by Synthesis that Unrearranged Hydroxymethylbilane is the Product from Deaminase and the Substrate for Cosynthetase in the Biosynthesis of Uro'gen-III." Journal of the Chemical Society-Chemical Communications (24): 1155-1158.

Battersby, A. R., Fookes, C. J. R., Hart, G., Matcham, G. W. J. and Pandey, P. S. (1983a). "Biosynthesis of Porphyrins and Related Macrocycles. Part 21.[1] The Interaction of Deaminase and its Product (Hydroxymethylbilane) and the Relationship between Deaminase and Cosynthase." Journal of the Chemical Society-Perkin Transactions 1 (12): 3041-3047.

Battersby, A. R., Fookes, C. J. R., Matcham, G. W. J. and McDonald, E. (1979b). "Order of Assembly of the Four Pyrrole Rings during Biosynthesis of the Natural Porphyrins." Journal of the Chemical Society-Chemical Communications (12): 539-541.

Battersby, A. R., Fookes, C. J. R., Matcham, G. W. J., McDonald, E. and Hollenstein, R. (1983b). "Biosynthesis of Porphyrins and Related Macrocycles. 20.[1,2] Purification of Deaminase and Studies on its Mode of Action." Journal of the Chemical Society-Perkin Transactions 1 (12): 3031-3040.

Battersby, A. R., McDonald, E., Wurziger, H. K. W. and James, K. J. (1975). "Stereochemistry of Biosynthesis of Vinyl Groups of Protoporphyrin-IX: A Short Synthesis of Porphobilinogen." Journal of the Chemical Society-Chemical Communications (13): 493-494.

Bauer, C. E., Bollivar, D. W. and Suzuki, J. Y. (1993). "Genetic Analyses of Photopigment Biosynthesis in Eubacteria: a Guiding Light for Algae and Plants." Journal of Bacteriology **175** (13): 3919-3925.

Bax, A. and Grzesiek, S. (1993). "Methodological Advances in Protein NMR." Accounts of Chemical Research **26** (4): 131-138.

Beale, S. I. (1999). "Enzymes of chlorophyll biosynthesis." Photosynthesis Research **60** (1): 43-73.

Begley, T. P. and Young, H. (1989). "Protochlorophyllide Reductase. 1. Determination of the Regiochemistry and the Stereochemistry of the Reduction of Protochlorophyllide to Chlorophyllide." Journal of the American Chemical Society **111** (8): 3095-3096.

Belyaeva, O. B., Timofeev, K. N. and Litvin, F. F. (1988). "The primary reactions in the protochlorophyll(ide) photoreduction as investigated by optical and ESR spectroscopy." Photosynthesis Research **15** (3): 247-256.

Benkovic, S. J. and Hammes-Schiffer, S. (2003). "A Perspective on Enzyme Catalysis." Science **301** (5637): 1196-1202.

Benli, M., Schulz, R. and Apel, K. (1991). "Effect of light on the NADPH-protochlorophyllide oxidoreductase of *Arabidopsis thaliana*." Plant Molecular Biology **16** (4): 615-625.

Benz, J., Wolf, C. and Rudiger, W. (1980). "Chlorophyll biosynthesis: Hydrogenation of geranylgeraniol." Plant Science Letters **19** (3): 225-230.

Birve, S. J., Selstam, E. and Johansson, L. B. A. (1996). "Secondary structure of NADPH:protochlorophyllide oxidoreductase examined by circular dichroism and prediction methods." Biochemical Journal **317**: 549-555.

Bollivar, D. W., Jiang, Z. Y., Bauer, C. E. and Beale, S. I. (1994a). "Heterologous Expression of the *BchM* Gene Product from *Rhodobacter capsulatus* and Demonstration that it Encodes *S*-adenosyl-L-methionine:Mg-Protoporphyrin IX Methyltransferase " Journal of Bacteriology **176** (17): 5290-5296.

Bollivar, D. W., Suzuki, J. Y., Beatty, J. T., Dobrowolski, J. M. and Bauer, C. E. (1994b). "Directed mutational analysis of bacteriochlorophyll *a* biosynthesis in *Rhodobacter capsulatus*." Journal of Molecular Biology **237** (5): 622-640.

Breckau, D., Mahlitz, E., Sauerwald, A., Layer, G. and Jahn, D. (2003). "Oxygen-dependent Coproporphyrinogen III Oxidase (HemF) from *Escherichia coli* Is Stimulated by Manganese." Journal of Biological Chemistry **278** (47): 46625-46631.

Brindley, A. A., Raux, E., Leech, H. K., Schubert, H. L. and Warren, M. J. (2003). "A Story of Chelatase Evolution Identification and Characterization of a Small 13-15 kDa "Ancestral" Cobaltchelatase ($CbiX^S$) In the Archea." Journal of Biological Chemistry **278** (25): 22388-22395.

Bryant, D. A. and Frigaard, N. U. (2006). "Prokaryotic photosynthesis and phototrophy illuminated." Trends in Microbiology **14** (11): 488-496.

Burke, D. H., Alberti, M. and 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." Journal of Bacteriology **175** (8): 2414-2422.

Burton, G., Fagerness, P. E., Hosozawa, S., Jordan, P. M. and Scott, A. I. (1979). "^{13}C N.M.R. Evidence for a New Intermediate, Pre-uroporphyrinogen, in the Enzymatic Transformation of Porphobilinogen into Uroporphyrinogens I and III." Journal of the Chemical Society-Chemical Communications (5): 202-204.

Camadro, J. M., Thome, F., Brouillet, N. and 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." Journal of Biological Chemistry **269** (51): 32085-32091.

Cavaleiro, J. A., Kenner, G. W. and Smith, K. M. (1974). "Pyrroles and related compounds. XXXII. Biosynthesis of protoporphyrin-IX from coproporphyrinogen-3." J Chem Soc Perkin 1 **10** (0): 1188-1194.

Cavallaro, G., Decaria, L. and Rosato, A. (2008). "Genome-Based Analysis of Heme Biosynthesis and Uptake in Prokaryotic Systems." Journal of Proteome Research **7** (11): 4946-4954.

Chen, Z., Jiang, J. C., Lin, Z. G., Lee, W. R., Baker, M. E. and Chang, S. H. (1993). "Site-Specific Mutagenesis of *Drosophila* Alcohol Dehydrogease: Evidence for Involvement of Tyrosine-152 and Lysine-156 in Catalysis." Biochemistry **32** (13): 3342-3346.

Chew, A. G. M. and Bryant, D. A. (2007). "Chlorophyll Biosynthesis in Bacteria: The Origins of Structural and Functional Diversity." Annu Rev Microbiol **61**: 113-129.

Chisholm, S. W., Frankel, S. L., Goericke, R., Olson, R. J., Palenik, B., Waterbury, J. B., Westjohnsrud, L. and Zettler, E. R. (1992). "*Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*." Archives of Microbiology **157** (3): 297-300.

Coleman, M. L. and Chisholm, S. W. (2007). "Code and context: *Prochlorococcus* as a model for cross-scale biology." Trends in Microbiology **15**: 398-407.

Coomber, S. A., Chaudhri, M., Connor, A., Britton, G. and Hunter, C. N. (1990). "Localized transposom Tn5 mutagenesis of the photosynthetic gene cluster of *Rhodobacter sphaeroides*." Molecular Microbiology **4** (6): 977-989.

Cornah, J. E., Roper, J. M., Singh, D. P. and 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.)." Biochemical Journal **362**: 423-432.

Cornilescu, G., Delaglio, F. and Bax, A. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." Journal of Biomolecular Nmr **13** (3): 289-302.

Corradi, H. R., Corrigall, A. V., Boix, E., Mohan, C. G., Sturrock, E. D., Meissner, P. N. and Acharya, K. R. (2006). "Crystal Structure of Protoporphyrinogen Oxidase from *Myxococcus xanthus* and Its Complex with the Inhibitor Acifluorfen*." Journal of Biological Chemistry **281** (50): 38625-38633.

Crockett, N., Alefounder, P. R., Battersby, A. R. and Abell, C. (1991). "Uroporphyrinogen III synthase: Studies on its mechanism of action, molecular biology and biochemistry." Tetrahedron **47** (31): 6003-6014.

Dahlin, C., Sundqvist, C. and Timko, M. P. (1995). "The *in vitro* assembly of the NADPH-protochlorophyllide oxidoreductase in pea chloroplasts." Plant Molecular Biology **29** (2): 317-330.

Dailey, H. A., Finnegan, M. G. and Johnson, M. K. (1994). "Human ferrochelatase Is an Iron-Sulfur Protein." Biochemistry **33** (2): 403-407.

Dailey, T. A. and Dailey, H. A. (2002). "Identification of [2Fe-2S] Clusters in Microbial Ferrochelatases." Journal of Bacteriology **184** (9): 2460-2464.

Davison, P. A., Schubert, H. L., Reid, J. D., Iorg, C. D., Heroux, A., Hill, C. P. and Hunter, C. N. (2005). "Structural and Biochemical Characterization of Gun4 Suggests a Mechanism for Its Role in Chlorophyll Biosynthesis." Biochemistry **44** (21): 7603-7612.

Day, A. L., Parsons, B. M. and Dailey, H. A. (1998). "Cloning and Characterization of *Gallus* and *Xenopus* Ferrochelatases: Presence of the [2Fe-2S] Cluster in Nonmammalian Ferrochelatase[1]." Archives of Biochemistry and Biophysics **359** (2): 160-169.

Dean, D. R., Bolin, J. T. and Zheng, L. M. (1993). "Nitrogenase Metalloclusters: Structures, Organization, and Synthesis." Journal of Bacteriology **175** (21): 6737-6744.

Durin, G. (2006). Mechanisme catalytique de la protochlorophyllide oxydoreductase : etude du couplage entre activite fonctionnelle et dynamique de la proteine. Physique. France, Universite Joseph Fourier - Grenoble 1 Science and Geodraphie: 146.

Dvorsky, R., Sevcik, J., Caves, L. S. D., Hubbard, R. E. and Verma, C. S. (2000). "Temperature Effects on Protein Motions: A Molecular Dynamics Study of RNase-*Sa*." Journal of Physical Chemistry B **104** (44): 10387-10397.

Elder, G. H., Evans, J. O., Jackson, J. R. and Jackson, A. H. (1978). "Factors Determining Sequence of Oxidative Decarboxylation of the 2-and 4-Propionate Substituents

of Coproporphyrinogen III by Coproporphyrinogen Oxidase in Rat Liver." Biochemical Journal **169** (1): 215-223.

Ensor, C. M. and Tai, H. H. (1991). "Site-directed mutagenesis of the conserved tyrosine 151 of human placental NAD^+-dependent 15-hydroxyprostaglandin dehydrogenase yields a catalytically inactive enzyme." Biochemical and Biophysical Research Communications **176** (2): 840-845.

Erskine, P. T., Coates, L., Newbold, R., Brindley, A. A., Stauffer, F., Beaven, G. D. E., Gill, R., Coker, A., Wood, S. P., Warren, M. J., Shoolingin-Jordan, P. M., Neier, R. and Cooper, J. B. (2005). "Structure of yeast 5-aminolaevulinic acid dehydratase complexed with the inhibitor 5-hydroxylaevulinic acid." Acta Crystallographica Section D-Biological Crystallography **61**: 1222-1226.

Erskine, P. T., Newbold, R., Brindley, A. A., Wood, S. P., Shoolingin-Jordan, P. M., Warren, M. J. and Cooper, J. B. (2001). "The X-ray structure of yeast 5-aminolaevulinic acid dehydratase complexed with substrate and three inhibitors." Journal of Molecular Biology **312** (1): 133-141.

Erskine, P. T., Newbold, R., Roper, J., Coker, A., Warren, M. J., Shoolingin-Jordan, P. M., Wood, S. P. and Cooper, J. B. (1999a). "The Schiff base complex of yeast 5-aminolaevulinic acid dehydratase with laevulinic acid." Protein Science **8** (6): 1250-1256.

Erskine, P. T., Norton, E., Cooper, J. B., Lambert, R., Coker, A., Lewis, G., Spencer, P., Sarwar, M., Wood, S. P., Warren, M. J. and Schoolingin-Jordan, P. M. (1999b). "X-ray structure of 5-aminolevulinic acid dehydratase from Escherichia coli complexed with the inhibitor levulinic acid at 2.0 angstrom resolution." Biochemistry **38** (14): 4266-4276.

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F. T., Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. and Steffen, W. (2000). "The global carbon cycle: A test of our knowledge of earth as a system." Science **290** (5490): 291-296.

Feldman, G. "SeaWiFS Project." Retrieved 28th June, 2011, from http://oceancolor.gsfc.nasa.gov/SeaWiFS/.

Ferreira, G. C. and Dailey, H. A. (1988). "Mouse protoporphyrinogen oxidase - Kinetic parameters and demonstration of inhibition by bilirubin." Biochemical Journal **250** (2): 597-603.

Fiedor, L., Kania, A., Mysliwa-Kurdziel, B., Orzel, L. and Stochel, G. (2008). "Understanding chlorophylls: Central magnesium ion and phytyl as structural determinants." Biochimica Et Biophysica Acta-Bioenergetics **1777** (12): 1491-1500.

Fields, P. A. (2001). "Review: Protein function at thermal extremes: balancing stability and flexibility." Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology **129** (2-3): 417-431.

Fodje, M. N., Hansson, A., Hansson, M., Olsen, J. G., Gough, S., Willows, R. D. and Al-Karadaghi, S. (2001). "Interplay between an AAA module and an integrin I domain may

regulate the function of magnesium chelatase." Journal of Molecular Biology **311** (1): 111-122.

Forreiter, C., Vancleve, B., Schmidt, A. and Apel, K. (1991). "Evidence for a General Light-Dependent Negative Control of Nadph-Protochlorophyllide Oxidoreductase in Angiosperms." Planta **183** (1): 126-132.

Fuesler, T. P., Wong, Y. S. and Castelfranco, P. A. (1984). "Localization of Mg-chelatase and Mg-protoporphyrin-IX monomethyl ester (oxidative) cyclase activities within isolated, developing cucumber chloroplasts." Plant Physiology **75** (3): 662-664.

Fujita, Y. (1996). "Protochlorophyllide reduction: A key step in the greening of plants." Plant and Cell Physiology **37** (4): 411-421.

Fujita, Y. and 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." Journal of Biological Chemistry **275** (31): 23583-23588.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and A., B. (2005). "Protein Identification and Analysis Tools on the ExPASy Server; ." The Proteomics Protocols Handbook. Retrieved 29 June, 2011.

Gibson, K. D., Laver, W. G. and Neuberger, A. (1958). "Initial Stages in the Biosynthesis of Porphyrins .2. The Formation of δ-Aminolaevilinic Acid from Glycine and Succinyl-Coenzyme A by Particles from Chicken Erythrocytes." Biochemical Journal **70**: 71-81.

Gibson, K. D., Neuberger, A. and Scott, J. J. (1955). "Purification and Properties of δ-Aminolevilinic Acid Dehydratase." Biochemical Journal **61** (4): 618-629.

Gibson, K. D., Tait, G. H. and Neuberger, A. (1963). "Studies on Biosynthesis of Porphyrin and Bacteriochlorophyll by *Rhodopseudomonas spheroides* .4. *S*-Adenosylmethionine-Magnesium Protoporphyrin Methyltransferase." Biochemical Journal **88** (2): 325-&.

Gibson, L. C. D. and Hunter, C. N. (1994). "The bacteriochlorophyll biosynthesis gene, *bchM*, of *Rhodobacter sphaeroides* encodes *S*-adenosyl-L-methionine:Mg protoporphyrin IX methyltransferase." Febs Letters **352** (2): 127-130.

Gibson, L. C. D., Willows, R. D., Kannangara, C. G., Von Wettstein, D. and Hunter, C. N. (1995). "Magnesium-protoporphyrin chelatase of *Rhodobacter-sphaeroides*: Reconstitution of activity by combining the products of the *bchH, -I* and *-D* genes expressed in *Escherichia coli*." Proceedings of the National Academy of Sciences of the United States of America **92** (6): 1941-1944.

Gorchein, A., Gibson, L. C. D. and Hunter, C. N. (1993). "Gene expression and control of enzymes for synthesis of magnesium protoporphyrin monomethyl ester in Rhodobacter sphaeroides." Biochemical Society Transactions **21** (2): 201S.

Goslings, D., Meskauskiene, R., Kim, C. H., Lee, K. P., Nater, M. and Apel, K. (2004). "Concurrent interactions of heme and FLU with Glu tRNA reductase (HEMA1), the target of metabolic feedback inhibition of tetrapyrrole biosynthesis, in dark- and light-grown Arabidopsis plants." Plant Journal **40** (6): 957-967.

Gough, S. P., Petersen, B. O. and Duus, J. O. (2000). "Anaerobic chlorophyll isocyclic ring formation in *Rhodobacter capsulatus* requires a cobalamin cofactor." Proceedings of the National Academy of Sciences of the United States of America **97** (12): 6908-6913.

Granick, S. (1954). "Enzymatic Conversion of δ-Aminolevulinic Acid to Porphobilinogen." Science **120** (3130): 1105-1106.

Griffiths, W. T. (1978). "Reconstitution of Chlorophyllide Formation by Isolated Etioplast Membranes." Biochemical Journal **174** (3): 681-692.

Griffiths, W. T., McHugh, T. and Blankenship, R. E. (1996). "The light intensity dependence of protochlorophyllide photoconversion and its significance to the catalytic mechanism of protochlorophyllide reductase." Febs Letters **398** (2-3): 235-238.

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993a). "^{13}C Line Narrowing by ^{2}H Decoupling in ^{2}H/^{13}C/^{15}N Enriched Proteins. Application to Tripple Resonance 4D *J* Connectivity of Sequential Amides." Journal of the American Chemical Society **115** (10): 4369-4370.

Grzesiek, S. and Bax, A. (1993b). "Amino acid type determination in the sequential assignment of uniformly ^{13}C/^{15}N-enriched proteins." Journal of Biomolecular Nmr **3** (2): 185-204.

Grzybowska, E., Gora, M., Plochocka, D. and Rytka, J. (2002). "*Saccharomyces cerevisiae* Ferrochelatase Forms a Homodimer." Archives of Biochemistry and Biophysics **398** (2): 170-178.

Han, B., Liu, Y., Ginzinger, S. and Wishart, D. (2011). "SHIFTX2: significantly improved protein chemical shift prediction." Journal of Biomolecular Nmr **50** (1): 43-57.

Hansson, M. and Hederstedt, L. (1994). "Purification and characterization of a water-soluble ferrochelatase from *Bacillus subtilis*." European Journal of Biochemistry **220** (1): 201-208.

Helfrich, M., Ross, A., King, G. C., Turner, A. G. and Larkum, A. W. D. (1999). "Identification of [8-vinyl] -protochlorophyllide *a* in phototrophic prokaryotes and algae: chemical and spectroscopic properties[1]." Biochimica Et Biophysica Acta-Bioenergetics **1410** (3): 262-272.

Heyes, D. J., Heathcote, P., Rigby, S. E. J., Palacios, M. A., van Grondelle, R. and Hunter, C. N. (2006a). "The First Catalytic Step of the Light-driven Enzyme Protochlorophyllide Oxidoreductase Proceeds via a Charge Transfer Complex*." Journal of Biological Chemistry **281** (37): 26847-26853.

Heyes, D. J. and Hunter, C. N. (2002a). "Site-directed mutagenesis of Tyr-189 and Lys-193 in NADPH:protochlorophyllide oxidoreductase from *Synechocystis*." Biochemical Society Transactions **30**: 601-604.

Heyes, D. J. and Hunter, C. N. (2004). "Identification and Characterization of the Product Release Steps within the Catalytic Cycle of Protochlorophyllide Oxidoreductase." Biochemistry **43** (25): 8265-8271.

Heyes, D. J. and Hunter, C. N. (2005). "Making light work of enzyme catalysis: protochlorophyllide oxidoreductase." Trends in Biochemical Sciences **30** (11): 642-649.

Heyes, D. J. and Hunter, C. N. (2009a). Biosynthesis of Chlorophyll and Bacteriochlorophyll. In. Warren, M. J. and Smith, A. G. Tetrapyrroles: Birth, Life and Death. New York, Landes Bioscience/Springer: 235-249.

Heyes, D. J., Hunter, C. N., van Stokkum, I. H. M., van Grondelle, R. and Groot, M. L. (2003a). "Ultrafast enzymatic reaction dynamics in protochlorophyllide oxidoreductase." Nature Structural Biology **10** (6): 491-492.

Heyes, D. J., Kruk, J. and Hunter, C. N. (2006b). "Spectroscopic and kinetic characterization of the light-dependent enzyme protochlorophyllide oxidoreductase (POR) using monovinyl and divinyl substrates." Biochemical Journal **394**: 243-248.

Heyes, D. J., Martin, G. E. M., Reid, R. J., Hunter, C. N. and Wilks, H. M. (2000). "NADPH:protochlorophyllide oxidoreductase from *Synechocystis*: overexpression, purification and preliminary characterisation." Febs Letters **483** (1): 47-51.

Heyes, D. J., Ruban, A. V. and Hunter, C. N. (2003b). "Protochlorophyllide Oxidoreductase: "Dark" Reactions of a Light-Driven Enzyme." Biochemistry **42** (2): 523-528.

Heyes, D. J., Ruban, A. V., Wilks, H. M. and Hunter, C. N. (2002b). "Enzymology below 200 K: The kinetics and thermodynamics of the photochemistry catalyzed by protochlorophyllide oxidoreductase." Proceedings of the National Academy of Sciences of the United States of America **99** (17): 11145-11150.

Heyes, D. J., Sakuma, M. and Scrutton, N. S. (2007). "Laser Excitation Studies of the Product Release Steps in the Catalytic Cycle of the Light-driven Enzyme, Protochlorophyllide Oxidoreductase." Journal of Biological Chemistry **282** (44): 32015-32020.

Heyes, D. J. and Scrutton, N. S. (2009b). "Conformational changes in the catalytic cycle of protochlorophyllide oxidoreductase: what lessons can be learnt from dihydrofolate reductase?" Biochemical Society Transactions **37**: 354-357.

Hinchigeri, S. B., Hundle, B. and Richards, W. R. (1997). "Demonstration that the BchH protein of *Rhodobacter capsulatus* activates *S*-adenosyl-L-methionine:magnesium protoporphyrin IX methyltransferase." Febs Letters **407** (3): 337-342.

Holtorf, H. and Apel, K. (1996a). "The regulation of NADPH-protochlorophyllide oxidoreductases A and B in green barley plants kept under a diural light/dark cycle." Planta **199** (2): 289-295.

Holtorf, H. and Apel, K. (1996b). "Transcripts of the two NADPH protochlorophyllide oxidereductase genes *PorA* and *PorB* are differentially degraded in etiolated barley seedlings." Plant Molecular Biology **31** (2): 387-392.

Holtorf, H., Reinbothe, S., Reinbothe, C., Bereza, B. and Apel, K. (1995). "Two routes for chlorophyllide synthesis that are differentially regulated by light in barley (*Hordeum vulgare* L.)." Proceedings of the National Academy of Sciences of the United States of America **92** (8): 3254-3258.

Ikura, M., Kay, L. E. and Bax, A. (1990). "A Novel Approach for Sequential Assignment of ^{1}H, ^{13}C, and ^{15}N Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimentional NMR Spectroscopy. Application to Calmodulin." Biochemistry **29** (19): 4659-4667.

Ilag, L. L., Kumar, A. M. and Soll, D. (1994). "Light Regulation of Chlorophyll Biosynthesis at the Level of 5-Aminolevulinate Formation in Arabidopsis." Plant Cell **6** (2): 265-275.

Ishijima, S., Uchlbori, A., Takagi, H., Maki, R. and Ohnishi, M. (2003). "Light-induced increase in free Mg^{2+} concentration in spinach chloroplasts: Measurement of free Mg^{2+} by using a fluorescent probe and necessity of stromal alkalinization." Archives of Biochemistry and Biophysics **412** (1): 126-132.

Jacobs, J. M. and Jacobs, N. J. (1987). "Oxidation of protoporphyrinogen to protoporphyrin, a step in chlorophyll and heme-biosynthesis Purification and partial characterization of the enzyme from barley organelles." Biochemical Journal **244** (1): 219-224.

Jacobs, N. J. and Jacobs, J. M. (1975). "Fumarate as alternative electron acceptor for late steps of anaerobic heme synthesis in *Escherichia coli*." Biochemical and Biophysical Research Communications **65** (1): 435-441.

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

Jaffe, E. K. (1995a). "Porphobilinogen Synthase, The First Source of Heme's Asymmetry." Journal of Bioenergetics and Biomembranes **27** (2): 169-179.

Jaffe, E. K. (1995b). "Porphobilinogen synthase, the first source of heme symmetry." Journal of Bioenergetics and Biomembranes **27** (2): 169-179.

Jaffe, E. K., Abrams, W. R., Kaempfen, H. X. and Harris, K. A. (1992). "5-Chloroevulinate Modification of Porphobilinogen Synthase Identifies a Potential Role for the Catalytic Zinc." Biochemistry **31** (7): 2113-2123.

Jaffe, E. K., Volin, M., Myers, C. B. and Abrams, W. R. (1994). "5-Chloro [1,4-^{13}C]levulinic Acid Modification of Mammalian and Bacterial Porphobilinogen Synthase Suggests an Active Site Containing Two Zn(II)." Biochemistry **33** (38): 11554-11562.

Jensen, P. E., Gibson, L. C. D., Henningsen, K. W. and 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." Journal of Biological Chemistry **271** (28): 16662-16667.

Jensen, P. E., Gibson, L. C. D. and 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* PCC6803." Biochemical Journal **334**: 335-344.

Jensen, P. E., Gibson, L. C. D. and Hunter, C. N. (1999). "ATPase activity associated with the magnesium-protoporphyrin IX chelatase enzyme of *Synechocystis* PCC6803: evidence for ATP hydrolysis during Mg^{2+} insertion, and the MgATP-dependent interaction of the ChlI and ChlD subunits." Biochemical Journal **339**: 127-134.

Jensen, P. E., Reid, J. D. and Hunter, C. N. (2000). "Modification of cysteine residues in the ChlI and ChlH subunits of magnesium chelatase results in enzyme inactivation." Biochemical Journal **352**: 435-441.

Jones, R. M. and Jordan, P. M. (1993). "Purification and properties of the uroporphyrinogen decarboxylase from *Rhodobacter sphaeroides*." Biochemical Journal **293**: 703-712.

Jordan, P. M. and Berry, A. (1981). "Mechanism of action of porphobilinogen deaminase The participation of stable enzyme substrate covalent intermediates between porphobilinogen and the porphobilinogen deaminase from *Rhodopseudomonas spheroides*." Biochemical Journal **195** (1): 177-181.

Jordan, P. M., Burton, G., Nordlov, H., Schneider, M. M., Pryde, L. and Scott, A. I. (1979a). "Pre-uroporphyrinogen: a Substrate for Uroporphyrinogen III Cosynthetase." Journal of the Chemical Society-Chemical Communications (5): 204-205.

Jordan, P. M. and Gibbs, P. N. B. (1985). "Mechanism of action of 5-aminolevulinic acid dehydratase from human erythrocytes." Biochemical Journal **227** (3): 1015-1020.

Jordan, P. M. and Seehra, J. S. (1979b). "Biosynthesis of uroporphyrinogen III: Order of assembly of the four porphobilinogen molecules in the formation of the tetrapyrrole ring." Febs Letters **104** (2): 364-366.

Jordan, P. M. and Warren, M. J. (1987). "Evidence for a dipyrromethane cofactor at the catalytic site of *Escherichia coli* porphobilinogen deaminase." Febs Letters **225** (1-2): 87-92.

Kagan, R. M. and Clarke, S. (1994). "Widespread Occurance of Three Sequence Motifs in Diverse S-Adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for These Enzymes." Archives of Biochemistry and Biophysics **310** (2): 417-427.

Kannangara, C. G., Gough, S. P., Bruyant, P., Hoober, J. K., Kahn, A. and Vonwettstein, D. (1988). "Transfer RNA-glu as a cofactor in δ-aminolaevilinic biosynthesis-Steps that regulate chlorophyll synthesis." Trends in Biochemical Sciences **13** (4): 139-143.

Kelley, L. A. and Sternberg, M. J. E. (2009). "Protein structure prediction on the Web: a case study using the Phyre server." Nature Protocols **4** (3): 363-371.

Kestin, J. and Whitelaw, J. H. (1966). Transport Properties of Water Substance. Sixth International Conference on the Properties of Steam, Trans ASME - Journal of Engineering for Power.

Kikuchi, G., Shemin, D. and Bachmann, B. J. (1958). "The enzymatic synthesis of δ-aminolaevulinic acid." Biochimica Et Biophysica Acta **28** (1): 219-220.

Klement, H., Helfrich, M., Oster, U., Schoch, S. and Rudiger, W. (1999). "Pigment-free NADPH:protochlorophyllide oxidoreductase from *Avena sativa* L Purification and substrate specificity." European Journal of Biochemistry **265** (3): 862-874.

Koch, M., Breithaupt, C., Kiefersauer, R., Freigang, J., Huber, R. and Messerschmidt, A. (2004). "Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis." Embo Journal **23** (8): 1720-1728.

Kohen, A., Cannio, R., Bartolucci, S. and Klinman, J. P. (1999). "Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase." Nature **399** (6735): 496-499.

Kumar, A. M., Csankovszki, G. and Soll, D. (1996). "A second and differentially expressed glutamyl-tRNA reductase gene from *Arabidopsis thaliana*." Plant Molecular Biology **30** (3): 419-426.

Kushlan, D. M. and Lemaster, D. M. (1993). "Resolution and sensitivity enhancement of heteronuclear correlation for methylene resonances via ^{2}H enrichment and decoupling." Journal of Biomolecular Nmr **3** (6): 701-708.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during Assembly of Head of Bacteriophage T4." Nature **227** (5259): 680-&.

Lake, V., Olsson, U., Willows, R. D. and Hansson, M. (2004). "ATPase activity of magnesium chelatase subunit I is required to maintain subunit D *in vivo*." European Journal of Biochemistry **271** (11): 2182-2188.

Lash, T. D. (1991). "Action of uroporphyrinogen decarboxylase on uroporphyrinogen III: a reassessment of the clockwise decarboxylation hypothesis." Biochemical Journal **278**: 901-903.

Layer, G., Heinz, D. W., Jahn, D. and Schubert, W.-D. (2004). "Structure and function of radical SAM enzymes." Current Opinion in Chemical Biology **8** (5): 468-476.

Layer, G., Moser, J., Heinz, D. W., Jahn, D. and Schubert, W. D. (2003). "Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes." Embo Journal **22** (23): 6214-6224.

Layer, G., Reichelt, J., Jahn, D. and Heinz, D. W. (2010). "Structure and function of enzymes in heme biosynthesis." Protein Science **19** (6): 1137-1161.

Lebedev, N., Karginova, O., McIvor, W. and 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 **40** (42): 12562-12574.

Lebedev, N. and Timko, M. P. (1998). "Protochlorophyllide photoreduction." Photosynthesis Research **58** (1): 5-23.

Lebedev, N. and Timko, M. P. (1999). "Protochlorophyllide oxidoreductase B-catalyzed protochlorophyllide photoreduction *in vitro*: Insight into the mechanism of chlorophyll formation in light-adapted plants." Proceedings of the National Academy of Sciences of the United States of America **96** (17): 9954-9959.

Lebedev, N., vanCleve, B., Armstrong, G. and 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** (12): 2081-2090.

Lee, D. S., Flachsova, E., Bodnarova, M., Demeler, B., Martasek, P. and Raman, C. S. (2005). "Structural basis of hereditary coproporphyria." Proceedings of the National Academy of Sciences of the United States of America **102** (40): 14232-14237.

Leustek, T., Smith, M., Murillo, M., Singh, D. P., Smith, A. G., Woodcock, S. C., Awan, S. J. and Warren, M. J. (1997). "Siroheme Biosynthesis in Higher Plants - Analysis of an *S*-Adenosyl-L-Methionine-Dependent Uroporphyrinogen III Methyltransferase from *Arabidopsis thaliana*." Journal of Biological Chemistry **272** (5): 2744-2752.

Lo, M. C., Aulabaugh, A., Jin, G. X., Cowling, R., Bard, J., Malamas, M. and Ellestad, G. (2004). "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery." Analytical Biochemistry **332** (1): 153-159.

Lockhart, P. J., Larkum, A. W. D., Steel, M. A., Waddell, P. J. and Penny, D. (1996). "Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis." Proceedings of the National Academy of Sciences of the United States of America **93** (5): 1930-1934.

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. and ShoolinginJordan, P. M. (1996). "The Three-Dimensional Structure of *Escherichia coli* Porphobilinogen Deaminase at 1.76-Å Resolution." Proteins-Structure Function and Genetics **25** (1): 48-78.

Louie, G. V., Brownlie, P. D., Lambert, R., Cooper, J. B., Blundell, T. L., Wood, S. P., Warren, M. J., Woodcock, S. C. and Jordan, P. M. (1992). "Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site." Nature **359** (6390): 33-39.

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

Martins, B. M., Grimm, B., Mock, H. P., Huber, R. and Messerschmidt, A. (2001). "Crystal Structure and Substrate Binding Modeling of the Uroporphyrinogen-III Decarboxylase from *Nicotiana tabacum* Implications for the Catalytic Mechanism." Journal of Biological Chemistry **276**: 44108-44116.

Masuda, S., Ikeda, R., Masuda, T., Hashimoto, H., Tsuchiya, T., Kojima, H., Nomata, J., Fujita, Y., Mimuro, M., Ohta, H. and Takamiya, K.-i. (2009). "Prolamellar bodies formed by cyanobacterial protochlorophyllide oxidoreductase in Arabidopsis." The Plant Journal **58** (6): 952-960.

Masuda, T. (2008). "Recent overview of the Mg branch of the tetrapyrrole biosynthesis leading to chlorophylls." Photosynthesis Research **96** (2): 121-143.

Masuda, T., Fusada, N., Oosawa, N., Takamatsu, K., Yamamoto, Y. Y., Ohto, M., Nakamura, K., Goto, K., Shibata, D., Shirano, Y., Hayashi, H., Kato, T., Tabata, S., Shimada, H. and Takamiya, K. (2003). "Functional Analysis of Isoforms of NADPH:Protochlorophyllide Oxidoreductase (POR), PORB and PORC, in *Arabidopsis thaliana*." Plant and Cell Physiology **44** (10): 963-974.

Masuda, T. and Takamiya, K. (2004). "Novel insights into the enzymology, regulation and physiological functions of light-dependent protochlorophyllide oxidoreductase in angiosperms." Photosynthesis Research **81** (1): 1-29.

Mathews, M. A. A., Schubert, H. L., Whitby, F. G., Alexander, K. J., Schadick, K., Bergonia, H. A., Phillips, J. D. and Hill, C. P. (2001). "Crystal structure of human uroporphyrinogen III synthase." Embo Journal **20** (21): 5832-5839.

McEwen, B., Seyyedi, M., Younis, S. and Sundqvist, C. (1996). "Formation of short-wavelength chlorophyll(ide) after brief irradiation is correlated with the occurrence of protochlorophyll(ide)$_{636-642}$ in dark-grown epi- and hypocotyls of bean (*Phaseolus vulgaris*)." Physiologia Plantarum **96** (1): 51-58.

McFarlane, M. J., Hunter, C. N. and Heyes, D. J. (2005). "Kinetic characterisation of the light-driven protochlorophyllide oxidoreductase (POR) from *Thermosynechococcus elongatus*." Photochemical & Photobiological Sciences **4** (12): 1055-1059.

Medlock, A., Swartz, L., Dailey, T. A., Dailey, H. A. and Lanzilotta, W. N. (2007). "Substrate interactions with human ferrochelatase." Proceedings of the National Academy of Sciences of the United States of America **104** (6): 1789-1793.

Menon, B. R. K., Davison, P. A., Hunter, C. N., Scrutton, N. S. and Heyes, D. J. (2010). "Mutagenesis Alters the Catalytic Mechanism of the Light-driven Enzyme Protochlorophyllide Oxidoreductase." Journal of Biological Chemistry **285** (3): 2113-2119.

Menon, B. R. K., Waltho, J. P., Scrutton, N. S. and Heyes, D. J. (2009). "Cryogenic and Laser Photoexcitation Studies Identify Multiple Roles for Active Site Residues in the Light-

driven Enzyme Protochlorophyllide Oxidoreductase." Journal of Biological Chemistry **284** (27): 18160-18166.

Mosinger, E., Batschauer, A., Schafer, E. and Apel, K. (1985). "Phytochrome control of *in vitro* transcription of specific genes in isolated-nuclei from barley (*Hordeum-vulgare*)." European Journal of Biochemistry **147** (1): 137-142.

Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G. and Fujita, Y. (2010). "X-ray crystal structure of the light-independent protochlorophyllide reductase." Nature **465** (7294): 110-U124.

Nagata, N., Tanaka, R., Satoh, S. and 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 **17** (1): 233-240.

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. and Tabata, S. (2002). "Complete Genome Structure of the Thermophilic Cyanobacterium *Thermosynechococcus elongatus* BP-1." DNA Research **9** (4): 123-130.

Nomata, J., Kitashima, M., Inoue, K. and Fujita, Y. (2006). "Nitrogenase Fe protein-like Fe-S cluster is conserved in L-protein (BchL) of dark-operative protochlorophyllide reductase from *Rhodobacter capsulatus*." Febs Letters **580** (26): 6151-6154.

Nomata, J., Swem, L. R., Bauer, C. E. and Fujita, Y. (2005). "Overexpression and characterization of dark-operative protochlorophyllide reductase from *Rhodobacter capsulatus*." Biochimica Et Biophysica Acta-Bioenergetics **1708** (2): 229-237.

Obeid, J. and White, P. C. (1992). "Tyr-179 and Lys-183 are essential for enzymatic-activity of 11β-hydroxysteroid dehydrogenase." Biochemical and Biophysical Research Communications **188** (1): 222-227.

Oliver, R. P. and Griffiths, W. T. (1980). "Identification of the polypeptides of NADPH-protochlorophyllide oxidoreductase." Biochemical Journal **191** (1): 277-280.

Oliver, R. P. and Griffiths, W. T. (1981). "Covalent labeling of the NADPH:protochlorophyllide oxidoreductase from etioplast membranes with [^{3}H]*N*-phenylmaleimide." Biochemical Journal **195** (1): 93-101.

Oosawa, N., Masuda, T., Awai, K., Fusada, N., Shimada, H., Ohta, H. and Takamiya, K. (2000). "Identification and light-induced expression of a novel gene of NADPH-protochlorophyllide oxidoreductase isoform in *Arabidopsis thaliana*." Febs Letters **474** (2-3): 133-136.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X. Q., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. and Jornvall, H. (2003). "Short-chain dehydrogenases/reductases (SDR): the 2002 update." Chemico-Biological Interactions **143**: 247-253.

Oster, U., Bauer, C. E. and Rudiger, W. (1997). "Characterization of Chlorophyll *a* and Bacteriochlorophyll *a* Synthases by Heterologous Expression in *Escherichia coli*." Journal of Biological Chemistry **272** (15): 9671-9676.

Parham, R. and Rebeiz, C. A. (1995). "Chloroplast Biogenesis 72: A [4-Vinyl] Chlorophyllide *a* Reductase Assay Using Divinyl Chlorophyllide *a* as an Exogeneous Substrate." Analytical Biochemistry **231** (1): 164-169.

Partensky, F., Hess, W. R. and Vaulot, D. (1999). "*Prochlorococcus*, a Marine Photosynthetic Prokaryote of Global Significance." Microbiology and Molecular Biology Reviews **63** (1): 106-127.

Pattanayak, G. K. and Tripathy, B. C. (2002). "Catalytic Function of a Novel Protein Protochlorophyllide Oxidoreductase C of *Arabidopsis thaliana*." Biochemical and Biophysical Research Communications **291** (4): 921-924.

Perham, R. N., Scrutton, N. S. and Berry, A. (1991). "New enzymes for old: Redesigning the coenzyme and substrate specificities of glutathione reductase." Bioessays **13** (10): 515-525.

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997). "Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proceedings of the National Academy of Sciences of the United States of America **94** (23): 12366-12371.

Phillips, J. D., Whitby, F. G., Kushner, J. P. and Hill, C. P. (2003). "Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase." Embo Journal **22** (23): 6225-6233.

Phillips, J. D., Whitby, F. G., Warby, C. A., Labbe, P., Yang, C., Pflugrath, J. W., Ferrara, J. D., Robinson, H., Kushner, J. P. and Hill, C. P. (2004). "Crystal structure of the Oxygen-dependant Coproporphyrinogen Oxidase (Hem13p) of *Saccharomyces cerevisiae*." Journal of Biological Chemistry **279** (37): 38960-38968.

Pinta, V., Picaud, M., Reiss-Husson, F. and 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." Journal of Bacteriology **184** (3): 746-753.

Porra, R. J. and Falk, J. E. (1961). "Protein-bound porphyrins associated with protoporphyrin biosynthesis." Biochemical and Biophysical Research Communications **5** (3): 179-&.

Porra, R. J. and Falk, J. E. (1964). "The Enzymatic Conversion of Coproporphyrinogen III into Protoporphyrin IX." Biochemical Journal **90** (1): 69-&.

Porra, R. J. and Jones, O. T. G. (1963). "Studies on Ferrochelatase 2. An Investigation of Role of Ferrochelatase in Biosynthesis of Various Haem Prosthetic Groups." Biochemical Journal **87** (1): 186-&.

Porra, R. J., Schafer, W., Gadon, N., Katheder, I., Drews, G. and Scheer, H. (1996). "Origin of the two carbonyl oxygens of bacteriochlorophyll *a* - Demonstration of two different pathways for the formation of ring E in *Rhodobacter sphaeroides* and *Roseobacter denitrificans*, and a common hydratase mechanism for 3-acetyl group formation." European Journal of Biochemistry **239** (1): 85-92.

Porra, R. J., Schafer, W., Katheder, I., Scheer, H., Gadon, N. and Drews, G. (1995). "Evidence for two different pathways for the formation of isocyclic ring E of bacteriochlorophyll *a* in *Rhodobacter sphaeroides* and *Roseobacter denitrificans* using ^{18}O-labelling and mass spectrometry." Photosynthesis: From Light to Biosphere, Vol Iii: 881-885.

Porra, R. J., Urzinger, M., Winkler, J., Bubenzer, C. and Scheer, H. (1998). "Biosynthesis of the 3-Acetyl and 13^1-Oxo Groups of Bacteriochlorophyll *a* in the Facultative Aerobic Bacterium, *Rhodovulum sulfidophilum* The presence of both oxygenase and hydratase pathways for isocyclic ring formation." European Journal of Biochemistry **257** (1): 185-191.

Qin, X. H., Sun, L., Wen, X., Yang, X., Tan, Y., Jin, H., Cao, Q. Y., Zhou, W. H., Xi, Z. and Shen, Y. Q. (2010). "Structural insight into unique properties of protoporphyrinogen oxidase from *Bacillus subtilis*." Journal of Structural Biology **170** (1): 76-82.

Rand, K., Noll, C., Schiebel, H. M., Kemken, D., Dulcks, T., Kalesse, M., Heinz, D. W. and Layer, G. (2010). "The oxygen-independent coproporphyrinogen III oxidase HemN utilizes harderoporphyrinogen as a reaction intermediate during conversion of coproporphyrinogen III to protoporphyrinogen IX." Biological Chemistry **391** (1): 55-63.

Raskin, V. I. and Schwartz, A. (2002). "The charge-transfer complex between protochlorophyllide and NADPH: an intermediate in protochlorophyllide photoreduction." Photosynthesis Research **74** (2): 181-186.

Raux-Deery, E., Leech, H. K., Nakrieko, K. A., McLean, K. J., Munro, A. W., Heathcote, P., Rigby, S. E. J., Smith, A. G. and Warren, M. J. (2005). "Identification and Characterization of the Terminal Enzyme of Siroheme Biosynthesis from *Arabidopsis thaliana* - A Plastid-Located Sirohydrochlorin Ferrochelatase Containing a 2Fe-2S Center." Journal of Biological Chemistry **280** (6): 4713-4721.

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

Raux, E., Schubert, H. L. and Warren, M. J. (2000). "Biosynthesis of cobalamin (vitamin B_{12}): a bacterial conundrum." Cellular and Molecular Life Sciences **57** (13-14): 1880-1893.

Reed, M. A. C., Hounslow, A. M., Sze, K. H., Barsukov, I. G., Hosszu, L. L. P., Clarke, A. R., Craven, C. J. and Waltho, J. P. (2003). "Effects of Domain Dissection on the Folding and Stability of the 43 kDa Protein PGK Probed by NMR." Journal of Molecular Biology **330** (5): 1189-1201.

Reid, J. D. and Hunter, C. N. (2002). "Current understanding of the function of magnesium chelatase." Biochemical Society Transactions **30**: 643-645.

Reid, J. D. and Hunter, C. N. (2004). "Magnesium-dependent ATPase Activity and Cooperativity of Magnesium Chelatase from *Synechocystis* sp. PCC6803." Journal of Biological Chemistry **279** (26): 26893-26899.

Reid, J. D., Siebert, C. A., Bullough, P. A. and Hunter, C. N. (2003). "The ATPase Activity of the ChlI Subunit of Magnesium Chelatase and Formation of a Heptameric AAA^+ Ring." Biochemistry **42** (22): 6912-6920.

Reinbothe, C., Buhr, F., Pollmann, S. and Reinbothe, S. (2003a). "*In vitro* Reconstitution of Light-harvesting POR-Protochlorophyllide Complex with Protochlorophyllides *a* and *b*." Journal of Biological Chemistry **278** (2): 807-815.

Reinbothe, C., Lebedev, N., Apel, K. and Reinbothe, S. (1997). "Regulation of chloroplast protein import through a protochlorophyllide-responsive transit peptide." Proceedings of the National Academy of Sciences of the United States of America **94** (16): 8890-8894.

Reinbothe, C., Lebedev, N. and Reinbothe, S. (1999). "A protochlorophyllide light-harvesting complex involved in de-etiolation of higher plants." Nature **397** (6714): 80-84.

Reinbothe, C., Lepinat, A., Deckers, M., Beck, E. and 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." Journal of Biological Chemistry **278** (2): 816-822.

Reinbothe, S., Reinbothe, C., Apel, K. and Lebedev, N. (1996). "Evolution of Chlorophyll Biosynthesis - The Challenge to Survive Photooxidation." Cell **86** (5): 703-705.

Rosé, S., Frydman, R. B., Delossantos, C., Sburlati, A., Valasinas, A. and Frydman, B. (1988). "Spectroscopic Evidence for a Porphobilinogen Deaminase Tetrapyrrole Complex that Is an Intermediate in the Biosynthesis of Uroporphyrinogen III." Biochemistry **27** (13): 4871-4879.

Roy, A., Kucukural, A. and Zhang, Y. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nature Protocols **5** (4): 725-738.

Rüdiger, W. (1987). Chlorophyll synthetase and its implication for regulation of chlorophyll biosynthesis In. Biggins, J. Progress in Photosynthesis Research. Dordrecht, Martinus Nijhoff Publishers: 461-467.

Rüdiger, W. (1997). "Chlorophyll metabolism: From outer space down to the molecular level." Phytochemistry **46** (7): 1151-1167.

Sakakibara, H., Takei, K. and Sugiyama, T. (1996). "Isolation and characterization of a cDNA that encodes maize uroporphyrinogen III methyltransferase, an enzyme involved in the synthesis of siroheme, which is a prosthetic group of nitrite reductase." Plant Journal **10** (5): 883-892.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wuthrich, K. (1998). "TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins." Proceedings of the National Academy of Sciences of the United States of America **95** (23): 13585-13590.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. New York, Cold Spring Harbor.

Sano, S. and Granick, S. (1961). "Mitochondrial Coproporphyrinogen Oxidase and Protoporphyrin Formation." Journal of Biological Chemistry **236** (4): 1173-&.

Santoro, J. and King, G. C. (1992). "A Constant-Time 2D Overbodenhausen Experiment for Inverse Correlation of Isotopically Enriched Species." Journal of Magnetic Resonance **97** (1): 202-207.

Schägger, H. and Vonjagow, G. (1991). "Blue Native Electrophoresis for Isolation of Membrane Protein Complexes in Enzymatically Active Form." Analytical Biochemistry **199** (2): 223-231.

Scheer, H. (1991). Structure and occurrence of chlorophylls. In. Scheer, H. Chlrophylls. Cleveland, Ohio, CRC Press: 3-30.

Scheumann, V., Klement, H., Helfrich, M., Oster, U., Schoch, S. and Rudiger, W. (1999). "Protochlorophyllide b does not occur in barley etioplasts." Febs Letters **445** (2-3): 445-448.

Schmid, R. and Shemin, D. (1955). "The Enzymatic Formation of Porphobilinogen from δ-Aminolevilinic Acid and Its Conversion to Protoporphyrin." Journal of the American Chemical Society **77** (2): 506-507.

Schoch, S. (1978). "Esterification of Chlorophyllide *a* in Greening Bean-Leaves." Zeitschrift Fur Naturforschung C-a Journal of Biosciences **33** (9-10): 712-714.

Schoch, S. and Schafer, W. (1978). "Tetrahydrogeranylgeraniol, A Precursor of Phytol in Biosynthesis of Chlorophyll - Localization of Double Bonds." Zeitschrift Fur Naturforschung C-a Journal of Biosciences **33** (5-6): 408-412.

Schoefs, B. and Franck, F. (2003). "Protochlorophyllide Reduction: Mechanisms and Evolution." Photochemistry and Photobiology **78** (6): 543-557.

Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.3.

Seehra, J. S., Jordan, P. M. and Akhtar, M. (1983). "Anaerobic and aerobic coproporphyrinogen-III oxidases of *Rhodopseudomonas spheroides* - Mechanism and stereochemistry of vinyl group formation." Biochemical Journal **209** (3): 709-718.

Shedbalkar, V. P., Ioannides, I. M. and Rebeiz, C. A. (1991). "Chloroplast biogenesis. Detection of monovinyl protochlorophyll(ide) *b* in plants." Journal of Biological Chemistry **266** (26): 17151-17157.

Shepherd, M., Dailey, T. A. and Dailey, H. A. (2006). "A new class of 2Fe-2S -cluster-containing protoporphyrin (IX) ferrochelatases." Biochemical Journal **397**: 47-52.

Shepherd, M., Reid, J. D. and Hunter, C. N. (2003). "Purification and kinetic characterization of the magnesium protoporphyrin IX methyltransferase from *Synechocystis* PCC6803." Biochemical Journal **371**: 351-360.

Shoolingin-Jordan, P. M., Warren, M. J. and Awan, S. J. (1996). "Discovery that the assembly of the dipyrromethane cofactor of porphobilinogen deaminase holoenzyme proceeds initially by the reaction of preuroporphyrinogen with the apoenzyme." Biochemical Journal **316**: 373-376.

Siepker, L. J., Ford, M., Dekock, R. and Kramer, S. (1987). "Purification of bovine protoporphyrinogen oxidase: Immunological cross-reactivity and structural relationship to ferrochelatase." Biochimica Et Biophysica Acta **913** (3): 349-358.

Sigfridsson, E. and Ryde, U. (2003). "The importance of porphyrin distortions for the ferrochelatase reaction." Journal of Biological Inorganic Chemistry **8** (3): 273-282.

Smith, C. A., Suzuki, J. Y. and Bauer, C. E. (1996). "Cloning and characterization of the chlorophyll biosynthesis gene *chlM* from *Synechocystis* PCC 6803 by complementation of a bacteriochlorophyll biosynthesis mutant of *Rhodobacter capsulatus*." Plant Molecular Biology **30** (6): 1307-1314.

Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. and 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 Research **29** (5): 1097-1106.

Soll, J., Schultz, G., Rudiger, W. and Benz, J. (1983). "Hydrogenation of Geranylgeraniol Two Pathways Exist in Spinach Chloroplasts." Plant Physiology **71** (4): 849-854.

Stark, W. M., Hawker, C. J., Hart, G. J., Philippides, A., Petersen, P. M., Lewis, J. D., Leeper, F. J. and 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." Journal of the Chemical Society-Perkin Transactions 1 (23): 2875-2892.

Steinberg, T. H., Jones, L. J., Haugland, R. P. and Singer, V. L. (1996). "SYPRO Orange and SYPRO Red Protein Gel Stains: One-Step Fluorescent Staining of Denaturing Gels for Detection of Nanogram Levels of Protein." Analytical Biochemistry **239** (2): 223-237.

Su, Q. X., Frick, G., Armstrong, G. and Apel, K. (2001). "POR C of *Arabidopsis thaliana*: a third light- and NADPH-dependent protochlorophyllide oxidoreductase that is differentially regulated by light." Plant Molecular Biology **47** (6): 805-813.

Suzuki, J. Y. and Bauer, C. E. (1995). "Altered Monovinyl and Divinyl Protochlorophyllide Pools in *bchJ* Mutants of *Rhodobacter capsulatus* Possible Monovinyl Substrate Discrimination of Light-Independent Protochlorophyllide Reductase." Journal of Biological Chemistry **270** (8): 3732-3740.

Suzuki, J. Y., Bollivar, D. W. and Bauer, C. E. (1997). "Genetic Analysis of Chlorophyll Biosynthesis." Annual Review of Genetics **31**: 61-89.

Tait, G. H. (1969). "Coproporphyrinogenase Activities in Extracts of *Rhodopseudomonas spheroides*." Biochemical and Biophysical Research Communications **37** (1): 116-&.

Tait, G. H. (1972). "Coproporphyrinogenase Activities in Extracts of *Rhodopseudomonas spheroides* and *Chromatium* Strain D." Biochemical Journal **128** (5): 1159-&.

Tanaka, R. and Tanaka, A. (2007). "Tetrapyrrole Biosynthesis in Higher Plants." Annual Review of Plant Biology **58**: 321-346.

Tanaka, R., Yoshida, K., Nakayashiki, T., Masuda, T., Tsuji, H., Inokuchi, H. and Tanaka, A. (1996). "Differential Expression of Two *hemA* mRNAs Encoding Glutamyl-tRNA Reductase Proteins in Greening Cucumber Seedlings." Plant Physiology **110** (4): 1223-1230.

Tanaka, R., Yoshida, K., Nakayashiki, T., Tsuji, H., Inokuchi, H., Okada, K. and Tanaka, A. (1997). "The third member of the *hemA* gene family encoding glutamyl-tRNA reductase is primarily expressed in roots in *Hordeum vulgare*." Photosynthesis Research **53** (2-3): 161-171.

Teakle, G. R. and Griffiths, W. T. (1993). "Cloning, characterization and import studies on protochlorophyllide reductase from wheat (*Triticum aestivum*)." Biochemical Journal **296**: 225-230.

Teeter, M. M., Yamano, A., Stec, B. and Mohanty, U. (2001). "On the nature of a glassy state of matter in a hydrated protein: Relation to protein function." Proceedings of the National Academy of Sciences of the United States of America **98** (20): 11242-11247.

Townley, H. E., Griffiths, W. T. and Nugent, J. P. (1998). "A reappraisal of the mechanism of the photoenzyme protochlorophyllide reductase based on studies with the heterologously expressed protein." Febs Letters **422** (1): 19-22.

Townley, H. E., Sessions, R. B., Clarke, A. R., Dafforn, T. R. and Griffiths, W. T. (2001). "Protochlorophyllide Oxidoreductase: A Homology Model Examined by Site-Directed Mutagenesis." Proteins-Structure Function and Genetics **44** (3): 329-335.

Tripathy, B. C., Sherameti, I. and Oelmuller, R. (2010). "Siroheme An essential component for life on earth." Plant Signal Behav **5** (1): 14-20.

Valera, V., Fung, M., Wessler, A. N. and Richards, W. R. (1987). "Synthesis of 4R- and 4S- tritium labeled NADPH for the determination of the coenzyme stereospecificity of NADPH:protochlorophyllide oxidoreductase." Biochemical and Biophysical Research Communications **148** (1): 515-520.

Van Niel, C. B. (1962). "The Present Status of Comparative Study of Photosynthesis." Annual Review of Plant Physiology and Plant Molecular Biology **13**: 1-26.

Varughese, K. I., Xuong, N. H., Kiefer, P. M., Matthews, D. A. and Whiteley, J. M. (1994). "Structural and mechanistic characteristics of dihydropteridine reductase: A member of the Tyr-$(Xaa)_3$-Lys-containing family of reductases and dehydrogenases." Proceedings of the National Academy of Sciences of the United States of America **91** (12): 5582-5586.

Verdecia, M. A., Larkin, R. M., Ferrer, J. L., Riek, R., Chory, J. and Noel, J. P. (2005). "Structure of the Mg-Chelatase Cofactor GUN4 Reveals a Novel Hand-Shaped Fold for Porphyrin Binding." Plos Biology **3** (5): 777-789.

Vitkup, D., Ringe, D., Petsko, G. A. and Karplus, M. (2000). "Solvent mobility and the protein 'glass' transition." Nature Structural Biology **7** (1): 34-38.

Walker, C. J. and Griffiths, W. T. (1988). "Protochlorophyllide reductase: a flavoprotein?" Febs Letters **239** (2): 259-262.

Walker, C. J., Mansfield, K. E., Smith, K. M. and Castelfranco, P. A. (1989). "Incorporation of atmospheric oxygen into the carbonyl functionality of the protochlorophyllide isocyclic ring." Biochemical Journal **257** (2): 599-602.

Walker, C. J. and Weinstein, J. D. (1991a). "Furhter Characterization of the Magnesium Chelatase in Isolated Developing Cucumber Chloroplasts - Substrate Specificity, Regulation, Intactness, and ATP Requirements." Plant Physiology **95** (4): 1189-1196.

Walker, C. J. and Weinstein, J. D. (1991b). "*In vitro* assy of the chlorophyll biosynthetic enzyme Mg-chelatase: Resolution of the activity into soluble and membrane-bound fractions." Proceedings of the National Academy of Sciences of the United States of America **88** (13): 5789-5793.

Walker, C. J. and Weinstein, J. D. (1994). "The magnesium-insertion step of chlorophyll biosynthesis is a two-stage reaction." Biochemical Journal **299**: 277-284.

Walker, C. J. and Willows, R. D. (1997). "Mechanism and regulation of Mg-chelatase." Biochemical Journal **327**: 321-333.

Wang, K. F., Dailey, T. A. and Dailey, H. A. (2001). "Expression and characterization of the terminal heme synthetic enzymes from the hyperthermophile Aquifex aeolicus." Fems Microbiology Letters **202** (1): 115-119.

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

Warren, M. J. and Jordan, P. M. (1988). "Further evidence for the involvement of a dipyrromethane cofactor at the active-site of porphobilinogen deaminase." Biochemical Society Transactions **16** (6): 963-965.

Warren, M. J., Roessner, C. A., Santander, P. J. and Scott, A. I. (1990a). "The *Escherichia coli cysG* gene encodes *S*-adenosylmethionine-dependent uroporphyrinogen III methylase." Biochemical Journal **265** (3): 725-729.

Warren, M. J., Stolowich, N. J., Santander, P. J., Roessner, C. A., Sowa, B. A. and Scott, A. I. (1990b). "Enzymatic synthesis of dihydrosirohydrochlorin (precorrin-2) and of a novel pyrrocorphin by uroporphyinogen III methylase." Febs Letters **261** (1): 76-80.

Whitby, F. G., Phillips, J. D., Kushner, J. P. and Hill, C. P. (1998). "Crystal structure of human uroporphyrinogen decarboxylase." Embo Journal **17** (9): 2463-2471.

Wilks, H. M. and 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." Proceedings of the National Academy of Sciences of the United States of America **92** (3): 724-728.

Willows, R. D. (2003). "Biosynthesis of chlorophylls from protoporphyrin IX." Natural Product Reports **20** (3): 327-341.

Willows, R. D. and Beale, S. I. (1998). "Heterologous Expression of the *Rhodobacter capsulatus BchI, -D,* and *-H* Genes That Encode Magnesium Chelatase Subunits and Characterization of the Reconstituted Enzyme." Journal of Biological Chemistry **273** (51): 34206-34213.

Willows, R. D., Gibson, L. C. D., Kanangara, C. G., Hunter, C. N. and von Wettstein, D. (1996). "Three separate proteins constitute the magnesium chelatase of *Rhodobacter sphaeroides.*" European Journal of Biochemistry **235** (1-2): 438-443.

Willows, R. D., Hansson, A., Birch, D., Al-Karadaghi, S. and Hansson, M. (2004). "EM single particle analysis of the ATP-dependent BchI complex of magnesium chelatase: an AAA^+ hexamer." Journal of Structural Biology **146** (1-2): 227-233.

Willows, R. D. and Kriegel, A. M. (2008). Biosynthesis of bacteriochlorophylls in purple bacteria. . In The Purple Phototrophic Bacteria, Dordrecht ; London : Springer: 57-79.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992). "The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy." Biochemistry **31** (6): 1647-1651.

Wu, C. K., Dailey, H. A., Rose, J. P., Burden, A., Sellers, V. M. and Wang, B. C. (2001). "The 2.0 Å structure of human ferrochelatase, the terminal enzyme of heme biosynthesis." Nature Structural Biology **8** (2): 156-160.

Wyckoff, E. E., Phillips, J. D., Sowa, A. M., Franklin, M. R. and Kushner, J. P. (1996). "Mutational analysis of human uroporphyrinogen decarboxylase." Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology **1298** (2): 294-304.

Yamazaki, S., Nomata, J. and Fujita, Y. (2006). "Differential Operation of Dual Protochlorophyllide Reductases for Chlorophyll Biosynthesis in Response to Environmental Oxygen Levels in the Cyanobacterium *Leptolyngbya boryana.*" Plant Physiology **142** (3): 911-922.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. and Kay, L. E. (1994a). "A Suite of Tripple Resonance NMR Experiments for the Backbone Assignment of ^{15}N, ^{13}C,

^{2}H Labeled Proteins with High Sensitivity." Journal of the American Chemical Society **116** (26): 11655-11666.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. and Kay, L. E. (1994b). "An HNCA Pulse Scheme for the Backbone Assignment of ^{15}N, ^{13}C, ^{2}H-Labeled Proteins: Application to a 37-kDa *Trp* Repressor-DNA Complex." Journal of the American Chemical Society **116** (14): 6464-6465.

Yang, Z. M. and Bauer, C. E. (1990). "*Rhodobacter* capsulatus Genes Involved in Early Steps of the Bacteriochlorophyll Biosynthetic Pathway." Journal of Bacteriology **172** (9): 5001-5010.

Younis, S., Ryberg, M. and Sundqvist, C. (1995). "Plastid development in germinating wheat (*Triticum aestivum*) is enhanced by gibberellic acid and delayed by gabaculine." Physiologia Plantarum **95** (3): 336-346.

Zaman, Z. and Akhtar, M. (1976). "Mechanism and Stereochemistry of Vinyl-Group Formation in Haem Biosynthesis." European Journal of Biochemistry **61** (1): 215-223.

Závodszky, P., Kardos, J., Svingor, A. and Petsko, G. A. (1998). "Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins." Proceedings of the National Academy of Sciences of the United States of America **95** (13): 7406-7411.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." Bmc Bioinformatics **9**.

Zhong, L. B., Wiktorsson, B., Ryberg, M. and Sundqvist, C. (1996). "The Shibata shift; effects of *in vitro* conditions on the spectral blue-shift of chlorophyllide in irradiated isolated prolamellar bodies." Journal of Photochemistry and Photobiology B-Biology **36** (3): 263-270.

Zouni, A., Witt, H.-T., Kern, J., Fromme, P., Krauss, N., Saenger, W. and Orth, P. (2001). "Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution." Nature **409** (6821): 739-743.

Zsebo, K. M. and Hearst, J. E. (1984). "Genetic-Physical Mapping of a Photosynthetic Gene Cluster from *R. capsulata*." Cell **37** (3): 937-947.

Appendix A

Media and Buffer Recipes

A.1 Media, Agar, Antibiotics and IPTG

A.1.1 LB Agar

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

10 g	Bacto-tryptone	10 g	NaCl
5 g	Yeast extract	15 g	Bacto-agar

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.

A.1.2 LB Media

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

10 g	Bacto-tryptone	10 g	NaCl
5 g	Yeast extract		

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.

When LB medium was prepared with D_2O empty conical flasks were autoclaved with their bungs in place and left to dry in a 65 °C oven. Following production of the medium in D_2O it was filter sterilised through a 0.2 μm syringe filter and irradiated under UV light for 15 minutes.

A.1.3 M9 Media

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

6 g	Na_2HPO_4	0.5 g	NaCl
3 g	KH_2PO_4		

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]	2 ml	$(^{15}NH_4)_2SO_4$ (0.5 mg.ml^{-1}) [4]
3 g	Glucose [3]	1 ml	1M $MgSO_4$ (autoclaved)
1 ml	Thiamine (1 mg.ml^{-1}) [4]	0.1 ml	1M $CaCl_2$ (autoclaved) [5]

Notes:

[1] Deuterium Oxide (99.9 % purity) was used instead of Milli-Q water when conducting triple labelled growths. After adjusting the pH to 7.4 (uncorrected for D_2O) the medium was filtered through a 0.2 µm syringe filter and irradiated under UV light for 15 minutes.

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

550 mg	$CaCl_2.2H_2O$	45 mg	$CoCl_2.6H_2O$
140 mg	$MnSO_4.H_2O$	26 mg	$Na_2MoO_4.2H_2O$
40 mg	$CuSO_4.5H_2O$	40 mg	H_3BO_4
220 mg	$ZnSO_4.7H_2O$	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 or U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-D_7, 97-98%) was used here if required

[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.

A.1.4 VN Agar

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

10 g	Yeast extract	1 g	K_2HPO_4
5 g	$MgSO_4$	15 g	Bacto-agar

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.

A.1.5 VN Media

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

10 g	Yeast extract	1 g	K_2HPO_4
5 g	$MgSO_4$		

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.

A.2 Antibiotics and IPTG

A.2.1 Neomycin

Neomycin 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 as detailed in 2.3.

A.2.2 Chloramphenicol

Chloramphenicol was dissolved in ethanol to a concentration of 25 $mg.ml^{-1}$. 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 as detailed in 2.3.

A.2.3 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 proof container. When required the stock solution was added to a final concentration as detailed in 2.3.

A.2.4 IPTG (Isopropyl-β-D-galactosidase)

IPTG was dissolved in Milli-Q water to a concentration of 238 $mg.ml^{-1}$ (1M). 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.

A.3 Buffers and Reagents

A.3.1 RF1

For 200 ml, the following compounds were added to 150 ml Milli-Q water:

0.588 g	Potassium acetate	2.0 g	Magnesium chloride
2.42 g	Rubidium chloride	30 ml	Glycerol
0.294 g	Calcium chloride		

The pH was adjusted to 5.8 with dilute acetic acid and made up to 200 ml with Milli-Q water prior to being sterilised by filtration through a 0.2 μm syringe filter. The solution was stored at 4 °C.

A.3.2 RF2

For 100 ml, the following compounds were added to 50 ml Milli-Q water:

0.21 g	MOPS	0.121 g	Rubidium chloride
1.1 g	Calcium chloride	15 ml	Glycerol

The pH was adjusted to 6.5 with dilute NaOH and made up to 100 ml with Milli-Q water prior to being sterilised by filtration through a 0.2 μm syringe filter. The solution was stored at 4 °C.

A.3.3 Binding Buffer

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

6.06 g	TRIS	0.681 g	Imidazole
1.46 g	NaCl		

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.

A.3.4 Wash Buffer

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

6.06 g	TRIS	10.21 g	Imidazole
1.46 g	NaCl		

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.

A.3.5 Elution Buffer

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

6.06 g	TRIS	17.02 g	Imidazole
1.46 g	NaCl		

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.

A.3.6 Uber Elute Buffer

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

6.06 g	TRIS	68.09 g	Imidazole
1.46 g	NaCl		

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.

A.3.7 Low Salt Buffer

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

6.06 g	TRIS	1.46 g	NaCl

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

A.3.8 High Salt Buffer

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

6.06 g	TRIS	116.88 g	NaCl

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.

A.3.9 Gel Filtration Buffer

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

6.06 g	TRIS	1.46 g	NaCl

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 and de-gassing.

A.3.10 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	1.54 g DTT
5.84 g NaCl	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.

A.3.11 HPLC Buffer

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

6.06 g	TRIS	5.84 g	NaCl

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 and de-gassing.

A.3.12 Fluorescence Buffer

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

6.06 g	TRIS	1.48 g	NaCl
1.54g	DTT	Roche Complete Protease Inhibitor	

The pH was adjusted to 7.5 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.

Appendix B

Find_me Source Code

B.1 find_me

```
find_me_intermediate_2 $1 > temp_find_me_2

more temp_find_me_2 | sort -k1g
```

B.2 find_me_intermeditate

```
current $1 > temp_current

previous $1 > temp_previous


nawk '

FILENAME=="temp_current"  {val1[$1]=$0}

FILENAME=="temp_previous" {val2[$1]=$0}

END {

 for (x in val1) {

  if (x in val2) {print x, val1[x]}

  }

 }

' temp_current temp_previous
```

B.3 find_me_intermediate_2

```
find_me_intermediate $1 > temp_find_me

list_spins $1 > temp_assignment


nawk '

FILENAME=="temp_find_me"  {val1[$1]=$0}

FILENAME=="temp_assignment" {val2[$1]=$0}

END {

 for (x in val1) {

  if (x in val2) {print x, val1[x]}

    }

 }

' temp_find_me temp_assignment
```

B.4 find_me_intermediate_3

```
find_me_intermediate $1 > temp_find_me

list_spins $1 > temp_assignment



nawk '

FILENAME=="temp_find_me"  {val1[$1]=$0}

FILENAME=="temp_assignment" {val2[$1]=$0}

END {

 for (x in val1) {

  if (x in val2) {print ""}

  else {print x, val1[x]}

  }

 }

' temp_find_me temp_assignment
```

B.5 find_patt

```
nawk '

FILENAME == "seq_string" && NR==1  {seq=$1}

#FILENAME == "progress_string" {progress=$1}

END {

patt ="^"patt

for (i=1;i<=length(seq);i++) {

 s=substr(seq,i)

 n=match(s,patt)

 if (n!=0) {

#   print i " - " i+RLENGTH-1,substr(s,1,RLENGTH),substr(progress,i,RLENGTH)}

   print i " - " i+RLENGTH-1,substr(s,1,RLENGTH)}

 }

}' patt=$1 seq_string
```

B.6 find_pre_spin_system

```
nawk '(($5>="ca_low" && $5<="ca_high") || ($6>="ca_low" && $6<="ca_up")) &&  (($9>="cb_low" && $9<="cb_up") || ($9=="_"))'
```

B.7 find_spin_system

```
nawk '($4>="ca_low" && $7<="ca_up") && ((($7>="cb_low" && $7<="cb_up") ||  ($7=="_"))  || (( $8>="cb_low" && $8<="cb_up") || ($8=="_")))'
```

B.8 current

```
nawk '
FILENAME=="BACKBONE_ATOMS.txt" {
ss = id - 12
if ($1==ss) {
ca=$3; cb=$4
 if ((cb=="****")) {cb="_"}
 ca_up = ca + 1.5
 ca_low= ca - 1.5
 cb_up = cb + 1.5
 if ((cb=="_")) {cb_up="_"}
 cb_low= cb - 1.5
 if ((cb=="_")) {cb_low="_"}
{
   print ca, ca_up, ca_low, cb, cb_up, cb_low
   }
}
}
FILENAME=="spinsys_list" {
if (((($4>=ca_low && $4<=ca_up) || ($4=="_")) && ((($8>=cb_low && $8<=cb_up) || ($9>=cb_low &&
$9<=cb_up) || ($9=="_") && ($8=="_")))) {print $0}
}

' id=$1      BACKBONE_ATOMS.txt          spinsys_list
```

B.9 previous

```
nawk '

FILENAME=="BACKBONE_ATOMS.txt" {

ss = id - 13

if ($1==ss) {

ca=$3; cb=$4

 if ((cb=="****")) {cb="_"}

 ca_up = ca + 4.0

 ca_low= ca - 4.0

 cb_up = cb + 4.0

 if ((cb=="_")) {cb_up="_"}

 cb_low= cb - 4.0

 if ((cb=="_")) {cb_low="_"}

{

   print ca, ca_up, ca_low, cb, cb_up, cb_low

   }

}

}

FILENAME=="spinsys_list" {

#print "spin_system"

if (((($5>=ca_low && $5<=ca_high) || ($6>=ca_low && $6<=ca_up) || ($7>=ca_low && $7<=ca_up)) &&
(($10>=cb_low && $10<=cb_up) || ($11>=cb_low && $11<=cb_up) || ($12>=cb_low && $12<=cb_up) ||
(($10=="_") && ($11=="_") && ($12=="_")))) {print $0}

}

' id=$1       BACKBONE_ATOMS.txt          spinsys_list
```