UNIVERSITY OF SOUTHAMPTON

STUDIES ON RECOMBINANT *P. SATIVUM* AND HUMAN 5-AMINOLAEVULINIC ACID DEHYDRATASES

by James Humphrey Youell

BIOMOLECULAR SCIENCE GROUP
SCHOOL OF BIOLOGICAL SCIENCES
FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

August 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

BIOMOLECULAR SCIENCE GROUP, SCHOOL OF BIOLOGICAL SCIENCES,
FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

Doctor of Philosophy

STUDIES ON RECOMBINANT *P. SATIVUM* AND HUMAN 5-AMINOLAEVULINIC ACID DEHYDRATASES

by James Humphrey Youell

5-Aminolaevulinic acid dehydratase catalyses the dimerisation of two molecules of 5-aminolaevulinc acid to form the pyrrole, porphobilinogen. This is the first common step in the biosynthesis of haems, chlorophylls, corrins and other tetrapyrroles. A recombinant form of 5-aminolaevulinic acid dehydratase from *P. sativum*, that lacks the *N*-terminal chloroplast transit peptide, has been overexpressed in *Escherichia coli* and isolated by the use of a novel purification technique. This has yielded milligram amounts of enzyme that have enabled the characterisation and crystallisation of the enzyme to be accomplished. Experiments indicate the possibility of alternative quaternary forms of the enzyme that may be modulated by pH, temperature and enzyme concentration. Quantitative investigations into the nature of the chemical dimerisation of the substrate 5-aminolaevulinic acid to the dihydropyrazine have also been investigated under different buffer conditions and the influence on the activity of *P. sativum*, human and *E. coli* 5-aminolaevulinic acid dehydratases have been compared.

Purified *P. sativum* 5-aminolaevulinic acid dehydratase yielded crystals that diffract to 2.9Å and an adapted selenomethionine growth and purification protocol has yielded a selenomethionine form of the enzyme. Despite extensive attempts, using a variety of programmes, a crystal structure has not yet been solved and studies are continuing.

The structure of human recombinant erythrocyte 5-aminolaevulinic acid dehydratase has been solved and compared to that of the native species. The presence of a putative intermediate that is tightly bound at the active site has given further insight into the enzyme mechanism.

Acknowledgements

I would very much like to thank my supervisor, Professor Peter Shoolingin-Jordan, for his patient guidance and encouragement throughout the course of these studies.

I would also like to thank the members of Shoolingin-Jordan, Cooper and Wood groups for their contributions and guidance in respect to the varying aspects of this research. Namely, but in no particular order, Dr. Sarwar for his invaluable help with all things involving molecular biology, Dr. Cheung for the synthesis of so many organic compounds, Dr. Butler for providing crystals of the human erythrocyte ALAD, Dr. Erskine and Dr. Cooper for their expertise in the solving of the recombinant human erythrocyte ALAD structure and for related help with the crystallisation of the *P. sativum* ALAD, Dr. Coker for his expertise in refining the cryofreezing procedure for the *P. sativum* ALAD crystals and Dr. Findlow for help with the NMR studies on the dimerisation of ALA.

Thanks also to Rob, Jed, Katie, Leighton and everyone else who have made the research enjoyable during my time at Southampton.

Finally, special thanks to Caroline and my family who have managed to keep me motivated and enthusiastic through my studies.

Abbreviations

A	Acetate
ALA	5-Aminolaevulinic acid
ALAD	Aminolaevulinic acid dehydratase
ALAS	Aminolaevulinic acid synthase
Å	Ångström (10^{-10}m)
Abs	Absorbance
CCP4	Collaborative Computational Project Number 4
cDNA	Circular deoxyribonucleic acid
Da	Dalton
DEAE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
IPTG	Isopropyl-β-D-thiogalactoside
Km	Dissociation constant
Kb	Kilobase
Mes	2-[*N*-Morpholino]ethanesulphonic acid
Mr	Relative molecular mass
mRNA	Messenger ribonucleic acid
NMR	Nuclear Magnetic Resonance (spectroscopy)
P	Propionate
PAGE	Polyacrylamide gel electrophoresis
PBG	Porphobilinogen
PBGD	Porphobilinogen deaminase
pI	Isoelectric point
SDS	Sodium dodecylsulphate
TEMED	*N,N,N',N'*-Tetramethylenediamine
Tris	Tris (hydroxymethyl)aminomethane
u.v.	Ultraviolet

List of Figures

List of Tables

Chapter One: Introduction

1.1 Tetrapyrroles

Tetrapyrroles are natural, coloured molecules essential for vital processes such as respiration and photosynthesis. Natural tetrapyrroles can either be linear or cyclic; linear tetrapyrroles include the bilins and cyclic tetrapyrroles include haem, chlorophyll and corrins(Leeper 1989).

Tetrapyrroles are only biosynthesised in small quantities, the pathways are well regulated and biosynthetic intermediates never accumulate in large amounts since many of them are photoreactive. In the past, this has led to difficulty in the study of the enzymes in the tetrapyrrole pathway, however, molecular biology techniques and the availability of cloned genes or cDNAs now allows the production of milligram amounts of the enzymes, thus enabling detailed structural investigations to be carried out.

1.2 The production of uroporphyrinogen III

The tetrapyrrole biosynthetic pathway is broadly homologous in every living organism, with 5-aminolaevulinic acid (ALA) being an essential precursor. In three enzyme catalysed stages, eight molecules of ALA form uroporphyrinogen III from which all other tetrapyrroles are derived(Battersby *et al.*, 1980; Jordan, 1994). Figure 1.1 depicts the early reactions in the haem biosynthesis pathway, up to the synthesis of uroporphyrinogen III. This early part of the pathway is similar for many organisms. The uroporphyrinogen III molecule is often referred to as the common precursor molecule for tetrapyrroles, since uroporphyrinogen III can have several possible destinations due to branching of the biosynthetic pathway. For instance, it can be converted into protoporphyrin IX for the production of haem, chlorophylls and bacteriochlorophylls or methylated to embark on the sirohaem, vitamin B_{12} or factor F_{430} pathways. All of these pathways have been extensively studied and detailed reviews have been published(Battersby 1978; Leeper 1989; Raux *et al.*, 1999; Raux *et al.*, 2000; Roessner *et al.*, 2001; Scott, 2001; Banerjee and Ragsdale, 2003; Willows, 2003).

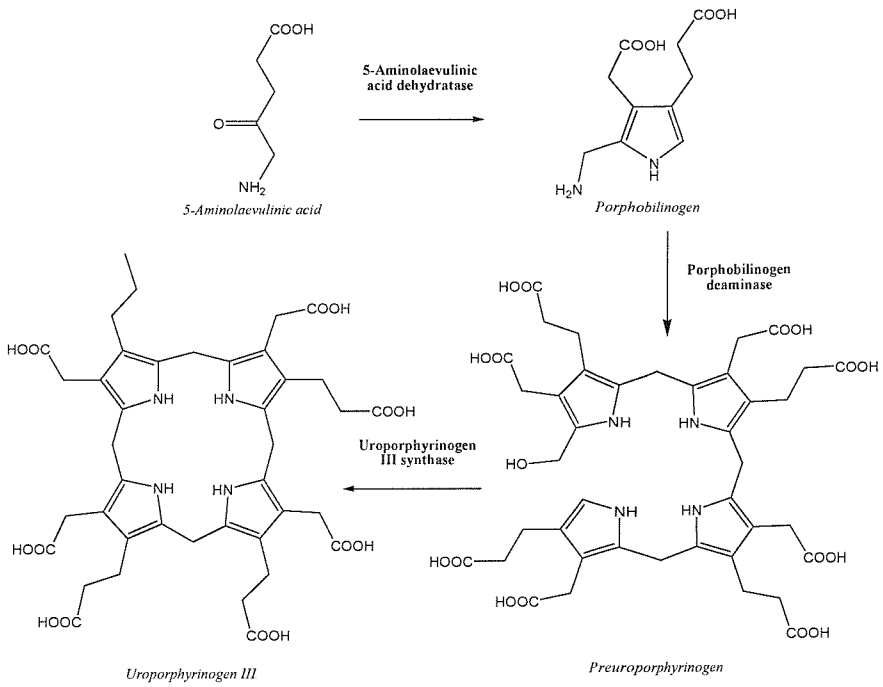
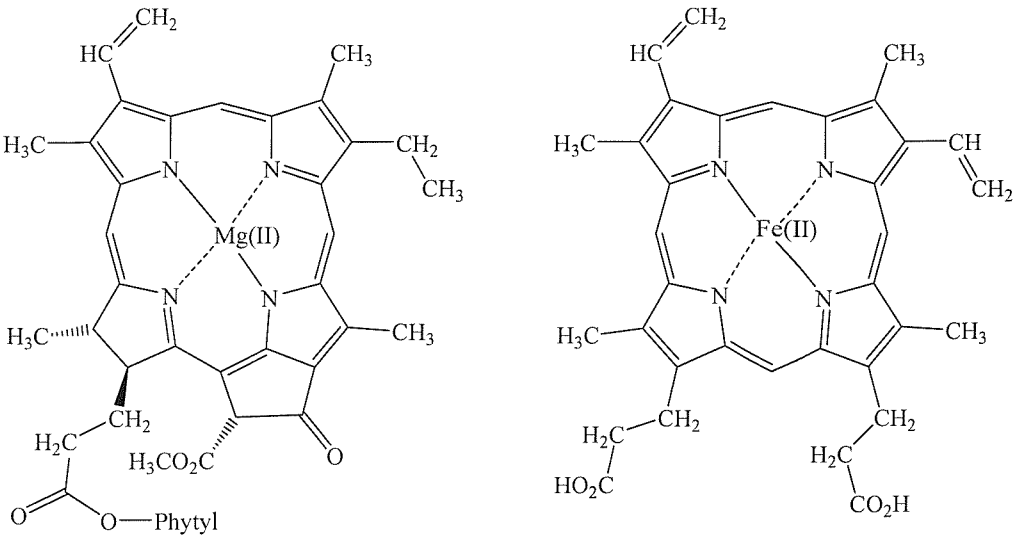


Figure 1.1 The tetrapyrrole pathway: ALA is converted to uroporphyrinogen III via three enzyme catalysed steps(Battersby, *et al.*, 1980; Battersby, 1986; Jordan, 1994).

1.3 Chlorophyll and haem

Two of the most common tetrapyrroles found in nature are chlorophyll and haem. From their structures (Figure 1.2), it is clear that both are derived from the tetrapyrrole intermediate, uroporphyrinogen III. Chlorophyll is essentially similar to haem, however it has an additional fifth isocyclic ring, one double bond has been reduced and it is esterified with a long chain alcohol (either geranol or phytol). Since their conjugated double bonds absorb visible light, haem and chlorophyll containing proteins are coloured (Battersby, *et al.,* 1980; Leeper, 1989; Warren and Scott, 1990; Mauzerall, 1998).

In higher plants, chlorophyll biosynthesis is localised exclusively in the plastids and all, except the last two enzymatic steps required for the synthesis of haem, take place in the chloroplasts (Cornah *et al.*, 2003). In mammals, the first and last three reactions take place in the mitochondria and the other reactions are carried out in the cytosol (Granick and Beale, 1978; Battersby *et al.,* 1980; Jordan, 1994).



Chlorophyll-a *Haem*

Figure 1.2 The structures of the chlorophyll and haem macrocycles. The production of chlorophyll in plants requires fifteen enzyme catalysed steps, whereas the synthesis of haem in mammals requires only ten enzyme catalysed steps(Jordan, 1994).

Chlorophyll is essential to plants in its role as a photoreceptor. Two forms are utilised by plants, chlorophyll *a* and chlorophyll *b*, between which there is only one difference, the formation of a formyl group from the C7 methyl group. This single difference allows the two molecules to absorb light from a broader spectrum of wavelengths.

In the same way, haem is just as essential and is produced in all animals and higher plants where it is a constituent of cytochromes, involved in electron transfer. Haem is also the prosthetic group in catalases, peroxidases and cytochromes P_{450}.

Most of the haem in the human body is derived from endogenous synthesis, with each tissue producing haem according to its own needs. Much of the haem is used for binding and storage of oxygen as the prosthetic group of haemoglobin and myoglobin. The bone marrow accounts for more than 70% of total haem synthesis within the body. Here, haem is incorporated into haemoglobin for erythrocyte precursor cells. Hepatic haem synthesis is the second most important site accounting for approximately 15% of the total haem synthesized. In the liver there is a high requirement for haem to be incorporated into mitochondrial cytochromes as well as into cytochrome P_{450}, catalases and cytochrome b_5 (Granick and Beale, 1978; Battersby *et al.*, 1980; Jordan, 1994).

1.4 ALA synthesis

The first step in the biosynthesis of haem and chlorophyll is the production of 5-aminolaevulinic acid (ALA), which is a ubiquitous reactive α-aminoketone. ALA is formed by one of two routes. In plants and most prokaryotes ALA is produced from glutamate via the C-5 pathway (Figure 1.3). This involves three enzymes: glutamyl-tRNA ligase, glutamyl-tRNA reductase and glutamate 1-semialdehyde aminotransferase. In animals, yeast and photosynthetic bacteria ALA is synthesised in a single reaction from succinyl-CoA and glycine, *via* the Shemin pathway (Figure 1.5), catalysed by 5-aminolaevulinic acid synthase (ALAS)(Jordan, 1994) .

1.4.1 The C-5 Pathway

The C-5 pathway, or glutamate pathway, was discovered in the 1970s and involves the participation of glutamyl-tRNA to convert glutamate to ALA (Figure 1.3). Glutamyl-tRNA ligase catalyses the ligation of the α-carboxyl group of glutamate with $tRNA^{glu}$, in a reaction requiring magnesium ions and ATP. The α-carboxyl-activated glutamate is then reduced to glutamate 1-semialdehyde by the enzyme glutamyl tRNA reductase in a NADPH dependent reaction. Finally, glutamate 1-semialdehyde aminotransferase, which is a pyridoxal 5'-phosphate dependent enzyme, transfers the amino group to produce ALA (Kannangara *et al.*, 1988; Beale and Weinstein, 1991).

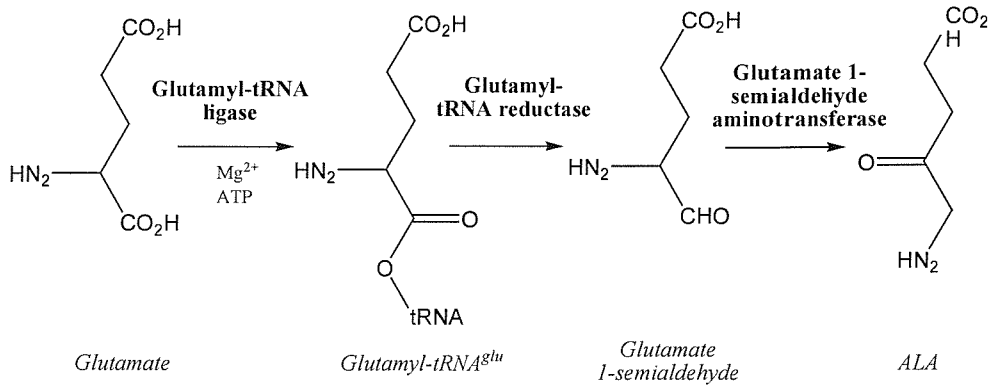


Figure 1.3 The C-5 pathway of ALA synthesis from glutamate, utilising tRNA.

1.4.2 The Shemin pathway

In mammals, ALA is derived from glycine and succinyl-CoA *via* the Shemin pathway (Figure 1.4), which is catalysed by 5-aminolaevulinic acid synthase (ALAS).

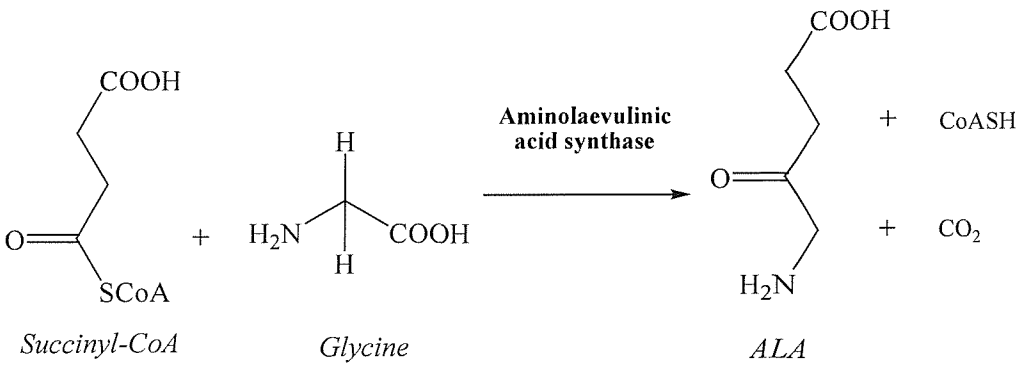


Figure 1.4 The reaction catalysed by 5-aminolaevulinic acid synthase (ALAS). Glycine and succinyl-CoA are condensed, in the presence of ALAS, located in the mitochondria. The Shemin pathway, although occurring in some bacterial species, is mainly restricted to animals and fungi.

ALAS, in common with all other pyridoxal 5'-phosphate dependant enzymes, exists as a homodimer in all species. There are two ALAS genes in mammals, one is expressed in erythrocytes (ALAS2) and the other, a housekeeping enzyme (ALAS1), is expressed in all tissues. Nucleotide sequences are similar for the two isoenzymes although there is no Similarity in the N-terminal region.

ALAS from *Rhodobacter sphaeriodes* has been well studied and is particularly interesting, since the organism can synthesize haem, bacteriochlorophyll and corrin.

ALAS is a pyridoxal 5'-phosphate dependent reaction with the coenzyme bound to the enzyme through a Schiff base with an invariant lysine residue. In murine ALAS this invariant residue has been identified as lysine 313. Mutation of this residue reduces the activity of the enzyme dramatically, although the cofactor can still be non-covalently bound (Dailey, 1990; Jordan, 1991; Chadwick and Ackrill, 1994; Ferreira, 1995).

In eukaryotes the fully functional enzyme is found in mitochondria. ALAS is first synthesised as a cytosolic protein that contains an *N*-terminal signal sequence thought to direct the enzyme to the mitochondria. A chaperone protein maintains the cytosolic ALAS in an unfolded state so that it can be transported through the mitochondrial membrane. Once inside the mitochondrion, the *N*-terminal signal sequence is cleaved by a protease and the final folding is catalysed in an ATP dependent process (May *et al.*, 1986; Volland and

Urban-Grimal, 1988; Gonzalez-Dominguez *et al.*, 2001). There is some evidence that ALA synthase forms a complex with succyinyl-CoA synthase.

1.4.3 Regulation of ALA biosynthesis

In plants, ALA synthesis is the pivotal control point of tetrapyrrole biosynthesis (Beale, 1991). As in other organisms, ALA synthesis appears to be inhibited by haem. This appears to be at the first stage of ALA synthesis from glutamate, at glutamyl-tRNA reductase (GluTR), and has been demonstrated using recombinant GluTR and exogenous haem, where inhibition requires the N-terminal 30 amino acids of the enzyme (Vothknecht *et al.*, 1998). Further research has confirmed the observation that inhibition takes place at GluTR due to the wide range of regulatory signals that have been observed to modulate expression of the *HEMA* gene that codes for the enzyme. *HEMA* genes from a variety of plant species have been shown to be regulated by hormones, the circadian clock, plastid signals and light throught the actions of the phytochrome and cryptochrome families of photoreceptors (Cornah, 2003). It has been proposed that the Mg branch of the tetrapyrrole biosynthetic pathway, specifically Mg-chelatase, may also have a regulatory effect upon GluTR expression and therefore ALA synthesis (Papenbrock *et al.*, 2000). Also, FLU, a nuclear-encoded protein that is localised to chloroplast membranes, has been shown to regulate ALA synthesis independently of haem (Meskauskiene *et al.*, 2001).

In animals, yeast and non-photosynthetic bacteria, ALAS is the rate-limiting step of haem biosynthesis as it determines the flux of substrates shunted though the pathway. Haem is known to regulate ALAS production and activity in a number of ways. The synthesis of ALAS mRNA can be suppressed in the presence of high intracellular concentrations of haem (Granick and Beale, 1978). Messenger RNA encoding ALAS has a short half-life, allowing a quick response to high haem levels. Translation of the ALAS can also be regulated by haem. It has been postulated that haem may react with a specific regulatory protein, which can inhibit translation. Haem is also proposed to have a role in determining the rate of ALAS transport through the mitochondrial membrane (May *et al.*, 1990). Mechanisms of regulation in the erythropoetic system and the liver are different. This regulation is essential not only for controlling the cellular haem levels, but also for preventing the build-up of pathway intermediates, several of which have cytotoxic, photodynamic properties in their oxidised forms.

Translation of the ALAS2 mRNA has been shown to be regulated by binding of certain regulatory proteins. Thus the ALAS2 mRNA contains an iron response element (IRE) that mediates translational control (Cox, 1991). In iron-depleted cells, the high affinity binding of iron-regulatory proteins 1 and 2 (IRP-1 and IRP2) to the IRE represses translation (Ferreira, 1999). Although the exact mechanism by which this occurs is unknown, it has been shown to be induced during erythroid cell differentiation (Fraser, 1987).

Additionally, five DNase hypersensitive sites have been identified in the ALAS2 gene in murine erythroleukemia cells (Schoenhaut, 1989). These sites have been proposed to be specific *cis*-regulatory elements, which are recognised by *trans*-activating transcriptional factors, including, GATA-1 binding factor (GATA-1)(Tsai, 1989), nuclear factor erythroid 2 (NF-E2) (Mignotte, 1989), and erythoid Krüppel-like factor (EKLF)(Miller, 1993).

Certain drugs can induce ALAS and therefore increase the flux of intermediates through the pathway. In patients with porphyria, these drugs can precipitate an acute attack as the defective enzyme interrupts the flux through the pathway and can cause early intermediates such as ALA and PBG to accumulate to dangerous levels. Other porphyrinogenic drugs may initiate attacks by lowering the intracellular haem pool, usually caused by the demand for the formation of cytochrome P_{450} holoenzymes (Tishler, 1999).

1.5 5-Aminolaevulinic acid dehydratase (ALAD)

5-Aminolaevulinic acid dehydratase (ALAD), also known as porphobilinogen synthase (PBGS), is a cytosolic enzyme which catalyses the condensation of two molecules of ALA to form the monopyrrole PBG. This is the first reaction common in the formation of all tetrapyrroles. The substrate, 5-ALA, is intrinsically reactive and can readily dimerise in a non-enzymic reaction forming a dihydropyrazine instead of the desired pyrrole product PBG (Figure 1.5) (Butler, 1992; Novo, 1996).

ALADs are highly conserved, usually octameric enzymes, with eight active sites per octamer. A reduced level of ALAD caused by either genetic factors or inhibition causes an acute porphyria (ALAD deficiency or Doss porphyria)(Sassa, 1998). ALAD was first isolated in 1955 from ox liver and avian erythrocytes. The enzyme has since been purified

from many other sources, such as human erythrocytes (Gibbs *et al.*, 1985), *E. coli* (Senior, *et al.*, 1996) and spinach (Schaumburg *et al.*, 1992) and is thought to be present in virtually all organisms. The availability of cDNA coding for ALAD has allowed recombinant strains of *E. coli* to be engineered to produce large amounts of the enzyme (Liedgens *et al.*, 1980; Spencer and Jordan, 1993; Senior *et al.*, 1996; Frankenberg *et al.*, 1999).

The structural gene for 5-aminolaevulinic acid dehydratase is *hemB* (*Hem2* in yeast). The gene has been cloned and sequenced in *E. coli* (Echelard *et al.*, 1988; Li *et al.*, 1988), yeast (Myers *et al.*, 1987), rat liver (Bishop, 1986), human liver (Wetmur, 1986) and also *P. sativum* (Boese *et al.*, 1991). The sequence of the *E. coli* gene showed two promoter regions, two Shine-Dalgarno sequences, and two start codons. The two start codons would encode two proteins, one with a subunit molecular weight of 35,505 and one with a subunit molecular weight of 36,763. The 5`-end of the *hemB* mRNA was mapped by primer extension analysis and this showed that transcription initiates at the second promoter, located immediately downstream of the first ATG codon. These results were confirmed by *in vitro* DNA-directed translation experiments followed by SDS-PAGE electrophoresis(Echelard *et al.*, 1988). The gene sequences for all bacterial and animal ALADs are very similar except for a few organism-specific codons, reflecting the overall similarity between species (Li *et al.*, 1988).

Similarly the human ALAD is coded for by an open reading frame of 990 bp, which yields a protein of 330 amino acids in length. The gene was originally cloned from a human liver library, and was identified as being identical to the microsequenced ALAD purified from human erythrocytes (Wetmur, 1986).

Plant gene sequences differ from the animal and bacterial ALAD genes due to the additional 5'-sequence that encodes the *N*-terminal plastid targeting peptide. The protein precursor has been shown to be encoded by a long open reading frame specifying 398 amino acids. This additional sequence coding for the *N*-terminal domain of the protein are thought to be required for the import of the protein into the chloroplasts and has also been suggested to be a membrane spanning domain for binding to the thylakoid membrane of the chloroplast (Boese *et al.*, 1991).

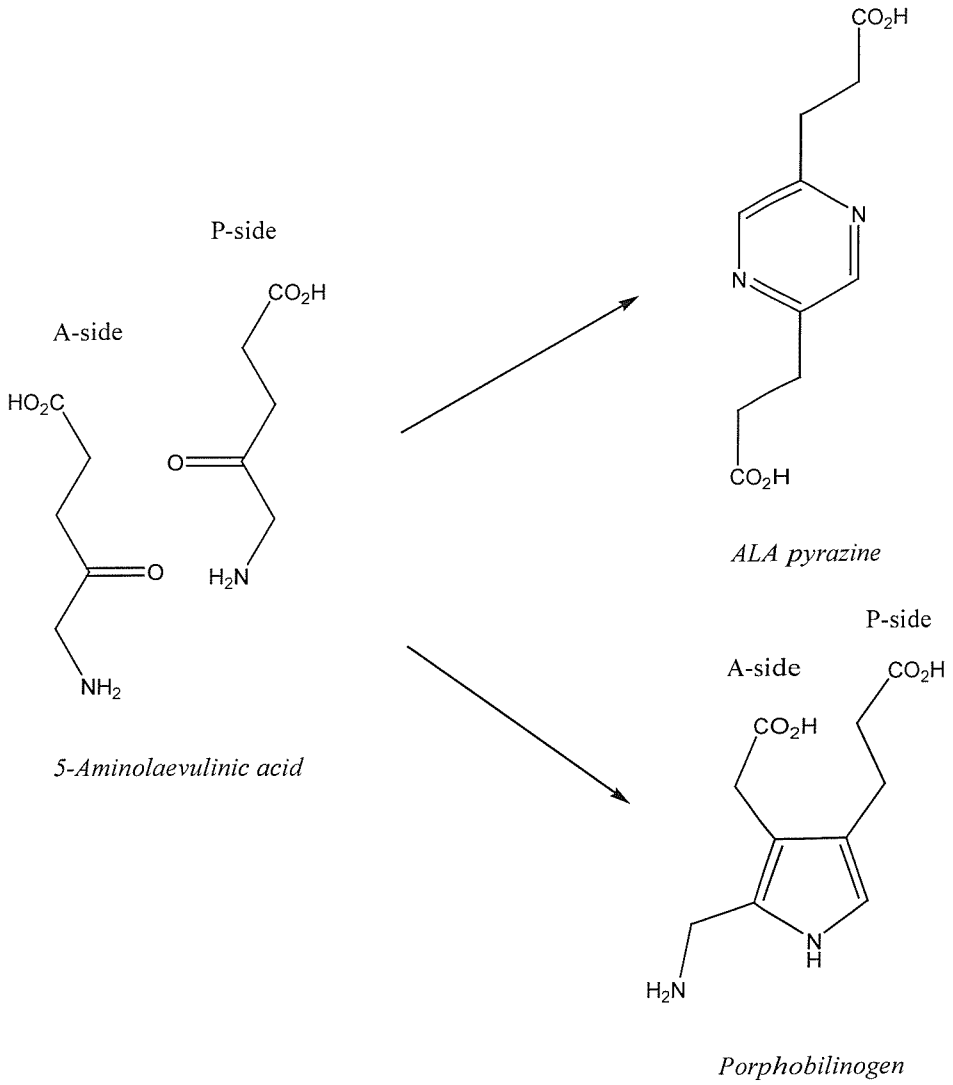


Figure 1.5 The dimerisation of ALA.

ALAD condenses the two molecules of ALA to form one molecule of PBG. The A-side and P-side ALA become the acetate and propionate side chains of PBG respectively. A dihydropyrazine is formed from the non-enzymic reaction between two molecules of ALA and is converted to the pyrazine.

1.6 The 3D structure of ALAD

The first high-resolution structure of a prokaryotic ALAD, yeast ALAD, was published in 1999 (Erskine *et al.*, 1997). The *E. coli* (Erskine *et al.*, 1999), *Pseudomonas aeruginosa* (Frankenberg *et al.*, 1999), and the native human enzymes (both the recombinant human enzyme and the human enzyme purified from human erythrocytes) have since been crystallised and the X-ray structures have been solved.

ALADs crystallise as octamers. It is regarded as a tetramer of dimers as interactions within the dimer are much stronger than those between the four dimers. The octamer has a solvent channel of 15-20 angstroms in diameter passing through the centre.

1.6.1 Structure of the *E. coli* ALAD monomer

Each monomer (Figure 1.6) adopts a TIM barrel or $(\alpha/\beta)_8$ fold, which is formed by an eight-membered cylindrical β-sheet with eight surrounding α helices. This fold shared by many proteins, including a number of aldolases, enzymes of similar in function to ALAD. The loop regions between the α and β segments are elaborated extensively by the insertion of extra secondary structure elements. These loop sections at the *C*-termini of the barrel motifs, form the active site(Erskine *et al.*, 1997; Erskine *et al.*, 1999; Frankenberg *et al.*, 1999; Erskine *et al.*, 2001).

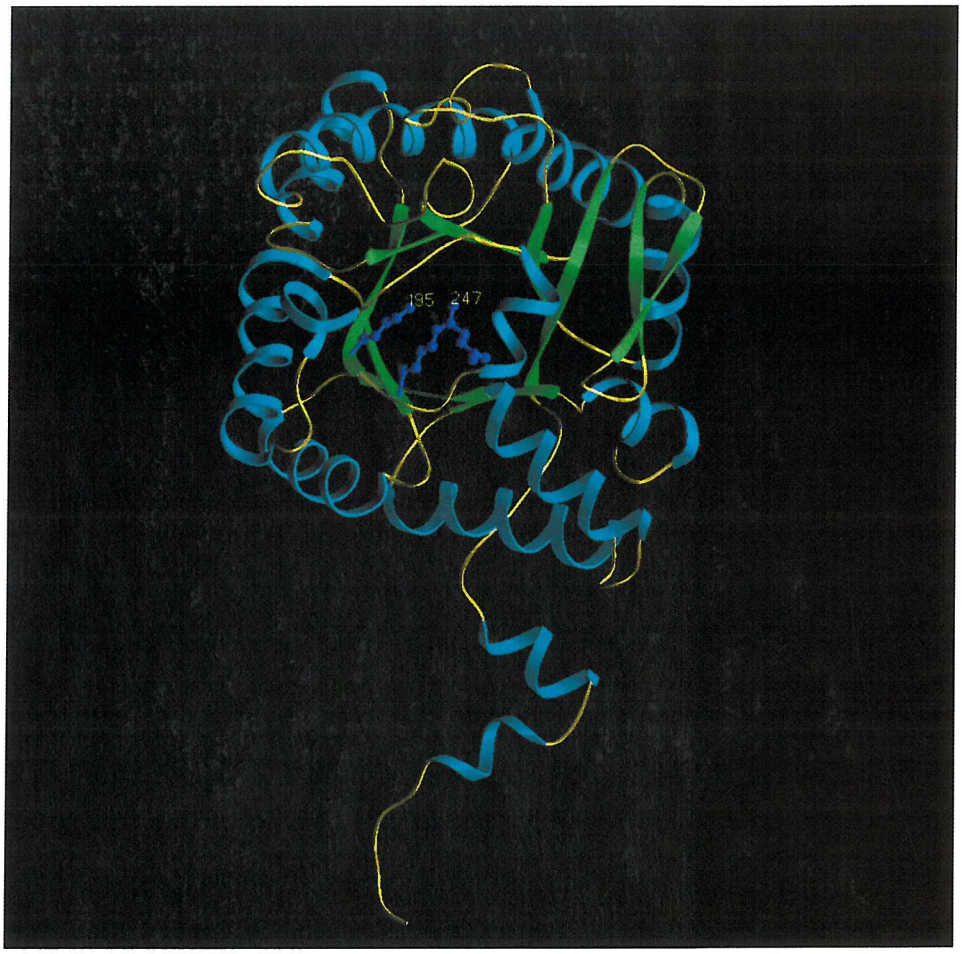


Figure 1.6 Structure of the *E. coli* ALAD monomer.
Catalytic lysines 195 and 247 are shown, with laevulinic acid, bound as a Schiff base, to lysine 247 (Erskine, *et al.*, 1999).

1.6.2 Structure of the *E. coli* ALAD dimer

In the dimer, the *N*-terminal arm from one monomer partly wraps around the barrel domain of the neighbouring monomer, resembling the number 69, with a perpendicular axis. The *N*-terminus of the arm region is partly exposed to the solvent side of the octamer, while the inner surface interacts with the α4 and β5 helices of the second subunit in the dimers (Figure 1.7) (Erskine *et al.*, 1997; Erskine *et al.*, 1999; Frankenberg *et al.*, 1999; Erskine *et al.*, 2001).

Figure 1.7 Structure of the *E. coli* ALAD dimer. The two constitutive monomers orientate so that the *N*-terminal arm region of each monomer wraps around the other in manner resembling the number 69 (Erskine *et al.*, 1999).

1.6.3 Structure of the *E. coli* ALAD octamer

In the ALAD octamer, the dimers give the gross appearance of ellipsoids, with the long axis inclined slightly with respect to the 4-fold symmetry axis. All of the dimers interact identically with each other, mediated by the *N*-terminal arm regions of each dimer, which

associate with one end of the β-barrel in the neighbouring subunit barrel. This end is effectively capped by the *N*-terminal arm region of the neighbouring subunit, leaving the other end of the barrel (effectively the active sites) exposed on the surface of the octamer (Figure 1.8) (Erskine *et al.*, 1997; Erskine *et al.*, 1999; Frankenberg *et al.*, 1999; Erskine *et al.*, 2001).

Figure 1.8 Structure of the *E. coli* ALAD octamer, comprising four dimer components, with the interactions between each dimer mediated between the *N*-terminal extensions of their constitutive monomers.

1.6.4 Structural Similarity between 5-aminolaevulinic acid dehydratases

There is great three dimensional similarity between the ALADs from different species (Figure 1.9), as has been seen from the structures that have been characterised so far and from Similarity mapping of primary amino acid sequences. For example, the *P. sativum* amino acid sequence shares 50% identity (66% similar residues) with the *E. coli* sequence, 42% identity (63% similar residues) with the human sequence, 38% identity (57% similar

residues) with the yeast sequence, and 47% identity (67% similar residues) with the *Pseudomonas aeruginosa* sequence (calculated using the BLAST network service(Altschul *et al.*, 1997)). The most conserved sequences appear within the active site, around which many residues seem invariant, except for the specific metal binding ligands for Mg^{2+} and Zn^{2+} (Erskine *et al.*, 1999).

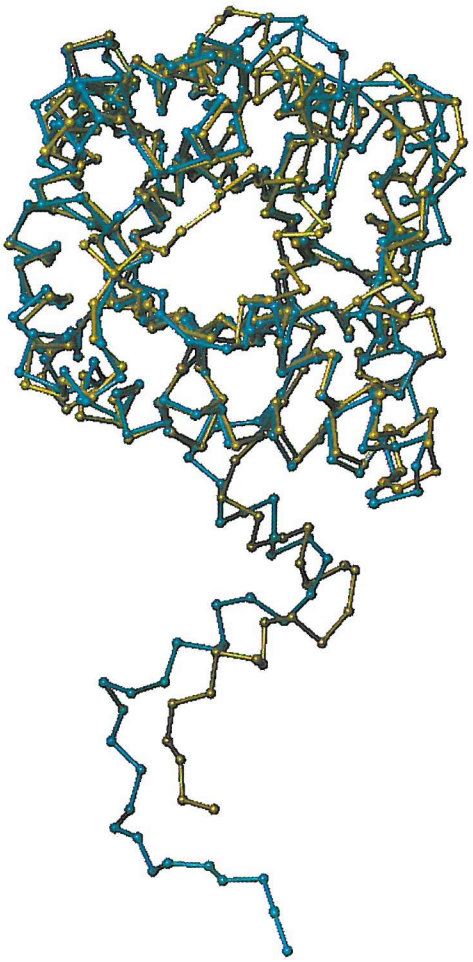


Figure 1.9 Overlay of *E. coli* and yeast ALAD monomers The *E. coli* structure is in yellow and the yeast structure in cyan lines. As can be seen, both monomers are almost superimposable, the major difference being the different length of the *N*-terminal arm region of each monomer (Erskine *et al.*, 1999).

1.6.5 The structure of the *E .coli* active site

Two molecules of ALA are condensed in a Knorr-type reaction by ALA dehydratase, to yield porphobilinogen (PBG) (Jordan, 1994). The enzyme binding sites of the two ALA substrates have been designated the A (giving rise to the acetic side chain of PBG) and the P (giving rise to the propionic side chain of PBG) sites.

Loops at the end of the β-barrel form a pronounced cavity, the base of which is dominated by two spatially adjacent lysine side chains at positions 195 and 247 in *E. coli* ALAD (Lys 210 and Lys 263 in yeast ALAD). Both lysines are located at the *C*-terminal ends of adjacent β-strands in the barrel (β7 and β8). The lysine side chains emerge from a hydrophobic pocket, predominantly filled with tyrosine side chains. However, the active site also contains a number of invariant polar residues which have possible roles in substrate binding and catalysis. Lys 247 forms the Schiff base to the P-side substrate and has a more hydrophobic environment than its neighbouring Lys195. Ser 193 and Tyr 312 have also been observed as being crucial to the binding of the carboxyl group of the substrate within the P site of the active site pocket in *E. coli* (see Figure 1.11) (Erskine *et al.*, 1997; Erskine *et al.*, 1999; Erskine *et al.*, 2001).

Metal binding sites

In *E. coli* ALAD, three cysteines at positions 120,122 and 130 are located in the loop connecting the β5 and α4 motifs with a water molecule coordinated to the catalytic Zn^{2+} with tetrahedral geometry. The catalytic Zn^{2+} is found close to the side chain of invariant Ser165 which, itself, hydrogen bonds to the invariant aspartate, Asp118 (see Figure 1.10). Both residues interact indirectly with Zn^{2+} through a network of water molecules(Erskine *et al.*, 1999). They both, most probably, participate in substrate binding or catalysis. The three cysteine residues (120, 122 and 130) structurally correspond to residues Ala129, Asp131 and Asp139 in *Pseudomonas aeruginosa* ALAD, an Mg^{2+} stimulated enzyme.

In *E. coli* ALAD, a structural Zn^{2+} is located in a pocket, found between the TIM barrel domain and the *N*-terminal arm, adjacent to the side-chain of Lys195. Several water molecules are present in this vicinity. This pocket contains a zinc ion which is co-ordinated octahedrally, by one side-chain oxygen of Glu232 and five solvent molecules. In *Pseudomonas aeruginosa* ALAD the second Mg^{2+} has been found, as described above, again, co-ordinated with five water molecules and Glu245 (Erskine *et al.*, 1997; Erskine *et al.*, 1999; Erskine *et al.*, 2001).

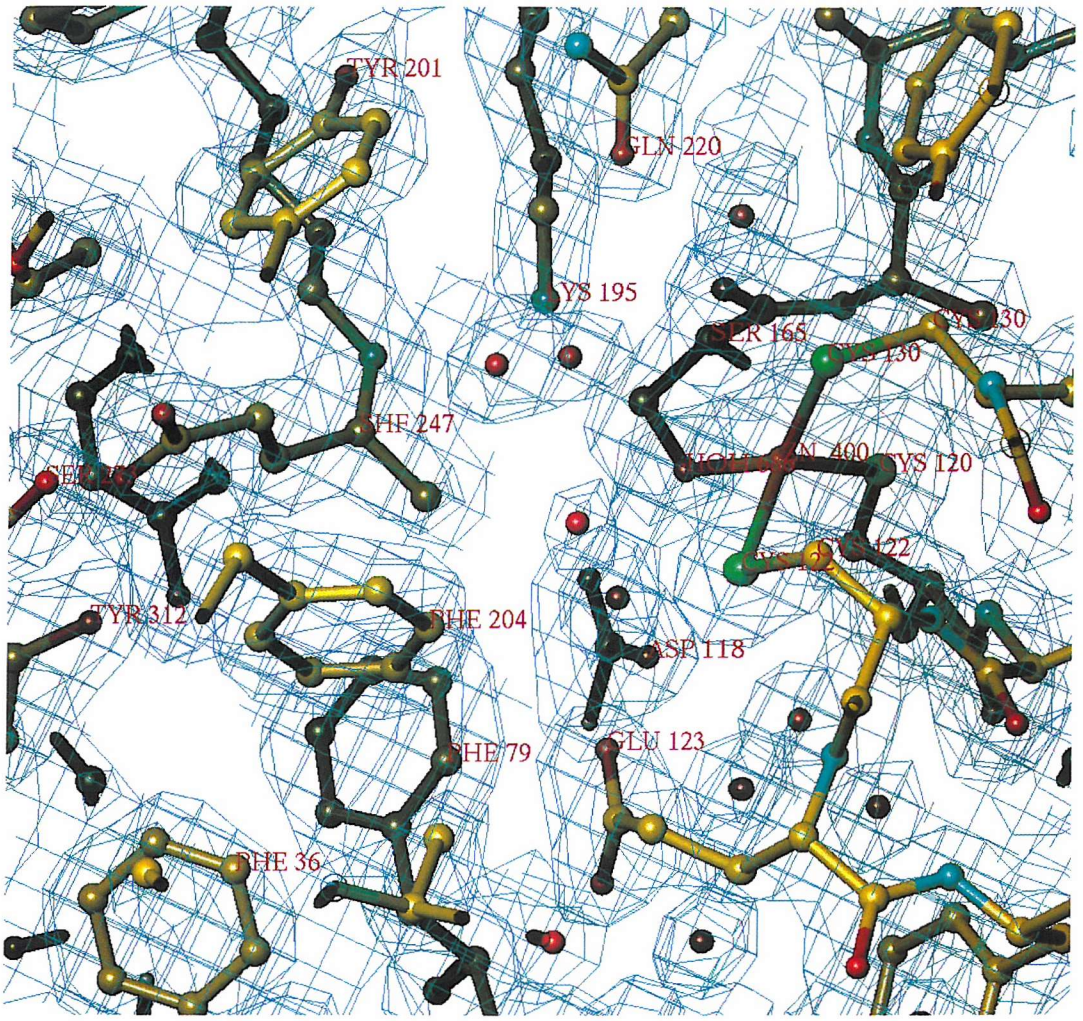
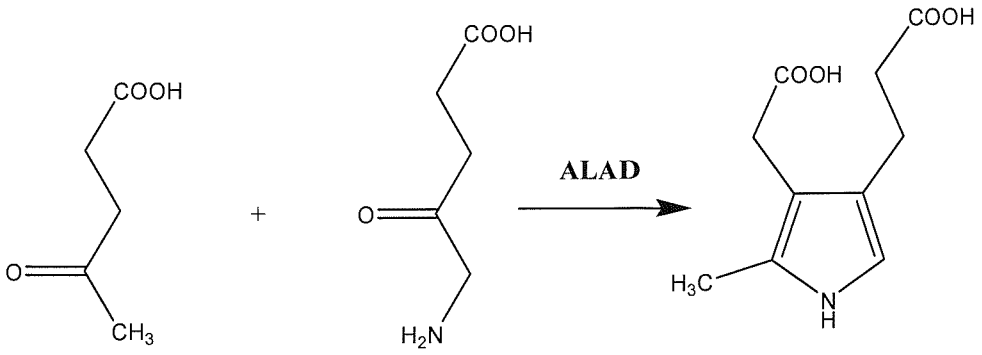


Figure 1.10 Substrate and metal binding sites within the *E. coli* ALAD active site. The active site is shown with laevulinic acid bound as a Schiff base to lysine 247 (SHF 247). Also interactions are visible between the carboxyl group of the laevulinic acid with Ser 217 and Tyr 312. The active site Zn^{2+} (ZN 400) is shown bound tetrahedrally between cysteine residues 120, 122, 130 and a water molecule, depicted as a red sphere (Erskine *et al.*, 1999).

1.7 The catalytic mechanism of ALAD

The first studies to determine the mechanism of ALAD were conducted by Nandi and Shemin (1968). The studies were based on the enzyme from *Rhodobacter sphaeroides*. A mixture of 5-aminolaevulinic acid (ALA) and laevulinic acid (LA) (a substrate analogue) were used in order to differentiate between the two binding sites. The experiment was based on the knowledge that ALAD can catalyse the formation of a mixed pyrrole derived from these two compounds. In this mixed pyrrole, it was observed that the LA component

of the product always formed the acetic acid ('A') side of PBG, whereas the ALA component always formed the propionic acid ('P') side (Nandi and Shemin, 1968). These results, along with observations that the LA also acts as a competitive inhibitor and can form a Schiff base with the enzyme, gave the impression that the substrate binding as the Schiff base gives rise to the 'A' side of PBG (Figure 1.11).



Laevulinic acid (LA) *5-Aminolaevulinic acid (ALA)* *'Mixed Pyrrole'*

Figure 1.11 The formation of the mixed pyrrole from LA and ALA catalysed by ALAD

This was generally accepted until, in 1980, single turnover experiments using isotopically labelled substrate were used to investigate substrate binding in the bovine, bacterial and human forms of ALAD (Jordan and Seehra, 1980; Jordan and Gibbs, 1985). Stoichiometric amounts of labelled substrate (either ^{13}C or ^{14}C) were incubated with the enzyme, so as to occupy preferentially the substrate binding site with the highest affinity. Then, unlabelled substrate was added so as to complete a single turnover of the enzyme to form PBG. The position of the labelled component of the PBG was then determined either by NMR or chemical degradation studies. The results clearly showed that the ALA initially bound to the enzyme formed the 'P' side of the PBG formed, and not the 'A' side as previously suggested (Jordan and Gibbs, 1985).

The original results of Nandi and Shemin could be explained, as the LA was observed to bind preferentially at the 'A' side because it lacks the amino function required for ring closure of the PBG molecule. In effect, it can only act as a competitive inhibitor of the 'P' side and not as a substrate.

In addition to the single turnover experiments, it was observed that the substrate which was bound at the 'P' side through a Schiff base could be trapped using sodium borohydride under similar conditions to those used in the single turnover experiments.

From these results it was concluded that the initial substrate binding occurred at the 'P' side of the active site *via* a Schiff base. The ligand that formed the Schiff base was subsequently identified using radioisotopically labelled ALA that was reduced onto the enzyme with sodium borohydride. Peptide sequencing confirmed it as Lys252 in the human ALAD (Gibbs and Jordan, 1986) (Lys 258 in *E. coli).*

Also required in the mechanism is an active site Zn^{2+}, which facilitates the binding of substrate at the 'A' site (Spencer and Jordan, 1994). However, removal of Zn^{2+} does not prevent the formation of the Schiff base at the 'P' side.

After formation of the Schiff base between the ALAD and ALA, catalysis can progress in one of two ways. Either, a C-C bond could be formed between the C4 of the 'P' side ALA and the C3 of the 'A' side, or, a C=N bond could form between the amino group of the 'P' side ALA and the C4 of the 'A' side substrate molecule (see Figure 1.12)

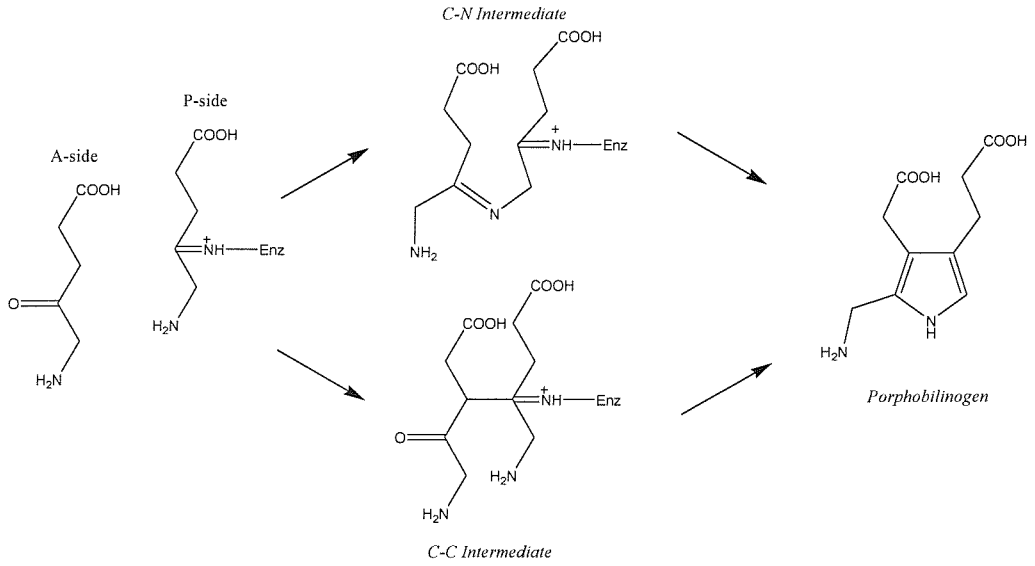


Figure 1.12 Two possible routes for the formation of porphobilinogen from 5-aminolaevulinic acid.

The greatest advancement since this work in the elucidation of the mechanism has been gained through crystallisation of ALAD. Structures from yeast, *E. coli*, *P. aeruginosa* and

human ALADs have all been co-crystallised with LA. All of the structures reveal an unoccupied 'A' site, with LA covalently bound to the 'P' site by a Schiff base between the 4-keto group and an active site lysine (Lys263 in yeast ALAD, Lys247 in *E. coli* ALAD, Lys260 in *P. aeruginosa* ALAD and Lys252 in human ALAD).

Similarly *E. coli* and yeast ALADs have been co-crystallised more recently with the irreversible inhibitor 4,7-dioxosebacic acid (DOSA) (Erskine *et al.*, 2001; Kervinen *et al.*, 2001). DOSA is a mimic of a possible C=N intermediate (Figure 1.13).

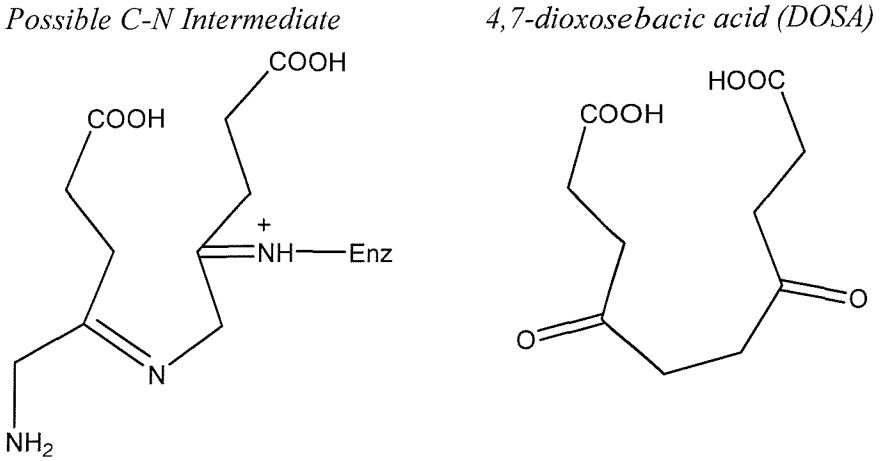


Figure 1.13 DOSA, a possible intermediate mimic modelled on the C=N mechanism. The structures revealed DOSA bound simultaneously to both the 'A' and 'P' sites active site lysines via Schiff bases.

One of the latest ALAD structures to be published is that of a 1.9Å structure of an inactive form of the *P. aeruginosa* ALAD mutant Asp139Asn in complex with the substrate analogue 5-fluorolaevulinic acid (5F-LA) (Frere *et al.*, 2002). In this structure the 5F-LA is found covalently bound to both active site lysines via their 4-oxo groups.

The structure co-crystallised with DOSA, and the structure co-crystallised with 5F-LA both show covalent binding with the 'P' site lysine, confirming previous experimental data. Additionally they show slow covalent binding, *via* Schiff base, with the 'A' side lysine, implying that a double Schiff base mechanism could be possible. It is always dangerous, however, to draw too rigid conclusions from inhibitor studies.

The complex between *P. aeruginosa* and the 5F-LA inhibitor molecules is the first published structure to show two identical substrate analogues bound to the active site. The

benefit of having a structure with both binding sites filled is that the active site residues are very well ordered, allowing a detailed model of the ligand-bound state and has identified all of the residues involved in binding and packing of the substrates around the 'A' and 'P' sites in a Mg^{2+} dependant enzyme. Based on the conformations observed, Frere *et al* (2002) proposed that deprotonation of the 'P' side substrate would be difficult, whereas the 'A' side substrate is predisposed for the formation of the C-C bond formation. Therefore it has been proposed that the first step is the formation of the C-C bond, followed by the formation of the inter-substrate Schiff base. The next step would be the *trans*-elimination of the C-N bond between the enzyme and the substrate, where the bond between enzyme and substrate would be broken. The last step would therefore be the deprotonation-aromatisation step, which would be expected to be very fast and would be the thermodynamic driving force for the reaction (Frere *et al.*, 2002). However, it must be remembered that this is not the native form of the enzyme as Asp139 has been mutated to an asparagine residue, and the amino group of ALA has been replaced with fluorine, so it is still unsure whether a wild-type enzyme would react in the same way.

With all of the data gained, insights into the mechanism have been provided, although the true order of catalysis, whether C-C or C=N bonds are formed first, remains elusive.

1.8 Similarity between ALAD from various species

Although ALADs seem structurally similar and share a high number of conserved residues between species (Figure 1.14), the actual kinetic and biochemical characteristics of ALAD between species can vary a large amount.

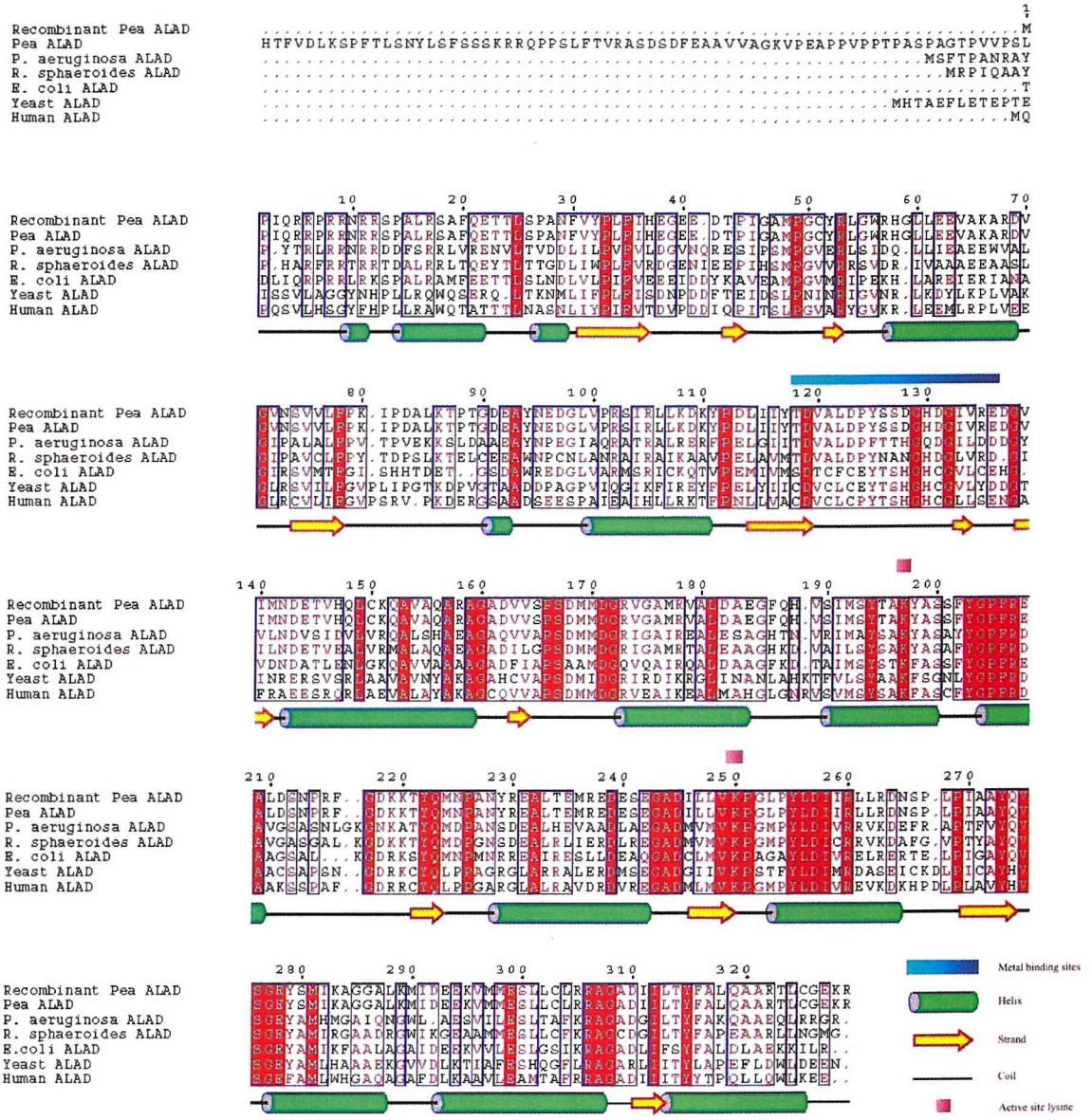


Figure 1.14 Sequence alignment and secondary structure prediction for *P. sativum* ALAD.

Numbering and secondary structure prediction shown is for the recombinant *P. sativum* sequence. The active site lysines are indicated by pink blocks and the metal binding sites are indicated by a blue block. Conserved residues are shown in blocks of red and similar sequences are outlined in blue with similar residues shown in red (Higgins, 1994).

Species	Subunit molecular weight	Oligomeric molecular weight	pK_a		k_m	pI	Metal Ion										
			pK_1	pK_2			Mg^{2+}	Zn^{2+}	Mn^{2+}	Co^{2+}	Ni^{2+}	K^+	NH_4^+	Rb^+	Li^+	PO^{4-}	Na^+
Eschericia coli (Jaffe *et al.*, 1995; Senior *et al.*, 1996)	36554 ± 17 (ESMS)	270000 ± 20000	7.5	8.9	800μM-1mM	5.0	S	A	S	-	-	-	-	-	-	-	-
Rhodobacter spheroids (Nandi *et al.*, 1968; Nandi and Shemin, 1968; Nandi and Shemin, 1968)	-	-	-	-	-	-	A	-	S	-	-	S	S	S	S	-	-
Pisum sativum (Boese *et al.*, 1991; Senior *et al.*, 1996; Kervinen *et al.*, 2000)	39000 ± 5 (ESMS)	309000 ± 20000	8.4	9.5	800μM-1mM	5.2	A	-	S	S	-	-	-	-	-	I	-
Pseudomonas aeruginosa (Frankenberg *et al.*, 1999)	37832 (ESMS)	280000 ± 10000	7.9	9.5	-	5.4	A	S	S	S	S	-	-	-	-	-	-
Bradyrhizobium japonicum (Chauhan and O'Brian, 1993; Chauhan *et al.*, 1997)	38710 (ESMS)	-	-	-	-	-	A	S	-	-	-	S	S	-	-	-	S

Table 1.1 A comparison of the characteristics of five types of 5-aminolaevulinic acid dehydratases, showing reported subunit and oligomeric molecular weight, pK_a, Km, pI and metal ion requirements and stimulators for each species.

The table shows characteristics for all five species including metal ions required for activity "A", metal ion stimulators "S" and inhibitory factors "I".

Species	Possible cooperativity	Concentration dependant specific activity	Half-site Reactivity
Human (Butler, 2003)	No	No	Yes
Eschericia coli (Jaffe *et al.*, 1995; Senior *et al.*, 1996)	No	No	Yes
Rhodobacter sphaeroides (Nandi *et al.*, 1968; Nandi and Shemin, 1968; Nandi and Shemin, 1968)	Yes	-	-
Pisum sativum (Boese *et al.*, 1991; Senior *et al.*, 1996; Kervinen *et al.*, 2000)	Yes	Yes	No
Pseudomonas aeruginos (Frankenberg *et al.*, 1999)	-	Yes	-

Table 1.2 Table showing reported characteristics for five types of ALAD.

As can be seen from the sequence alignment (Figure 1.14) there are a high number of conserved residues and, indeed, sequences throughout the ALAD species that have been characterised. These sequences typically occur at the interface between structural motifs, which form the active site.

The main differences between different species of ALADs seem to occur due to the type of metal ion bound by each type. Those ALADs that bind magnesium ion, in general, appear to have higher pH optima and moderately higher pI values. It has also been noted that magnesium binding ALADs have a lesser susceptibility to oxidation. Those that bind zinc ions appear to have lower pH optima and also lower pI values, and it has also been reported that they have a higher susceptibility to oxidation (Table 1.1). In addition there are also significant differences in the kinetic data reported for the different metal dependant

ALADs. Preliminary characterisation of magnesium-dependant enzymes have shown the possibility of cooperativity between the substrate units, and also concentration dependence. Zinc-dependant enzymes do not appear to follow these kinetics, but show evidence for half-site reactivity (Table 1.2).

1.9 Porphobilinogen deaminase and uroporphyrinogen III synthase

Uroporphyrinogen III is the universal precursor of all tetrapyrroles and its biosynthesis from PBG requires the sequential action of two enzymes, PBG deaminase and uroporphyrinogen III synthase (previously known as cosynthetase). In *E. coli* PBG deaminase is encoded for by the *hemC* gene.

The active site of the *E. coli* enzyme contains two integral PBG-like molecules linked together, termed the dipyrromethane cofactor, attached covalently as a thioether to the apoprotein by a cysteine residue (Cys242). In the first step of the reaction, a PBG molecule reacts covalently with the deaminase cofactor and one molecule of ammonia is released. The first condensation leads to the formation of ring A of uroporphyrinogen III. This step is repeated three more times and results in the formation of an open chain bilane, linked to the cofactor, which is displaced from the enzyme by water to yield the 1-hydroxymethylbilane, preuroporphyrinogen. Preuroporphyrinogen is unstable and, in the absence of uroporphyrinogen III synthase, cyclises at neutral pHs to yield uroporphyrinogen I. However in the presence of uroporphyrinogen III synthase, preuroporphyrinogen is rapidly converted to uroporphyrinogen III by inversion of ring D and cyclisation with the release of a water molecule (Frydman and Feinstein, 1974; Frydman *et al.*, 1976; Evans *et al.*, 1986; Jordan, 1994)(see Figure 1.15).

In plants, PBG deaminase and uroporphyrinogen III synthase are found loosely bound to the plastid membrane (Witty *et al.*, 1993), whereas in animal cells, uroporphyrinogen III is formed within the cytoplasm (Frydman and Feinstein, 1974).

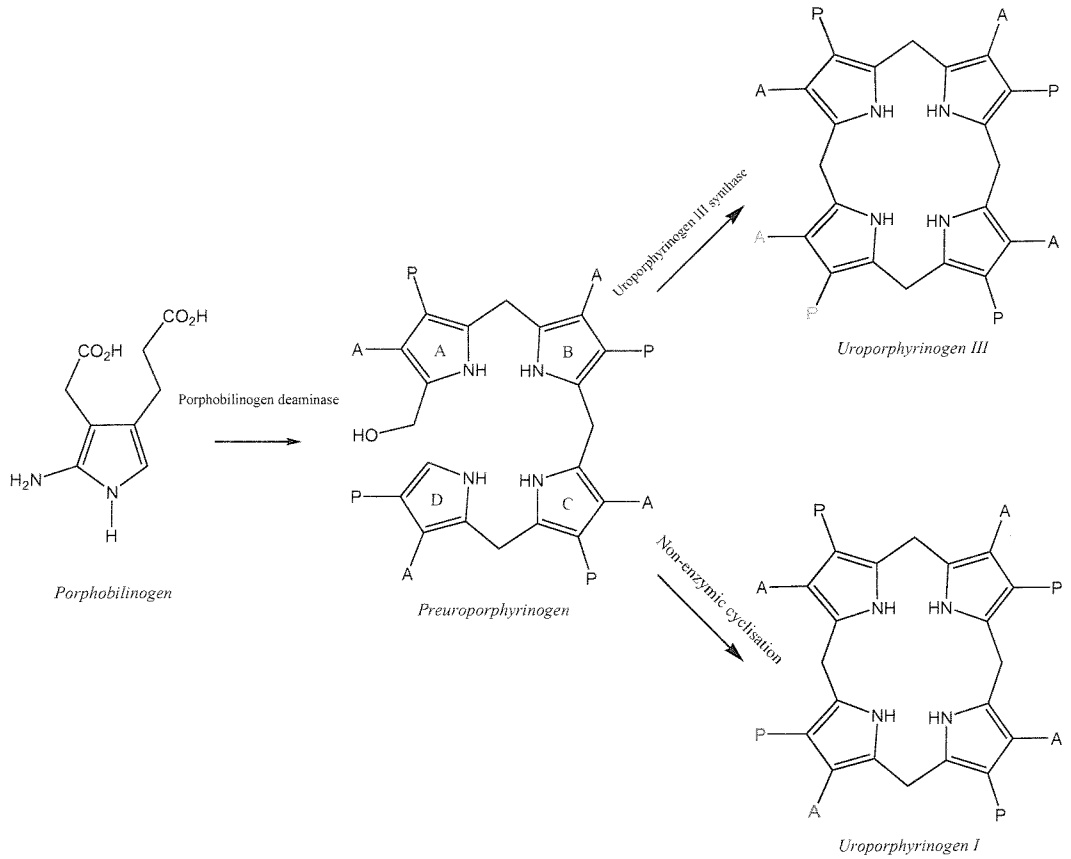
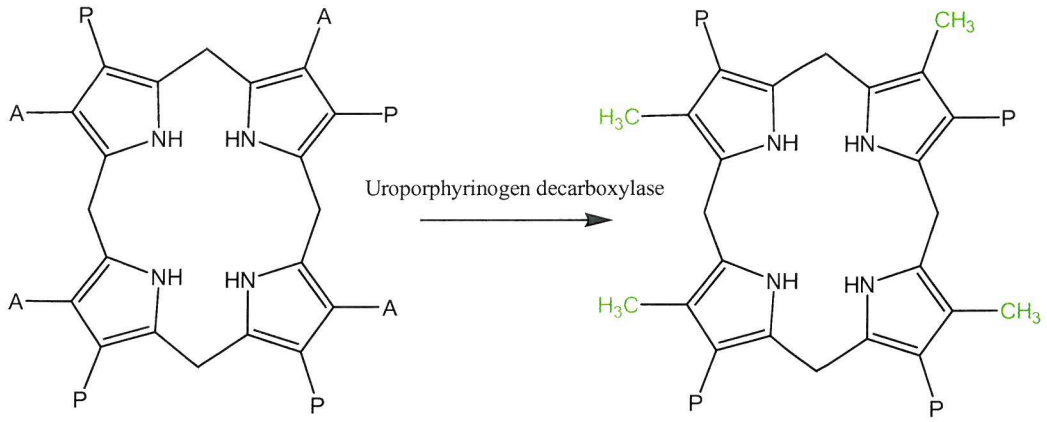


Figure 1.15 The role of porphobilinogen deaminase and uroporphyrinogen III synthase in the biosynthesis of uroporphyrinogen III.

Uroporphyrinogen III, as has been mentioned before, is the point at which the tetrapyrrole biosynthetic pathway branches to yield different end products.

1.10 Uroporphyrinogen decarboxylase

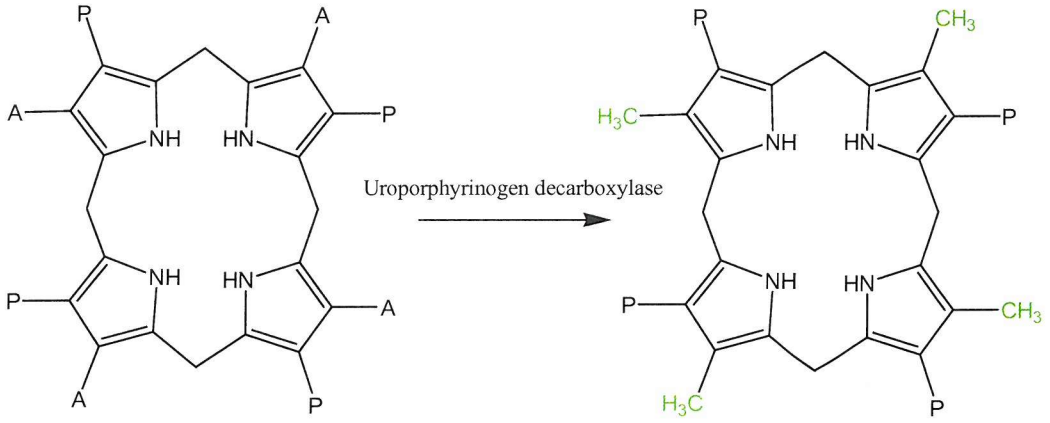
Uropophyrinogen decarboxylase decarboxylates the acetic acid side chains of all four pyrrole rings in uroporphyrinogen III, converting them to methyl groups (Figure 1.16).



Uroporphyrinogen III *Coproporphyrinogen III*

Figure 1.16 The decarboxylation of uroporphyrinogen III catalysed by uroporphyrinogen decarboxylase.

Uroporphyrinogen decarboxylase also acts upon uroporphyrinogen I, yielding coproporphyrinogen I (Figure 1.17). Coproporphyrinogen I has no known function, and its formation is thought to be a blind pathway.



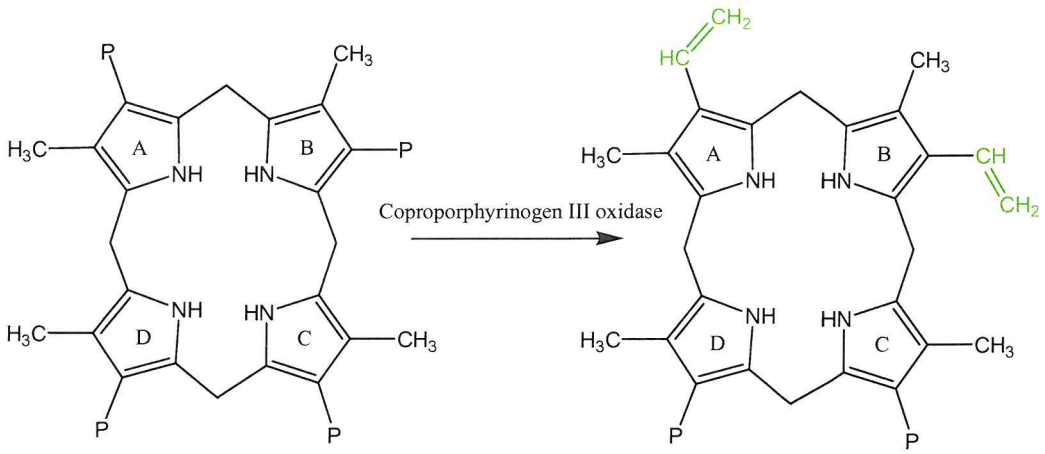
Uroporphyrinogen I *Coproporphyrinogen I*

Figure 1.17 Uroporphyrinogen decarboxylase also decarboxylates the acetic acid side chains of uroporphyrinogen I. Uroporphyrinogen I is a bi-product of the PBG deaminase reaction as discussed previously, although because uroporphyrinogen III synthase levels are high, very little is formed *in vivo* only small amounts of coproporphyrinogen I are formed.

1.11 Coproporphyrinogen III oxidase

Coproporphyrinogen III oxidase, unlike uroporphyrinogen decarboxylase, has a rigid substrate specificity and will not use coproporphyrinogen I as substrate. The reaction

involves the conversion of the propionic side chains of rings A and B to vinyl groups to form the product protoporphyrinogen IX (Figure 1.18):

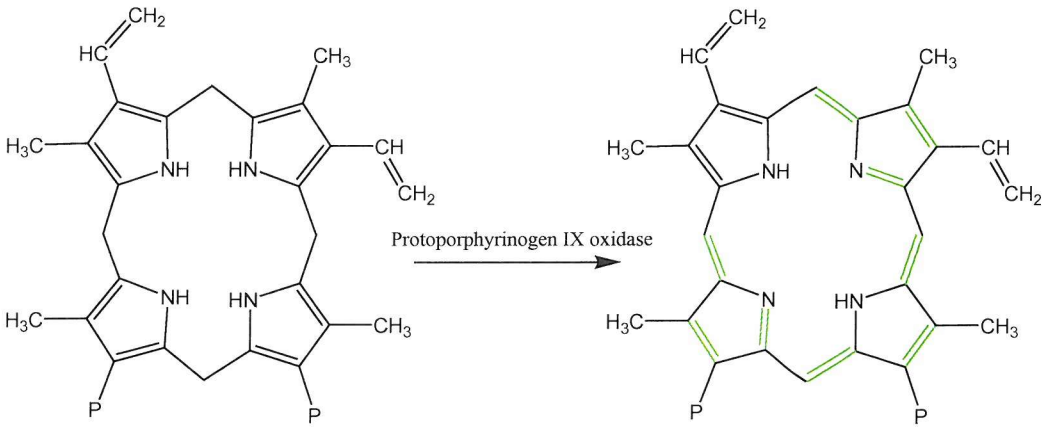


Coproporphyrinogen III *Protoporphyrinogen IX*

Figure 1.18 Coproporphyrinogen III oxidase converts the propionic side chains of rings A and B of corproporphyrinogen III to vinyl groups.

1.12 Protoporphyrinogen IX oxidase

Protoporphyrinogen IX oxidase converts the methylene bridges between the pyrrole rings to methine bridges. Resonance of the double bonds around the entire ring is now possible and results in greater stability of the structure (Willows, 2003):

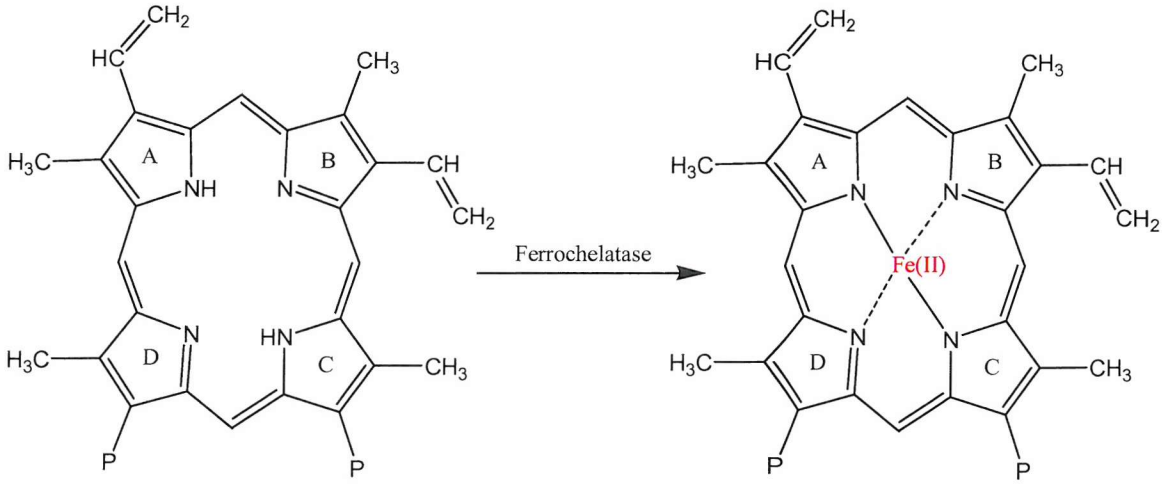


Protoporphyrinogen IX *Protoporphyrin IX*

Figure 1.19 The reaction catalysed by protoporphyrinogen IX oxidase.

1.13 Ferrochelatase

Ferrochelatase catalyses the insertion of ferrous iron into protoporphyrin IX (Figure 1.20) and represents the last stage in haem biosynthesis. The enzyme will also accept cobalt as a substrate and is inhibited by N-methylporphyrins (Ferreira, 1999).

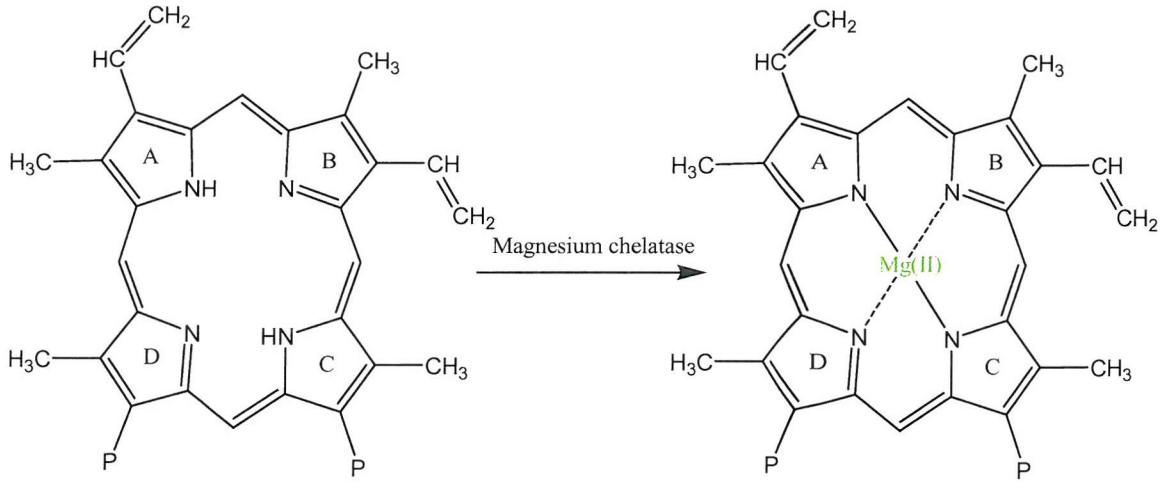


Protoporphyrin IX *Protoheme*

Figure 1.20 The reaction catalysed by ferrochelatase

1.14 Magnesium chelatase

Protoporphyrin IX can either proceed to haem by insertion of Fe^{2+} or embark on the magnesium branch along the chlorophyll biosynthetic pathway. Magnesium chelatase is the first committed step in chlorophyll biosynthesis, with the enzymes prior to this step also shared with the haem biosynthetic pathway. Magnesium chelatase inserts a magnesium ion into the centre of the tetrapyrrole ring and requires a complex of three proteins and several ATPs in order to do so. In developing chloroplasts from cucumber cotyledons, magnesium chelatase only catalyses the insertion of the magnesium ion (Figure 1.21), whereas in *R. sphaeroides*, magnesium chelatase catalyses the insertion of the metal ion and also the methylation of the propionic side chain of ring C to give magnesium protoporphyrin IX mononomethylester(Willows, 2003).

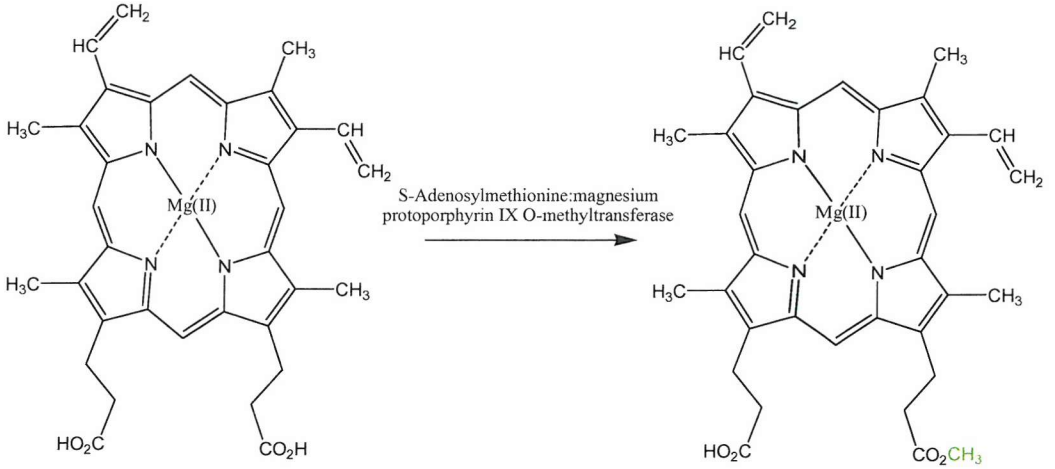


Protoporphyrin IX *Magnesium protoporphyrin IX*

Figure 1.21 Insertion of magnesium into protoporphyrin IX by magnesium chelatase in plants.

1.15 *S*-Adenosylmethionine:magnesium protoporphyrin IX *O*-methyltransferase

The next step in the pathway, in developing chloroplasts from cucumber cotyledons, is the *S*-adenosylmethionine-dependant methylation of the 13-propionate side-chain of magnesium protoporphyrin IX catalysed by *S*-adenosylmethionine:magnesium protoporphyrin IX *O*-methyltransferase. This reaction forms the product, magnesium protoporphyrin IX monomethyl ester (Figure 1.22)(Willows, 2003; Bollivar, 2003).



Magnesium protoporphyrin IX *Magnesium protoporphyrin IX monomethyl ester*

Figure 1.22 Selective methylation of the 13-propionate side chain of magnesium protoporphyrin IX.

1.16 Magnesium protoporphyrin IX monomethyl ester oxidative cyclase

An oxidative cyclisation is required to create the fifth ring of chlorophyll and this reaction is catalysed by magnesium protoporphyrin IX monomethyl ester oxidative cyclase (Willows, 2003):

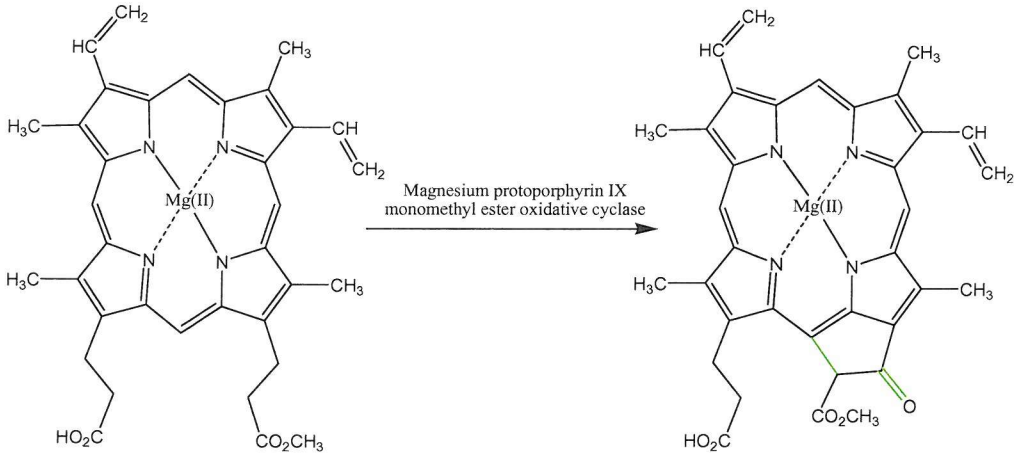


Figure 1.23 The oxidative cyclisation of magnesium protoporphyrin IX monomethyl ester to form the fifth constituent ring.

The reaction mechanism has been shown to proceed through two intermediates, an alcohol and a ketone as shown below to yield, finally, divinylprotochlorophyllide *a* (Leeper, 1989):

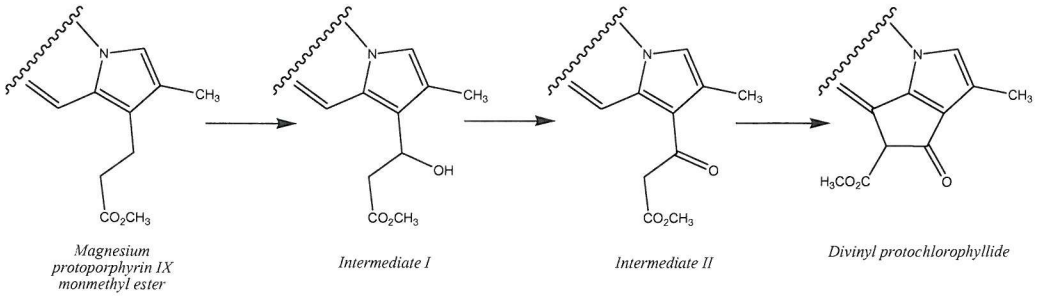


Figure 1.24 The reaction mechanism of magnesium protoporphyrin IX monomethyl ester

1.17 Divinyl protochlorophyllide 8-vinyl reductase

This reaction has been shown to occur at any stage between protoporphyrin IX to chlorophyllide *a*. The enzyme in this instance is shown to catalyse the formation of monovinyl protochlorophyllide from divinyl protochlorophyllide (Figure 1.25), with the vinyl group reduced to an ethyl group. The actual point at which this step occurs i.e. the reduction of the 8-vinyl group, is difficult to characterise and depends on numerous factors

such as species, developmental stage, time in the light or dark, age of the tissue and the light intensity (Willows, 2003).

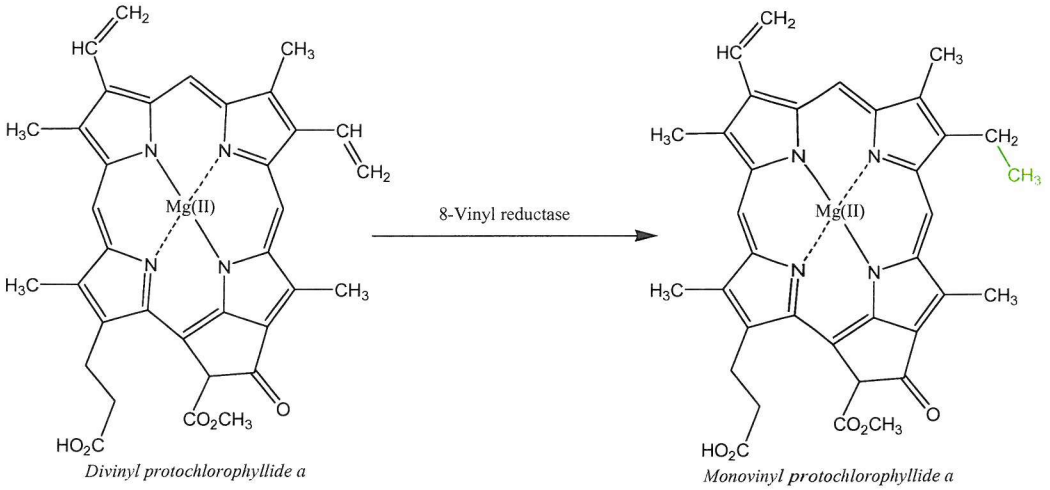


Figure 1.25 The reduction of the 8-vinyl group of divinylprotochlorophyllide a.

1.18 Protochlorophyllide oxidoreductases

Two types of enzymes have been identified that reduce the pyrrole D ring of monovinylprotochlorophyllide *a* (protochlorophyllide *a*) to form chlorophyllide *a* (Figure 1.26). The most studied of the two enzymes is the enigmatic, light-dependant NADPH-protochlorophyllide oxidoreductase (LPOR). LPOR requires light as a "substrate" and is found in all organisms that synthesis chlorophyll, but not those that synthesise bacteriochlorophyll. The second type of enzyme, known as the light-independent protochlorophyllide oxidoreductase (DPOR), is found in flowering plants but appears to be present in most other chlorophyll and bacteriochlorophyll synthesising organisms and allows chlorophyll production in the dark(Willows, 2003).

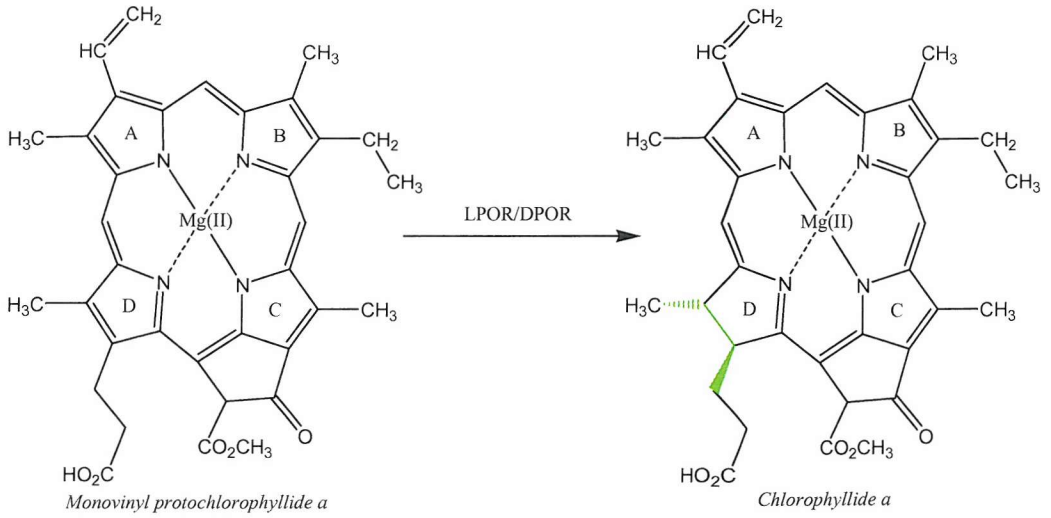


Figure 1.26 The reduction of the D pyrrole ring of monovinylprotochlorophyllide *a* by protochlorophyllide oxidoreductase.

1.19 Chlorophyll *a* synthase

Chlorophyll *a* synthesis is completed with the esterification of chlorophyllide *a* with phytol, a reaction catalysed by chlorophyll synthase (Figure 1.27), although esterification with geranyl-geraniol can also occur (Figure 1.28) (Willows, 2003).

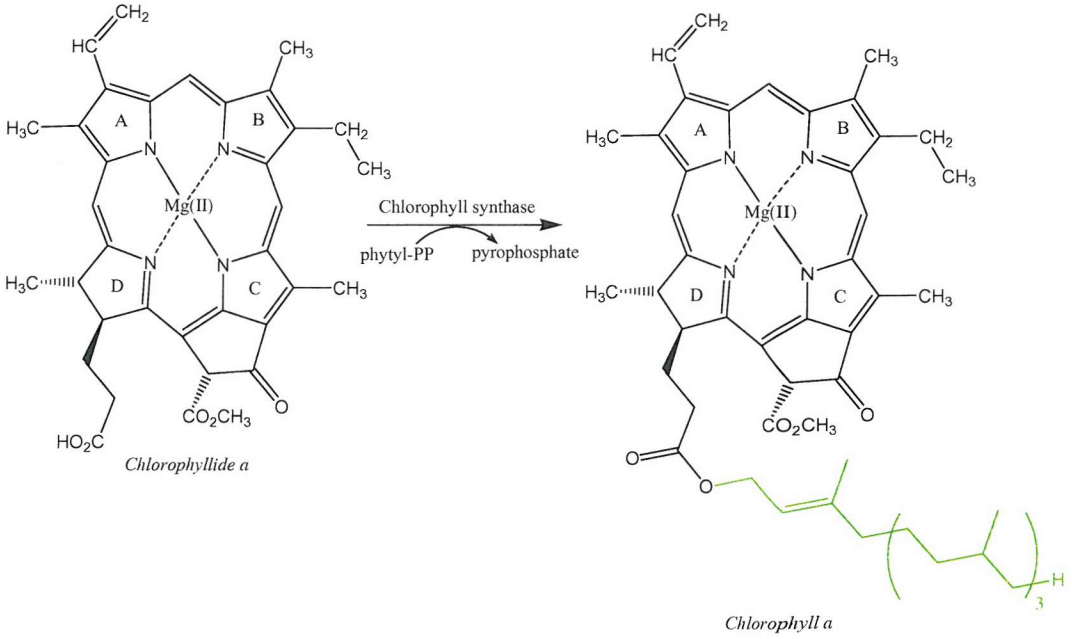


Figure 1.27 Esterification of chlorophyllide *a* with phytol catalysed by chlorophyll synthase.

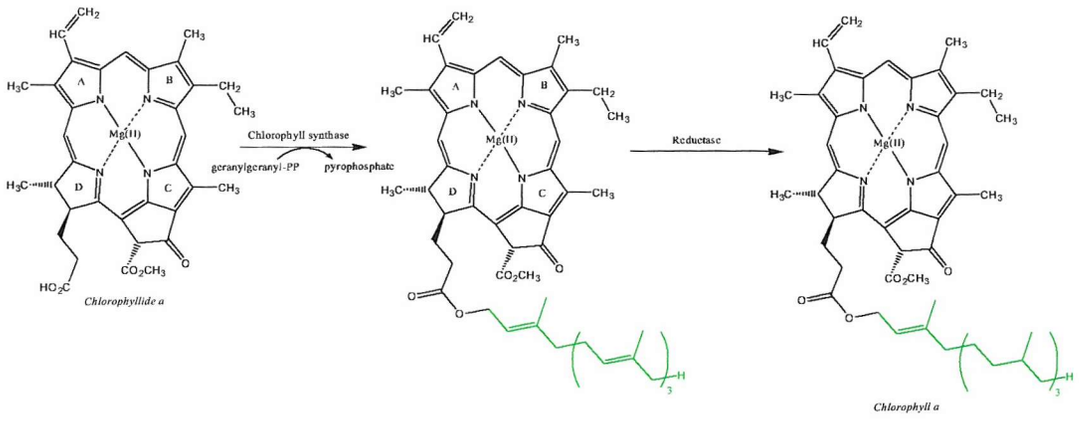


Figure 1.28 Esterification of chlorophyllide *a* with geranylgeraniol catalysed by chlorophyll synthase and reductase (uncharacterised).

1.20 Chlorophyll *b* synthesis

Chlorophyll *b* synthesis only occurs in oxygenic photosynthetic organisms, as molecular oxygen is required for the oxygen for the transformation of the methyl group at C7 to a formyl group. The process is catalysed *via* a hydroxy-intermediate of chlorophyllide *a* to chlorophyllide *b*, by chlorophyllide *a* oxygenase, an enzyme that contains a novel FeS centre (Eggink, 2004) (Figure 1.29).

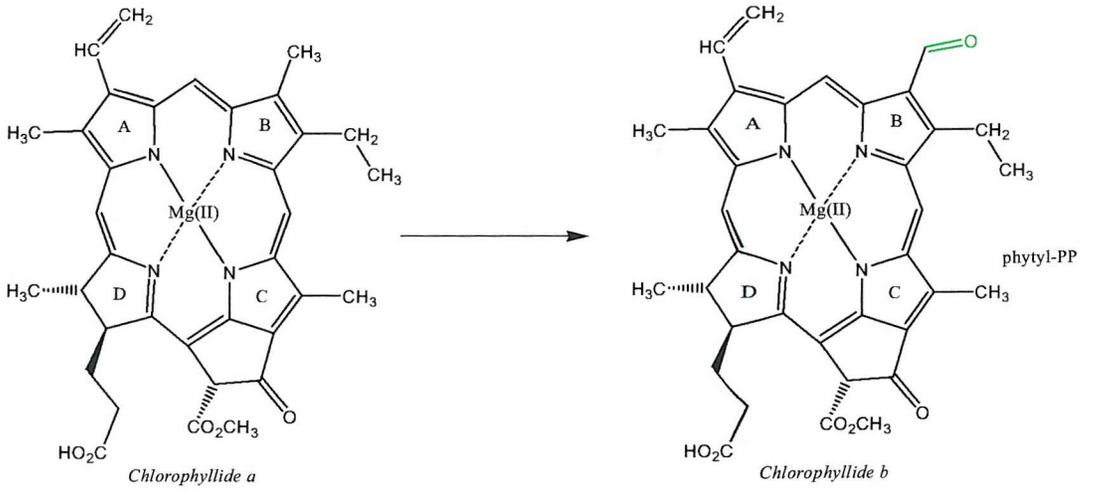


Figure 1.29 The conversion of chlorophyllide *a* to chlorophyllide *b* by chlorophyllide *a* oxygenase

Following this, the chlorophyllide *b* is esterified with either phytol or geranyl-geraniol in a similar fashion to chlorophyllide *a* (Figure 1.30), the difference between the end products being the 7-formyl oxygen component to chlorophyll *b*. This reaction is again catalysed by chlorophyll synthase, which can recognise both substrates (Rüdiger, 2003).

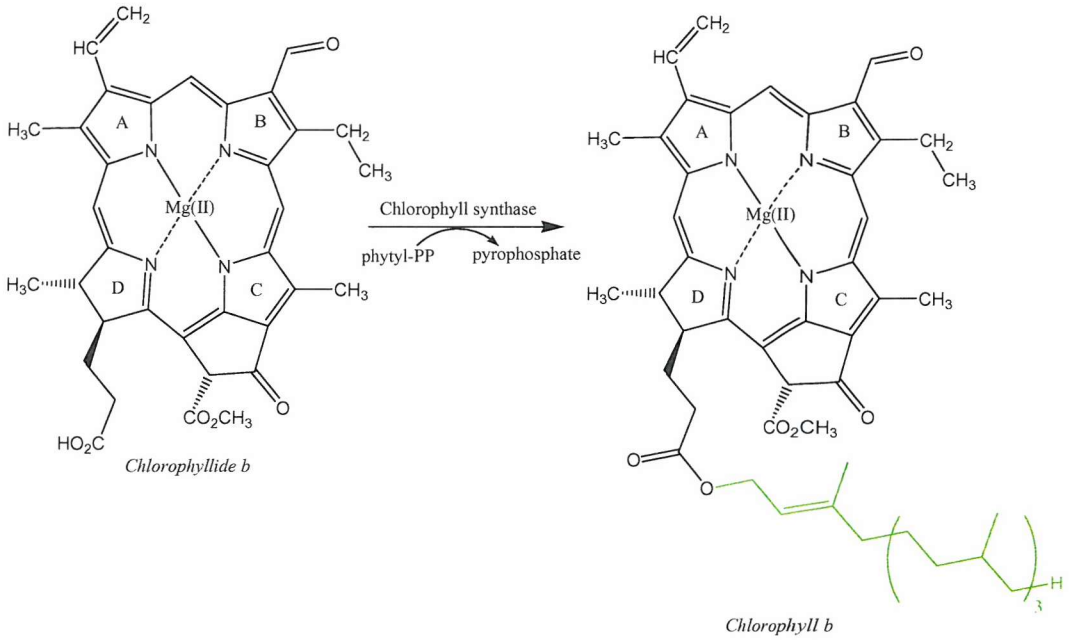


Figure 1.30 The conversion of chlorophyllide *b* to chlorophyll *b*, catalysed by chlorophyll synthase.

1.21 Aims and objectives

The initial purpose of the BBSRC CASE studentship, in conjunction with Syngenta, was to develop an expression system and purification protocol for a recombinant form of the mature *P. sativum* ALAD protein for crystallography trials. The plan was then to obtain diffraction quality crystals and to solve the three-dimensional structure of the enzyme. It was hoped that this would provide information about the nature of the active site cleft that would assist in the development of novel and species-specific inhibitors of ALAD that could be used as lead compounds for herbicides. This work is described in chapters 3 and 4.

The work was partially successful since an expression and purification protocol was successfully devised and applied to the purification of the recombinant protein. The characteristics of the *P. sativum* protein were compared with those of the *E. coli* enzyme,

the structure of which has already been solved. A large number of different crystallography techniques and approaches were undertaken and diffraction quality crystals were obtained. However, it became clear, after many attempts at obtaining phases for the data, that the structural model of *P. sativum* ALAD would not be immediately forthcoming.

The *E. coli* enzyme exhibits the highest protein sequence similarity to *P. sativum* ALAD and this was therefore used as a model for the *P. sativum* ALAD enzyme. Compounds were modelled graphically into the *E .coli* active site model and then chemically synthesised both by myself at Syngenta and by Dr Cheung (University of Southampton). Each of the compounds was tested for inhibition of the *P. sativum* ALAD and IC_{50} values were calculated for each compound (chapter 5).

As has been discussed, the exact mechanism by which the enzyme catalyses the formation of PBG from ALA is still not fully understood. Work by Erskine *et al.* (2003) showed the presence of a novel pyrrole like compound that appears to be a catalytic intermediate bound at the active site of yeast ALAD. The human ALAD structure (Mills-Davies, 2000) had originally shown the presence of such a compound bound at the active site, although the electron density was not detailed enough to identify the precise nature of the molecule. Previous studies had shown that the complex formed between the human ALAD and the compound was very stable, lasting throughout the purification process. Based on this, experiments were conducted to determine the nature of the molecule at the active site of the human ALAD. The experiments showed that the compound bound was indeed a true intermediate of the reaction and could be subject to turnover in the presence of ALA, LA or PBG to form the product PBG. These experiments are described in chapter 6.

Finally, data sets from a previously crystallised recombinant human ALAD were processed, molecular replacement was achieved and the structure was refined in an attempt to prove the structure of the bound intermediate. Unfortunately, the electron density at the active site was not sufficiently resolved to allow the characterisation of the bound compound, however, the new model did provide better evidence for certain areas of the structure that

had not been as well resolved in the model of the native human ALAD, especially the active site loop region. This work is described in chapter 7.

Chapter Two: Materials and methods

2.1 Sterilisation

All pipette tips, media, stock solutions, microfuge tubes and cryotubes were sterilised in a British Steriliser autoclave at 120°C and 1 bar for twenty minutes. Tips and tubes were then dried in a 65°C oven for at least twenty four hours.

2.2 Culture Media

Materials

Bactotryptone and yeast extract were purchased from Difco Laboratories, Bactoagar from LabM and sodium chloride from Sigma-Aldrich.

Luria broth (LB) media

Bactotryptone (10g), Bactoyeast extract (5g) and sodium chloride (5g) were dissolved in 1 litre of distilled water and autoclaved immediately.

Luria Broth plates

Bactoagar (15g) was added to a litre of the above media and autoclaved. Antibiotics were added once the media had cooled to 40°C and plates were then poured to a depth of 5mm and stored until required at 4°C for a maximum of two weeks.

Ampicillin Stock

Ampicillin, sodium salt, was dissolved in AnalaR water to a concentration of 100mg/mL, and filtered through a 0.2μm filter. This solution was always made freshly and used at a final concentration of 100μg/mL.

Isopropylthiogalactoside (IPTG) stock

IPTG stock solution (1mM) was made by dissolving 2.38g of IPTG in 10mL of AnalaR water, and filtering through a 0.2μm filter. This stock solution was always made immediately before it was required.

2.3 Bacterial cells strains and plasmids used in this study

The bacterial cell strains and plasmids used are displayed in Tables 2.1 and 2.2, and a map of the pT7-7 plasmid is shown in Figure 2.1.

Strain	Genotype	Origin
E. coli BL21(DE3)	F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3)	Novagen
E. coli BL21(DE3)pLysS (Derman, 1993)	F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cm^R)	Novagen
E. coli(DH5α) (Woodcock, 1989)	F- φ80*lacZ*M15 Δ(*lacZ*YA A-*arg*F)U169 *rac*A1 *end*A1 *hsd*R17(r_k^- m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Invitrogen
E. coli(DE3) Rosetta (Derman, 1993)	F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) $pRARE^2$ (CM^R)	Novagen
E. coli TB1 *hemB*$^-$ (Baldwin, 1984)	F^- *ara* Δ(*lac-pro*AB) *rps*L (Str^r)[f80 *dlac*D(*lacZ*)*M15*] *thi hsd*R (r_K^- m_K^+) *hemB*$^-$	In Lab

Table 2.1 Bacterial cell strains used in this study

Plasmid	Properties
pT7-7	T7 promotor system(Tabor, 1985)
pPALAD	pT7-7 carrying *hemB* (984kb fragment of the *P. sativum* ALAD gene)
pPALAD1	pPALAD with aspartate 131 mutated to asparagine
pPALAD2	pPALAD with aspartate 139 mutated to asparagine
pPALAD3	pPALAD with aspartate 176 mutated to asparagine
pHALAD	pT7-7 carrying human *hemB*
pCAR262	pUC19 carrying *E. coli hemB*

Table 2.2 Vector and recombinant plasmids used in this study.

2.4 Cell strain maintenance

All strains were stored at -20°C. Every six months, cell strains were maintained by growing 20mL overnight cultures of each strain in LB media. Aliquots of 500μl were then taken and mixed with 500μl of glycerol. Glycerol acts as a cryoprotectant for the cells and reduces cell damage at -20°C.

2.5 Plasmid maintenance

All plasmid DNA was stored at -20°C. At yearly intervals, plasmid DNA was transformed into *E. coli* (DH5α). Single colonies were picked and 20mL overnight cultures containing ampicillin were grown. The following morning, cell cultures were centrifuged at 3000 x g for 15 minutes before purification of the plasmid DNA using the Promega `Wizard *Plus* SV Minipreps DNA Purification System`.

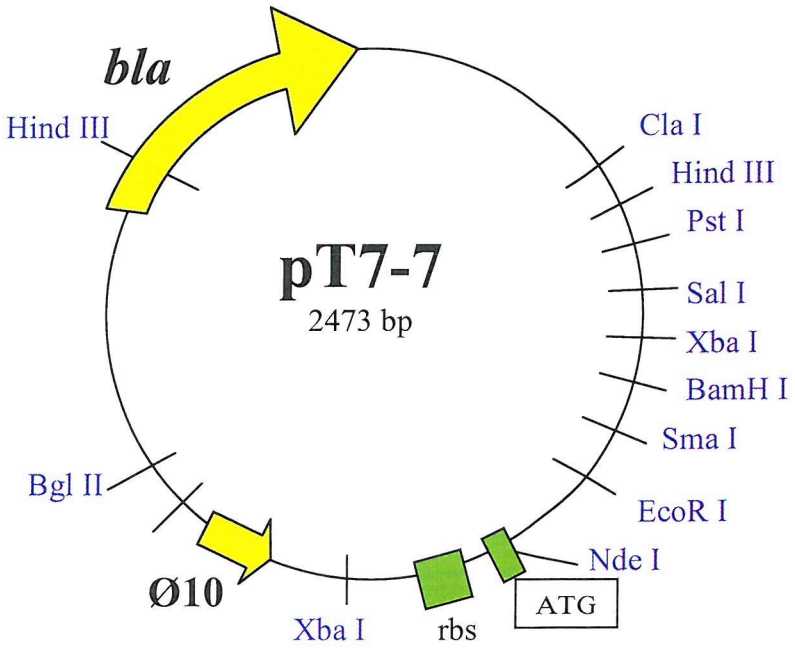


Figure 2.1 Map of the plasmid pT7-7. The *P. sativum* ALAD cDNA was ligated between the sites *NdeI* and *Bam*HI in the 5'- to 3'-orientation. The inserted cDNA is in-frame with the ATG start codon of the vector DNA and the optimal distance from the ribosome binding site (rbs), both being critical for efficient translation of the transcribed mRNA. The *bla* gene confers ampicillin resistance, allowing selection of recombinant plasmids.

2.6 DNA restriction digests

2.6.1 Materials

All materials required for the restriction digests were obtained from Promega UK Ltd, Delta House, Chilworth Research Centre, Southampton, UK and Amersham Biosciences UK Limited, Amersham Place, Little Chalfont Buckinghamshire, UK.

2.6.2 Method

10ng of plasmid DNA was mixed with 6µl of all4one + buffer and 1.5µl of *Nde*I restriction enzyme and the reaction was incubated for 1½ hours at 37°C. After this time 1.5µl of *Bam*HI restriction enzyme was added and the reaction was allowed to continue for a further 1½ hours at 37°C. The products of the digest were analysed by 1% agarose gel electrophoresis.

2.7 Agarose gel (1%) electrophoresis

2.7.1 Materials

Tris base and EDTA were obtained from Sigma-Aldrich and glacial acetic acid was obtained from Fisher.

Promega DNA ladders (1kb and 100bp), supplied by Promega, were used to estimate the size of DNA bands (Figure 2.2).

Figure 2.2 Promega 1kb and 100bp DNA ladders.

Ethidium bromide stock solution

Ethidium bromide (20mg) was dissolved in 2mL of AnalaR water to give a final concentration of 10mg mL^{-1} and stored at 4°C.

TAE buffer (50x)

242g of Tris base, 57.1mL of glacial acetic acid and 18.6g of EDTA were dissolved in a litre of AnalaR water. The solution was then filtered and stored at room temperature until needed.

2.7.2 Methods

Preparation of agarose gels

2mL of 50x TAE buffer was diluted with 100mL of AnalaR water, and 1g of agarose was added. The mixture was boiled using a hot-plate and, once cooled to 50°C, ethidium bromide solution was added to a final concentration of 0.5μg mL^{-1}. Gels were then poured into plates with the combs inserted and allowed to set at room temperature, after which the combs were removed. A further 2mL of 50x TAE was then diluted with 100mL of AnalaR water and poured into the gel tank, so as to cover the gel and the wells.

DNA sample preparation

Samples were prepared in AnalaR water with the addition of Promega 6x blue/orange loading buffer and a total of ~5ng DNA loaded per lane of gel, in a total sample volume of 10μl.

Electrophoresis

Gels were run horizontally at 100V until full separation occurred. The gels were visualised using UV transillumination.

2.8 Preparation of competent bacterial cells for transformation

LB media (10mL) was inoculated with a colony of *E. coli* BL-21 and incubated overnight at 37°C. An aliquot (200μl) of the overnight culture was added to 20mL of LB media and incubated at 37°C until the O.D.600nm was ~0.3. The cells were harvested by centrifugation at 3,000 x g for 10 minutes and the supernatant was discarded. The cells were resuspended in 10mL of cold 50mM $CaCl_2$ and left on ice for 1 hour followed by centrifugation for 10 minutes at 3,000 x g. The pellet was resuspended with 1-2mL of cold 50mM $CaCl_2$ and left in ice for up to 24 hours.

2.9 Transformation of competent cells

An aliquot (200µl) of competent cells was added to 1µl of the required plasmid DNA in a sterile Eppendorf tube and the mixture was left on ice for 25 minutes. The cells were then "heat shocked" at 42°C followed by incubation at 37°C for 1 hour, without shaking. An aliquot (100µl) of transformed cells was spread onto an LB agar plate, containing ampicillin (100µg mL^{-1}), and incubated at 37°C.

2.10 Expression studies

For recombinant plasmids containing the T7-promoter, a single colony was inoculated in 10mL of LB media supplemented with ampicillin (100µg/mL) and grown to an O.D of 1.0 at 600nm. Next, IPTG was added to a final concentration of 1mM so as to induce production of the desired protein and the cultures were grown for a further 3 hours at 37°C before harvesting by centrifugation (3000 x g/10 minutes at 4°C). The bacterial cell pellet was resuspended in 0.5mL of 20mM Tris/HCl buffer, pH8.2, containing 1mM DTT in an Eppendorf tube. The extract was sonicated at an amplitude of 10 microns in five bursts (30 seconds on / 60 seconds off). The crude sonicates were centrifuged at 4°C for 10 minutes at 3000 x g, before transferring the supernatant to a separate tube. The pellets were centrifuged again and any remaining supernatant was removed by aspiration. The supernatant was then assayed for protein and ALAD activity and analysed by SDS-PAGE.

2.11 Determination of protein concentration

2.11.1 Materials

Bio-Rad protein assay reagent was obtained from Bio-Rad laboratories, Bio-Rad House Maylands Avenue, Hemel Hempstead, Hertfordshire, UK.

2.11.2 Methods

Protein concentration was determined using the Bio-Rad protein assay. A 1mL assay consisted of 10µL of protein solution, 790µL of distilled water and 200µL of Bio-Rad reagent. Bio-Rad reagent is based on the Bradford assay (Bradford, 1976). The protein

binds to the dye, Coomassie brilliant blue G-250, in a concentration-dependent manner. This assay was routinely used for protein samples between 200 and 1,400 μg/mL. The colour was allowed to develop for 5 minutes at room temperature and the absorbance was measured at 595nm. The concentration was calculated using the conversion Abs 0.1 = 1.95μg of protein.

2.12 Assay for 5-aminolaevulinic acid dehydratase (ALAD) activity

2.12.1 Materials

All chemicals were obtained from Sigma-Aldrich Company Ltd., The Old Brickyard, New Road, Gillingham, Dorset, UK.

Modified Ehrlich's reagent

1g 4-Dimethylaminobenzaldehyde, 42mL of glacial acetic acid and 8mL of perchloric acid (60%) were mixed and the reagent was stored in the dark at 4°C and was used within one week.

Stop solution

10% Trichloracetic acid containing 0.1M mercuric chloride was stored in the dark in a foil-covered bottle at 4°C.

Stock 5-aminolaevulinic acid (5-ALA) solution

A stock solution of 100mM ALA was made by dissolving 0.335g of 5-ALA hydrochloride in 20mL of AnalaR water. Aliquots (500μl) were stored at -20°C until required.

***E. coli* ALAD assay buffer** (Norton, 1999)

50mM Tris/HCl buffer, pH 8.0, containing 100μM $ZnCl_2$ and 5mM 2-mercaptoethanol.

Human ALAD assay buffer (Butler, 2003)

50mM potassium phosphate buffer, pH 6.8, containing 100μM $ZnCl_2$ and 5mM dithiothreitol (DTT).

***P. sativum* ALAD assay buffer**

50mM bis-Tris/HCl propane buffer, pH 9.3, containing 50mM $MgCl_2$ and 10mM 2-mercaptoethanol.

2.12.2 Method for the assay of ALAD

An aliquot (10µL) of ALAD enzyme solution was added to the required assay buffer and pre-warmed for 5 minutes at 37°C and ALA solution was added to a final assay concentration of 10mM. After mixing, incubation was carried out at 37°C for 5 minutes. The reaction was terminated by the addition of an equal volume of stop solution. Ehrlich's reagent was then added in a volume equal to the total volume of the assay plus the stop solution. Samples were centrifuged in a bench-top centrifuge and the absorbance of the supernatant was measured at 555nm. Specific activity was calculated as µmoles PBG formed/mg of protein/hour, using the extinction co-efficient for the coloured complex formed by Ehrlich's reagent reacting with PBG at 555nm of 6.02×10^4 $M^{-1}cm^{-1}$.

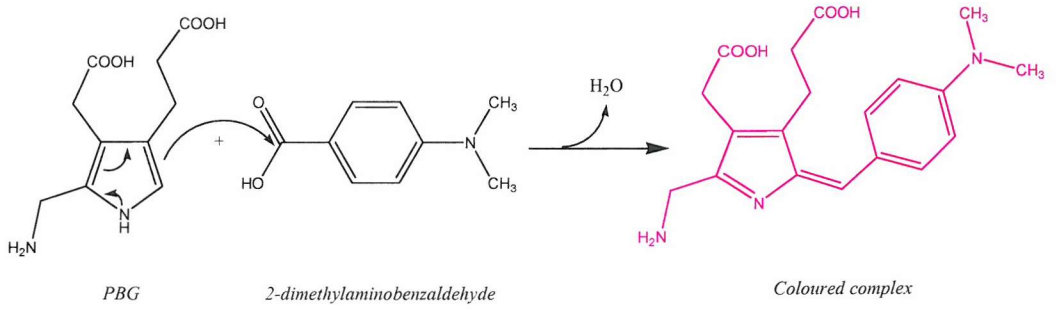


Figure 2.3 The reaction of PBG with Ehrlich's reagent (4-dimethylaminobenzaldehyde) to form a coloured complex.

2.13 Polyacrylamide gel electrophoresis (PAGE)

All denaturing polyacrylamide gel electrophoresis was carried out according to the method of Laemmli (1970). The protocol for preparation of native gels was similar, except that SDS and β-mercaptoethanol were omitted throughout.

2.13.1 Materials

All chemicals required for polyacrylamide gels electrophoresis were obtained from Sigma-Aldrich. The equipment employed was purchased from BIORAD.

Protein standard M_r markers

Throughout the course of the studies, the following M_r markers were used to estimate the size of sample proteins. These were prepared according to the manufacturer's instructions.

Protein	M_r
Bovine serum albumin	66,000
Egg albumin	45,000
Glyceraldehyde-3-phosphate dehydrogenase	36,000
Carbonic anhydrase	29,000
Trypsinogen	24,000
Soybean trypsin inhibitor	20,100
α-Lactalbumin	14,200

Table 2.3 Dalton VII marker protein M_r standards (Sigma-Aldrich).

Stock SDS running buffer (l0x)

SDS running buffer was prepared using 144g glycine, 30g Tris base and 10g SDS dissolved in distilled water and adjusted to pH8.3 using concentrated HCl. The volume was made up to 1 L with distilled water. Running buffer for non-denatured PAGE was made up in the same way, with the exclusion of SDS.

Disruption buffer (10x)

Disruption buffer was made from 1.2mL of 0.5M Tris/HCl buffer, pH 6.8, containing 1mL of glycerol, 2mL of 10% SDS and 0.5mL of 0.1% w/v bromophenol blue mixed with 4.8mL of H_2O to make a final volume of 9.5mL. Aliquots (450µl) of this stock solution were taken and mixed with 50µl of 2-mercaptoethanol prior to use. The same method was used to prepare disruption buffer for non-denaturing gels, however, SDS and 2-mercaptoethanol were omitted.

Staining solution for PAGE gels

Staining solution was made with 2.5g Coomassie brilliant blue dissolved in 400mL methanol and 70mL glacial acetic acid. The volume was then adjusted to 1L with distilled water.

Destain solution for PAGE

Destain solution was prepared by mixing 400mL methanol and 70mL acetic acid with 530mL distilled water.

2.13.2 Methods

Protein sample preparation

Protein samples were prepared by the addition of the appropriate amount of disruption buffer followed by boiling the mixture for 2 minutes.

Preparation of the polyacrylamide gels

The composition of the resolving and stacking gels are shown in Table 2.4 and Table 2.5, respectively.

Acrylamide (30%)	3.19mL
1.5M Tris/HCl, pH 8.8	1.88mL
Distilled H_2O	2.24mL
SDS 20%w/v	75µl
APS 10% w/v	75µl
TEMED	8µl

Table 2.4 Resolving gel for SDS PAGE gel (12% acrylamide)

Acrylamide (30%)	0.75mL
0.5M Tris/HCl pH 6.8	1.5mL
Distilled H_2O	3.6mL
SDS 20% w/v	75µl
APS 10% w/v	75µl
TEMED	8µl

Table 2.5 Stacking gel for SDS PAGE gel (12% acrylamide)

Electrophoresis

After the samples were loaded onto the gel, electrophoresis was carried out at 120V until the bromophenol blue dye front had reached the bottom of the gel. The gel was then submerged in stain solution containing Coomassie brilliant blue dye for 1 hr before being washed and submerged in destain solution. Gels were visualized by transillumination.

2.14 Purification of *E. coli* ALAD

Purification of the *E. coli* ALAD was based upon the method described by previously (Norton, 1999).

2.14.1 Growth of *E. coli* over-expressing the *E. coli hemB* gene.

The pCAR262 clone was transformed into the *E. coli* (CR261 strain) using the method described earlier. Single colonies from the resulting plates were used to inoculate 20mL growths of LB media containing ampicillin (100µg/mL) that were grown overnight at 30°C. These were then used to inoculate 800mL of LB media containing ampicillin (100µg/mL) in 2L baffled flasks. The flasks were incubated at 37°C, with shaking, at 160rpm for 12 hours after which time the cells were harvested by centrifugation at 6000 x g for 12mins at 4°C. The pellets were used immediately or were frozen at -20°C.

2.14.2 Sonication

The bacterial pellet from 3.2L of culture was resuspended in approximately 30mL of 20mM Tris/HCl buffer, pH 8.2, containing 5mM DTT, 2mM 2-mercaptoethanol, 100µM $MgCl_2$ and 200µM PMSF. The resuspended cells were then sonicated on ice for a total of twenty

cycles, each cycle consisting of 20 seconds sonication at an amplitude of 10 microns with a 40 second cooling period.

2.14.3 Heat treatment

The sonicated extract was placed in a round bottomed flask and rapidly heated to 60°C in a water bath at 90°C, with constant swirling. This temperature was maintained for 3 minutes, after which the solution was rapidly cooled by swirling the flask in an ice-water bath. The heat-treated sample was then ultracentrifuged at 100,000 x g for 25 minutes at 4°C. The pellet, containing denature protein, was discarded and the supernatant was retained.

2.14.4 Ion-exchange chromatography

The supernatant from above was loaded directly onto a DEAE-Sepharose ion-exchange column which had been previously equilibrated with 50mM Tris/HCl buffer, pH 8.2, containing 2mM 2-mercaptoethanol and 100μM $ZnCl_2$. The fractions were eluted with a gradient of 100-400mM NaCl dissolved in the aforementioned buffer in a total volume of 500mL. Fractions (5mL) were collected over the elution gradient, assayed for activity and protein and analysed on SDS PAGE. The fractions with the highest specific activity were pooled and precipitated by treating with 70% saturated ammonium sulphate and the pellet was collected by centrifugation at 100,000 x g for 25 minutes at 4°C.

2.14.5 Gel-filtration chromatography

The redissolved protein was loaded onto an Sephacryl S-300 gel filtration column, equilibrated in 50mM Tris/HCl buffer, pH 8.4, containing 2mM 2-mercaptoethanol and 50μM $ZnCl_2$ and the fractions were eluted in 5mL samples. The fractions were again assayed for activity, protein concentration and analysed by SDS PAGE and those with the highest specific activity were pooled and retained for further work.

2.14.6 Storage of pure *E. coli* ALAD

Pure *E. coli* ALAD was stored as an ammonium sulphate pellet at -20°C until required.

2.15 Purification of the human recombinant ALAD

The purification of the human ALAD was based upon previous work by Butler (2003).

2.15.1 Growth of *E. coli* over-expressing the recombinant human ALAD.

The pHALAD clone was transformed into the *E. coli* (DE3) Rosetta strain, using the method described earlier. Individual colonies of ampicillin-resistant *E. coli* expressing the ALAD were grown in 10mL of LB media containing ampicillin (10μg mL^{-1}) at 30°C overnight. The overnight cultures were each used to inoculate 800mL of LB media containing ampicillin (100μg mL^{-1}) in 2L baffled flasks. The flasks were incubated at 37°C, with shaking at 160rpm. until the $O.D._{600nm}$ was approximately 1.0. IPTG was then added to a final concentration of 1mM and the cultures were incubated for a further 25 hours at 25°C. Cells were harvested by centrifugation for 12 minutes at 6,000 x g at 4°C. The pellets were used immediately or were frozen at -20°C.

2.15.2 Sonication

The bacterial pellet from 3.2L of media was resuspended in approximately 80mL of 50mM potassium phosphate buffer, pH 6.8, containing 5mM 2-mercaptoethano, 50μM $ZnCl_2$ and 100μM PMSF. The resuspended cells were then sonicated on ice for a total of twenty cycles, each cycle consisting of 20 seconds sonication at an amplitude of 10 microns with a 40 second cooling period.

2.15.3 Heat treatment

The sonicated extract was placed in a round bottomed flask and rapidly heated to 60°C in a water bath at 90°C, with constant swirling. This temperature was maintained for 3 minutes, after which the solution was rapidly cooled by placing the flask in an ice-water bath. The heat-treated sample was then ultracentrifuged at 100,000 x g for 25 minutes at 4°C. The pellet was discarded and the supernatant was retained.

2.15.4 Ion-exchange chromatography on Bio-Rad Bio-Gel hydroxyapatite (HTP)

The supernatant from the heat treatment step was loaded onto a column of Bio-Rad Bio-Gel hydroxyapatite (5.5cm x 10cm) that had previously been equilibrated with the same buffer used in the previous step. After loading onto the column the enzyme was washed with 1L of the same buffer. A linear gradient (500mL total; 50mM-300mM potassium

phosphate buffer) was applied to the column and 5mL fractions were collected. The fractions were assayed and purity was judged by SDS-PAGE, before the most active fractions were pooled. A 70% saturated $(NH_4)_2SO_4$ precipitation was carried out on the pooled fractions and the precipitate was collected by centrifugation as before. The resulting pellet was resuspended in 1.5mL of 100mM potassium phosphate buffer, pH6.8, containing 5mM 2-mercaptoethanol, 50μM $ZnCl_2$ and 100μM PMSF.

2.15.5 Gel-filtration chromatography on a Sephacryl S-300 column

The resuspended pellet from the hydroxyapatite column step was applied to a Sephacryl S-300 gel filtration column (3cm x 1m) and the column was developed with 100mM potassium phosphate buffer, pH 6.8, containing 5mM 2-mercaptoethanol and 100μM PMSF. Fractions containing protein were assayed for ALAD activity and purity was judged by SDS PAGE, before the fractions were pooled and concentrated using a Vivaspin concentration vessel to a volume of 2mL.

2.15.6 Chromatography on a Mono-Q ion-exchange column

The concentrated pooled fractions from the S-300 column chromatography were loaded onto a Mono-Q column pre-equilibrated with 20mM Tris/HCl, pH 7.2, containing 5mM 2-mercaptoethanol 50μM $ZnCl_2$. Fractions were eluted using a concentration gradient (total volume 40mL; 0-400mM NaCl) with 0.5mL fractions collected. Again, fractions containing protein were concentrated and assayed for ALAD activity and purity was judged by SDS PAGE, before the fractions were pooled.

2.15.7 Storage of human recombinant ALAD

The pooled fractions were concentrated using a 70% ammonium sulphate precipitation step and stored as an ammonium sulphate pellet at -20°C until required.

2.16 Protein crystallisation methods

2.16.1 Siliconised coverslips

Coverslips for the hanging drop crystallisation method were prepared by washing them in methanol (1 min), dipping in dimethyldichlorosilane solution (1 min) and washing them again in methanol (1 min) before oven-drying.

2.16.2 Crystal screens

Crystal screens were conducted using two methods, the hanging drop method (Figure 2.4) (Rhodes, 2000) and the counter-diffusion method (Figure 2.5) (Gavira, 2002; Gavira *et al.*, 2002). Both methods rely upon the theoretical protein solubility curve (Figure 2.6) which shows the solubility over a time period of a given protein.

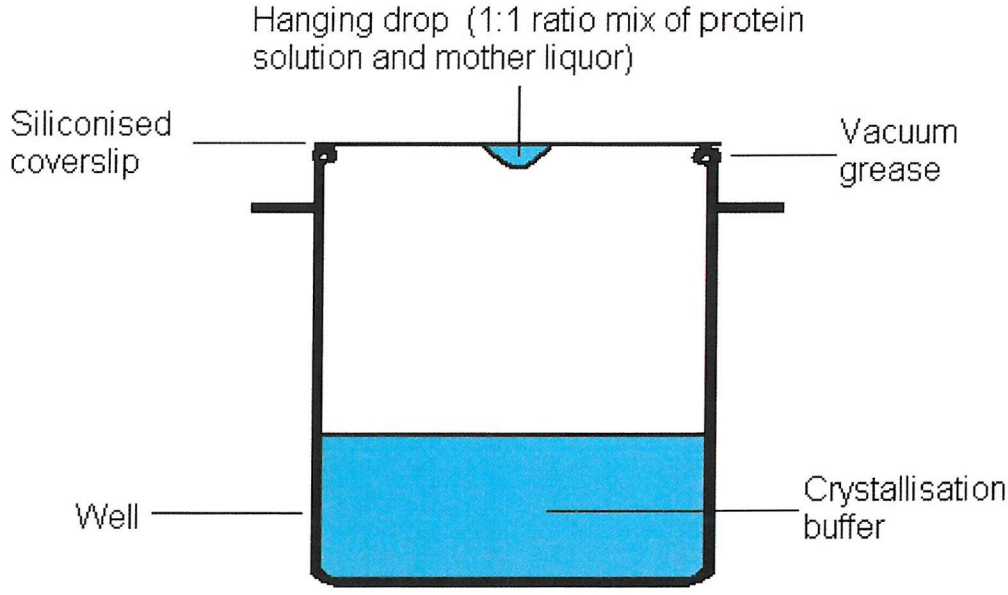


Figure 2.4 The hanging drop method. This method relies upon vapour diffusion to increase gradually the concentration of precipitant and the protein of the hanging drop to that of the crystallisation buffer. The hanging drop was composed of a mixture of 5µl protein solution and 5µl of mother liquor.

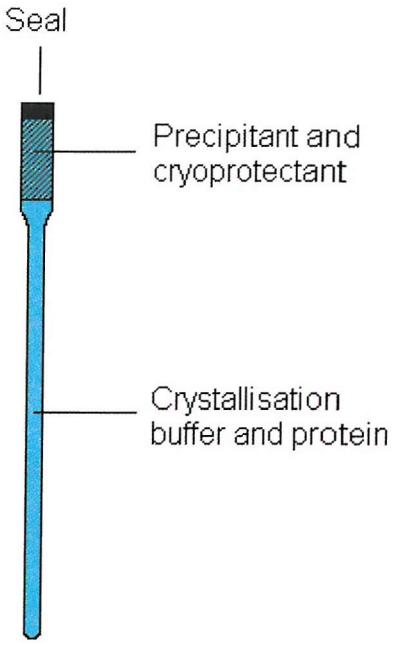


Figure 2.5 The counter-diffusion method. Crystallography relies upon the slow diffusion of the precipitant and cryoprotectant mixture into the crystallisation buffer and protein mix to create a concentration gradient throughout the Lindemann tube. This technique allows direct freezing of the crystals formed without handling, which eliminates damage to the crystals.

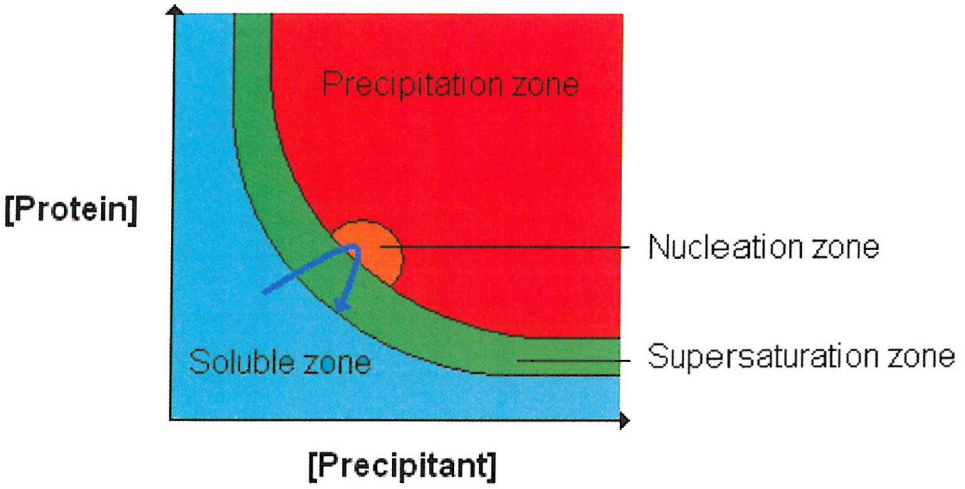


Figure 2.6 A protein solubility curve. The graph shows the variation of protein and precipitant concentration ranges of a crystallisation experiment.

2.16.3 Wet mounting of crystals

An alternative to freezing the crystals is that of wet mounting of the crystal (Figure 2.7). This technique can be less damaging to the crystals (McRee, 1993).

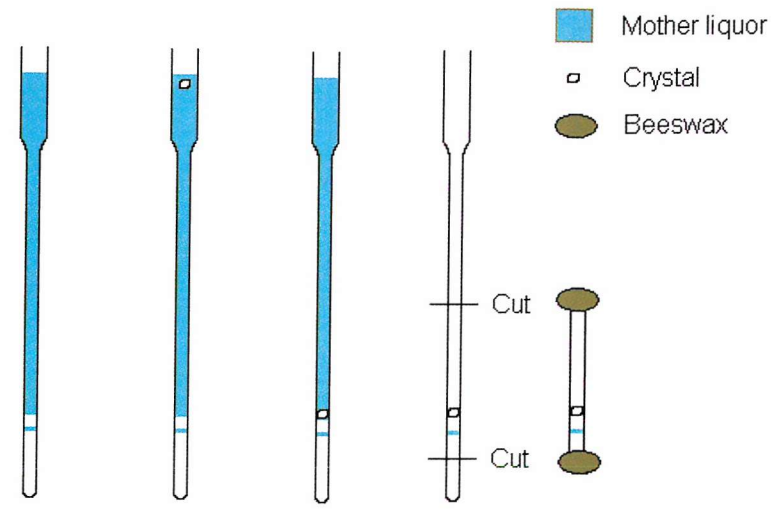


Figure 2.7 Wet mounting of crystals. A Lindemann tube is filled as shown with mother liquor, leaving a small air gap in between the two solvent phases. A crystal is added and allowed to drop to the bottom of the first solvent phase, then as much as possible of the first solvent phase is removed, the Lindemann tube is cut as shown and the ends of the tube are sealed with beeswax, to maintain an enclosed atmosphere.

2.16.4 Cryo-loop assembly

Cryo-loops (Figure 2.8) were used for freezing and transferral of crystals and could be directly mounted for exposure to the X-ray beam and therefore data collection.

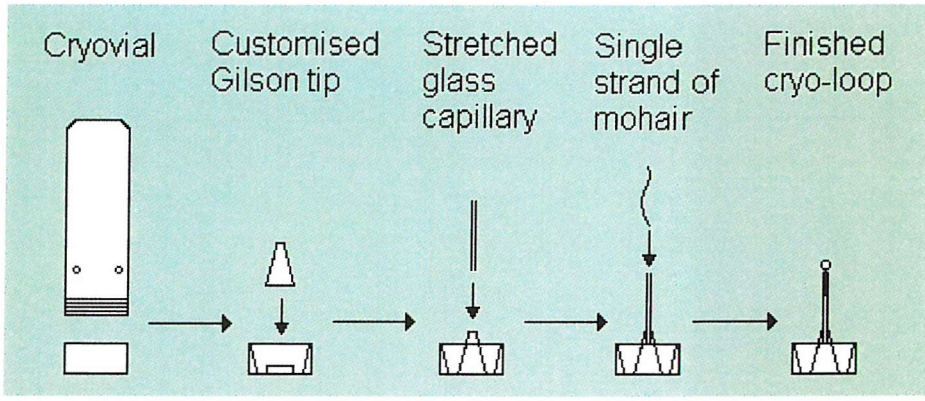


Figure 2.8 Cryo-loop assembly. All cryo-loops were made using superglue as shown above. Mohair was used as it gives minimal background diffraction when data collection takes place.

2.16.5 Freezing of crystals

Crystals were scooped from a drop (containing a chosen cryoprotectant) using a cryo-loop and frozen directly into liquid ethane and then transferred into liquid nitrogen. Frozen crystals were either mounted directly in the X-ray beam, or stored in liquid nitrogen in a Dewar flask. Freezing of crystals is used to minimise radiation damage to the crystal throughout the data collection process (Rhodes, 2000).

2.16.6 Screening of cryoprotectants

Minimum concentrations of cryoprotectants for the freezing of crystals were determined by making up established mother liquor solutions with the addition of varying concentrations of cryoprotectant from 0-30% and these were then frozen on a cryo-loop and diffraction images were gathered for each percentage concentration of cryoprotectant. The minimal concentration at which no ice rings were found was noted for each cryoprotectant. Concurrently similar tests were conducted using a guillotine freezing device (Dr. Coker, Southampton) for freezing the solutions. This was observed to reduce the concentrations required for each cryoprotectant by at least 1% in most cases.

Having discovered the minimum concentrations required for freezing the crystals, protocols for careful introduction of these conditions to the crystals were devised. Two different methods were used for transferring crystals into the desired cryoprotectant conditions:

(a) The first method relied on the same principle as vapour diffusion. The crystal was placed into a drop that contained half the concentration of cryoprotectant required and was inverted over a well containing the full amount of required cryoprotectant. The crystal as then left overnight at room temperature for the well to equilibrate.

(b) The second method required transferring the crystal through incrementally increasing concentrations of cryoprotectant solutions. This was generally achieved through 3% increases of cryoprotectant concentration with 30 minutes of equilibration time for each step.

Both of these protocols were based on the assumption that the crystals were originally suffering from osmotic shock.

2.16.7 Selenomethionine labelled *P. sativum* ALAD

Following the apparent inability to find a molecular replacement solution and obtain phases for the observed data, a protocol for growth and expression of selenomethione-labelled *P. sativum* ALAD was devised, based on a previous protocol by Ramakrishnan *et al.* (1994).

Materials

Minimal Media M9 (20X)

20 x M9 minimal media was prepared by mixing 10g of NH_4Cl, 30g KH_2PO_4 and 68g of Na_2HPO_4 with 500ml of AnalaR water. The solution was autoclaved and stored at room temperature until required.

Amino acid mix 1

100ml of the amino acid mix 1 was prepared by dissolving all the amino acids, except methionine, tyrosine, tryptophan and phenylalanine, at a concentration of 4mg ml^{-1} in AnalaR water. The solution was filter-sterilized and stored at -20°C until required.

Amino acid mix 2

100ml of the amino acid mix 2 was prepared by dissolving tyrosine, tryptophan and phenylalanine at a concentration of 4mg ml^{-1} in AnalaR water. The solution was filter-sterilized and stored at -20°C until required.

Seleno-L-methionine

30ml of seleno-L-methionine solution was prepared by dissolving seleno-L-methionine in 30ml of AnalaR water at a concentration of 10mg ml^{-1}. The solution was made immediately before use.

Vitamins

20ml of the vitamin mix was prepared by dissolving riboflavin, niacinamide and pyridoxine monohydrochloride at a concentration of 1mg ml^{-1} in AnalaR water. The solution was filter-sterilized and stored at -20°C until required.

Method

The final medium (6.4L) was prepared by mixing 12.8mL of 1M $MgSO_4$, 6.4L of 20 x M9 minimal media, 12.8mL of a 12.5mg ml^{-1} solution of $FeSO_4.7H_20$, 64mL of 40% glucose, 64mL of amino acid mix 1, 64ml of amino acid mix 2, 6.4ml of the vitamin solution. The solution was made to 6.4L with the addition of AnalaR water and the pH was checked to ensure it was ~7.4. Finally, ampicillin was added to a final concentration of 100µg ml^{-1}.

Cell strain utilised

So as to minimise leaky expression during the selenomethionine growth, the cell strain *E. coli* BL21(DE3)pLysS was used. This strain encodes small amounts of T7 lysozyme which inhibits basal levels of T7 RNA polymerase prior to induction.

Cell growth and expression

Sample cultures (20mL) expressing the shortened recombinant pea ALAD cDNA construct were grown overnight at 37°C. These sample cultures were then used to inoculate 800mL flasks of Luria broth media (for preparation see Chapter Two) and grown to an OD of 1. The cells were then harvested by centrifugation at 10000 rpm for 12 mins. The supernatant was discarded, and the bacterial pellets were resuspended in M9 minimal media. The resuspended pellets were then used to inoculate 800ml flasks containing the final minimal media mixture, without selenomethionine present. The cells were allowed to grow for 30 minutes, so that the cells would use up any methionine present from their growth in LB media, and then 3.2mL of the selenomethionine solution was added, in addition to 800µl of IPTG stock (for preparation see Chapter Two) to induce expression of the selenomethionine labelled protein. The culture was grown for a further 3 hours after induction, after which the cells were harvested by centrifugation at 12000rpm for 12 mins. The pellets were either used immediately, or stored at -20°C until required.

Seleno-L-methionine labelled protein purification

The labelled protein was purified, using the same protocol used for the unlabelled protein, and eluted in a similar fashion on both the ion-exchange and gel-filtration columns. Samples run on SDS-PAGE gave similar size bands as the unlabelled proteins. Attempts were made to obtain mass spectra for the selenomethionine-labelled protein, however, there was difficulty in attaining a spectrum.

2.17 Data Collection

2.17.1 X-ray sources

Two types of X-ray sources were utilised in the crystallographic studies for the collection of diffraction data.

The first, a rotating anode generator, generates X-rays through bombardment of a metal target (usually copper) with electrons, which are produced by heating of a filament. The electrons are then accelerated by an electric field towards the target (Figure 2.9).

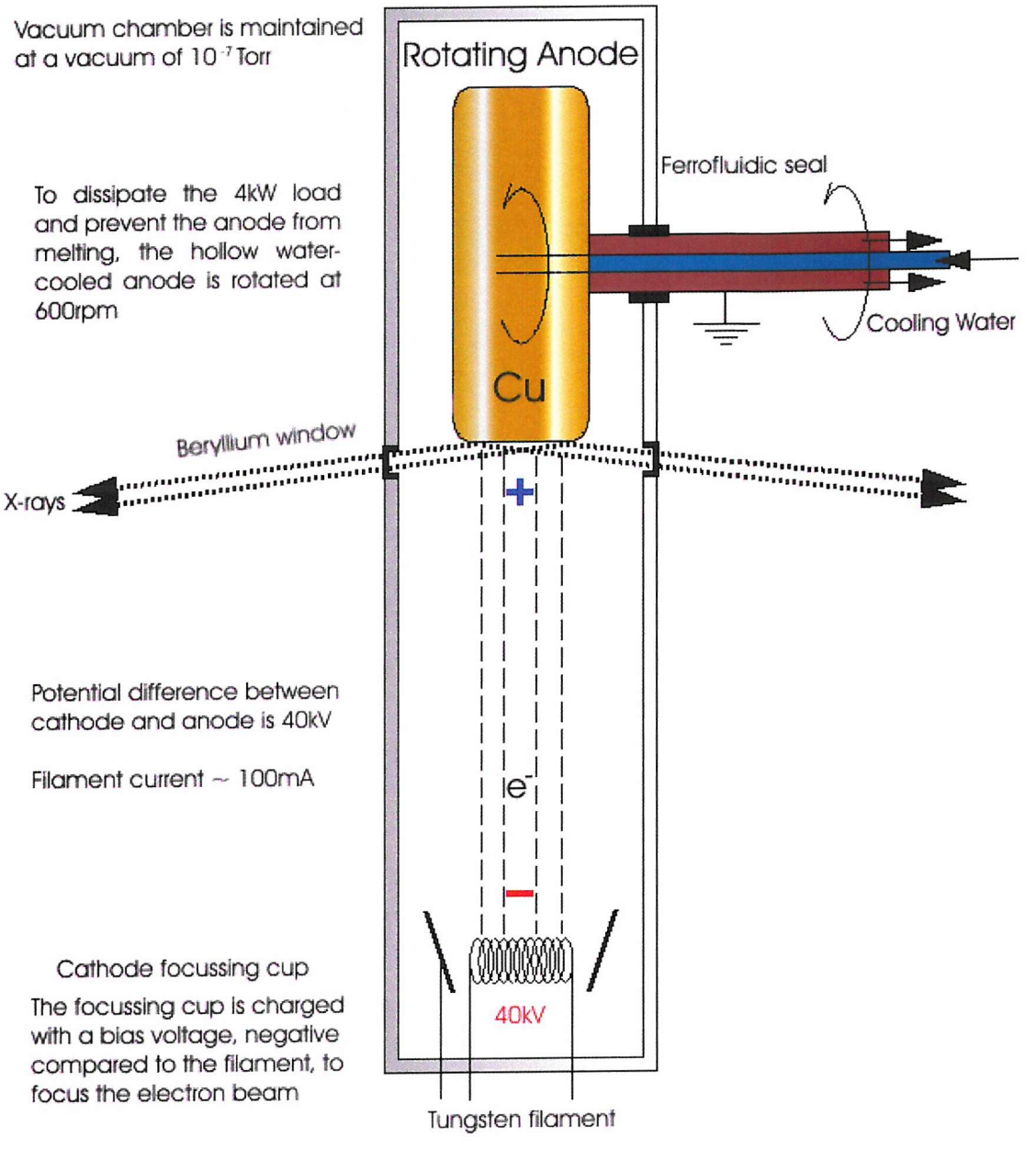


Figure 2.9 A schematic diagram of a rotating anode X-ray generator.

Electrons emitted from a tungsten filament are accelerated towards a rotating copper anode which causes emission of X-rays at a critical angle.

When high energy electrons collide with a metal target atom, as is the case with the rotating anode X-ray generator, electrons from an inner shell are ejected. When this happens an electron from a higher energy orbital drops down a level, and energy is emitted from the electron in the form of an X-ray photon (Figure 2.10).

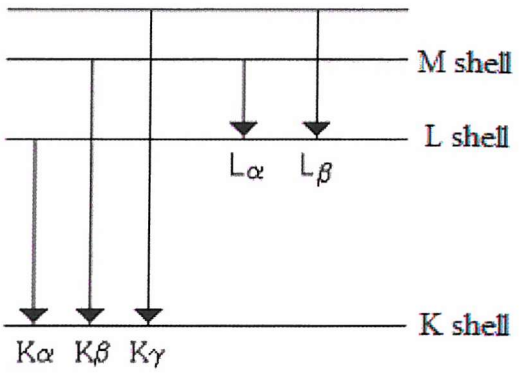


Figure 2.10 The emission of characteristic X-ray spectral lines.

For copper targets, commonly used in X-ray generators for investigating macromolecules, X-rays are emitted at λ=1.54Å (k_α) and λ=1.39Å (k_β). For X-ray diffraction, monochromatic X-rays of a single wavelength are required, and therefore the X-rays emitted from a rotating anode are effectively filtered using a nickel filter. The nickel filter absorbs the copper k_β emissions and allows k_α emissions to pass through and so a monochromatic X-ray at λ=1.54Å (k_α) is emitted (Figure 2.11) (Rhodes, 2000).

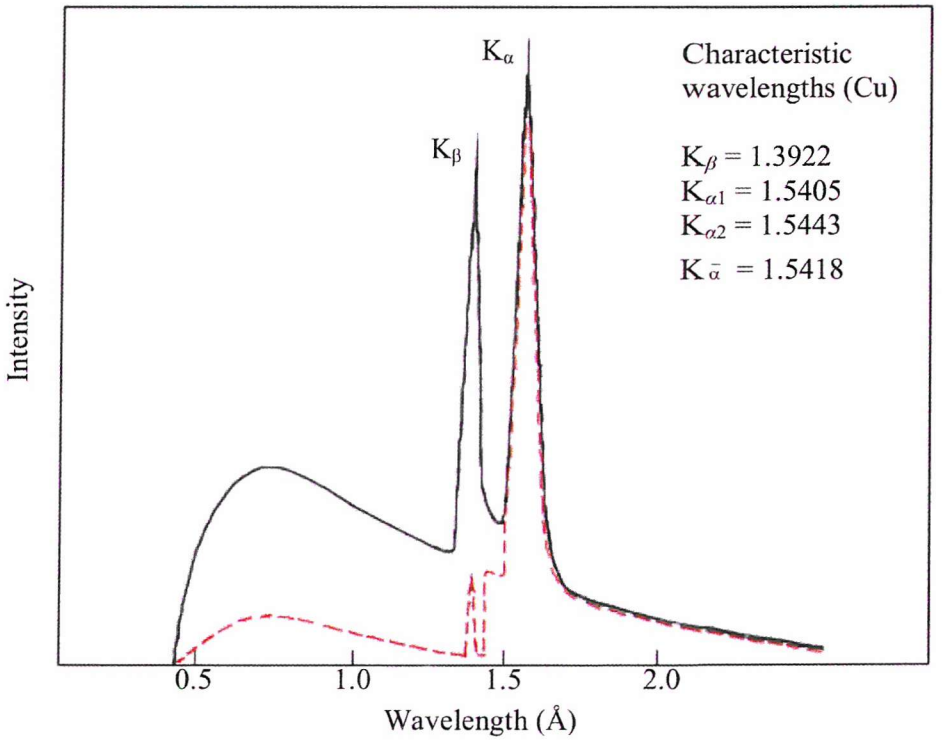


Figure 2.11 The X-ray spectrum emitted by a tube with a copper target before (-) and after (- - -) passage through a nickel filter.

The second type of X-rays used for macromolecular studies is generated by a synchrotron radiation source such as the ESRF Grenoble and provides the most powerful source of X-rays. Electrons or positrons are circulated in a giant ring at velocities near the speed of light and maintained at this speed through the use of powerful magnets. The electrons emit energy (synchrotron radiation) when forced into this circular motion by bending magnets (Figure 2.12). The intensity of synchrotron radiation is much greater than that produced by rotating anode generators, and so therefore shorter exposure times can be used on crystals to collect data (Rhodes, 2000).

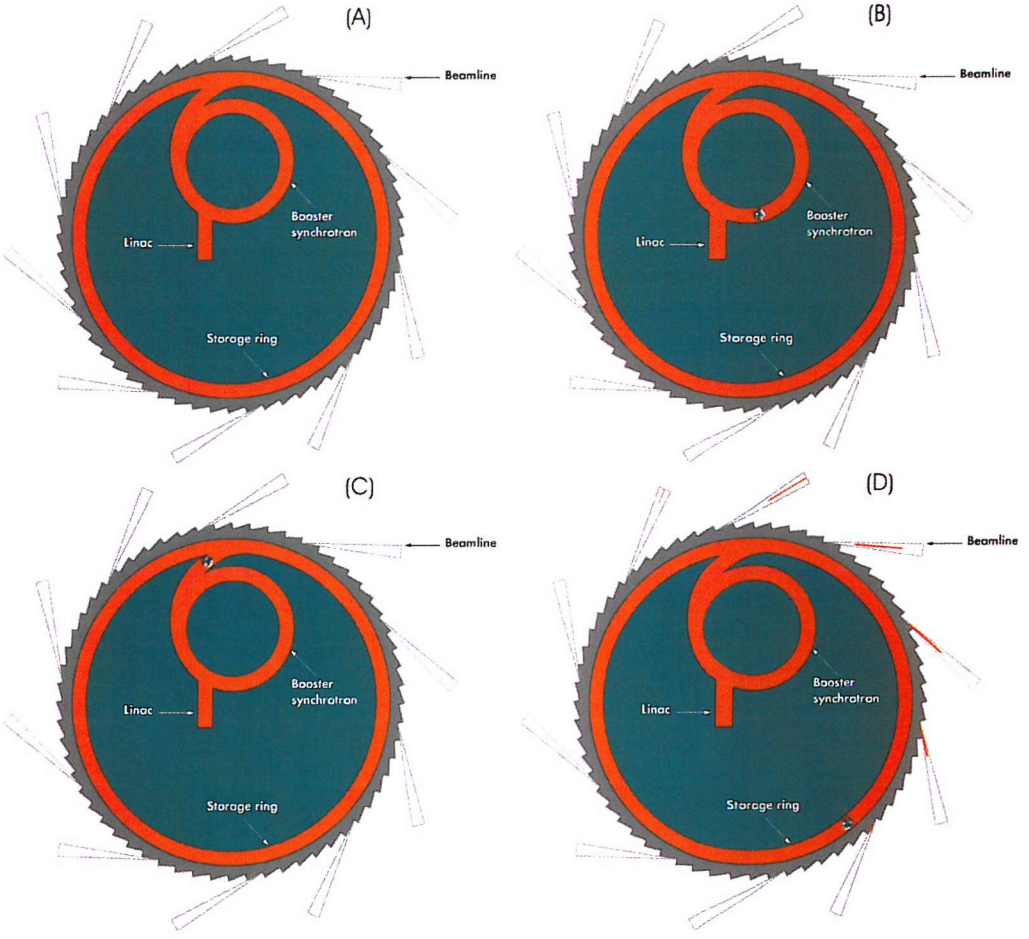


Figure 2.12 A representation of the synchrotron radiation source.

Electrons are accelerated in a booster synchrotron to around the speed of light, before being channelled into the main storage ring. The electrons are then maintained in a circular path by "bender" magnets, which cause the emission of synchrotron radiation. At certain points along the storage ring, beam lines branch off at tangents to the storage ring and this is where data collection is made at designated workstations (Figure reproduced from the ESRF website – http://www.esrf.fr).

2.17.2 Image plates

Image plate scanners are utilised to collect X-ray diffraction data. The most common of these is the MAR scanner, which has a plate radius of 90 or 150mm. The surface of the image plate is composed of barium fluoride doped with europium ions (Eu^{2+}), which are excited upon bombardment with X-rays into a metastable state (Eu^{2+}-Eu^{3+}). The scanner then uses a red He-Ne laser to generate light emission from the europium atoms. The intensity of the emitted light is proportional to the number of absorbed X-rays and can be amplified by a photomultiplier system and measured. Deletion of the image plate is achieved by a high intensity flash of visible light (McRee, 1993).

2.18 Protein crystals

A protein crystal is an orderly three-dimensional array of molecules, held together by noncovalent interactions. The crystal is divided into identical unit cells (Indicated by the line in the diagram). The array of points at the corners or vertices of the unit cell is called the lattice. Unit cells are the smallest and simplest volume elements that are completely representative of the whole crystal. Protein crystals are effectively an efficiently packed array of many unit cells stacked besides and on top of one another.

2.18.1 Space Groups

The contents of the unit cell are symmetric and therefore certain sets of reflections are equivalent. In theory only one set of equivalent reflections need to be measured and therefore knowledge of the unit cell symmetry allows the reduction of data collection time by minimising the chance of recording the same reflection more than once. The symmetry of a unit cell is described by its space group which is represented by a capital letter (referring to the unit cell's bravais lattice type) followed by numbers representing symmetry operations that are applicable to the unit cell that would not alter its appearance (Figure 2.13).

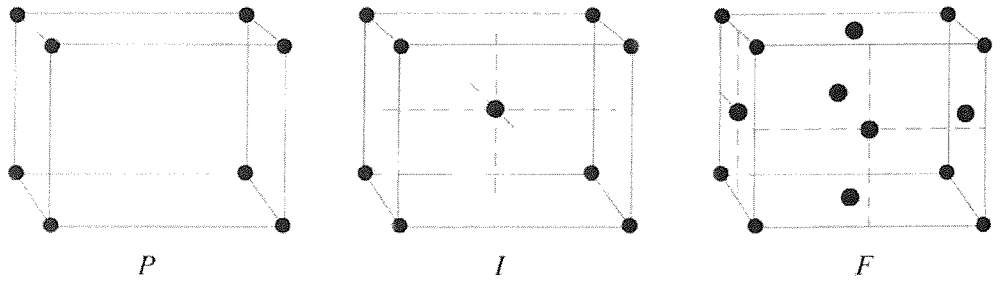


Figure 2.13 Three of the possible fourteen bravais lattice types. Lattice P describes a primitive lattice which contains one lattice point at each corner of the unit cell. Each lattice point is shared between eight neighbouring unit cells and therefore each unit cell contains eight times $1/8^{th}$ of a lattice point. Lattice I represents a body-centred lattice, with a lattice point at the centre of each unit cell and $1/8^{th}$ at each corner and therefore each unit cell contains two lattice points. Lattice F denotes a face centred lattice which has lattice points on the centre of each face, each of which is shared with neighbouring unit cell giving 1/2 of a lattice point at each face with $1/8^{th}$ at each corner, yielding a total of 4 lattice points per unit cell (Picture reproduced from 'Crystallography made crystal clear', Rhodes (2000)).

The unit cell contains a large asymmetric molecule (the protein) which contains no symmetry elements but which can be juxtaposed onto other identical entities by symmetry operations, which describe the contents of every unit cell. These symmetry operations incorporate translation, rotation and screw axis (the combination of translation and rotation). Translational effectively the movement of the molecule along x,y or z by a fraction of the unit cell edge. Rotation axes are represented by an integer n e.g. a 4 fold rotation generates the same structure after (360/n) or 90 degree rotation. The screw axis results from a combination of both translational and rotational elements n_m where n is an n fold screw axis with a translation of m/n of the unit translation. For example a 2_1 axis consists of a rotation of 180 degrees about the 2_1 axis which is on or parallel to one of the unit cell edges. This is then followed by a translation or movement along the axis parallel to the 2_1 axis equal to half the length of the unit cell edge.

2.18.2 Bragg's Law

Bragg's law refers to the simple equation $n\lambda=2d \sin\theta$ derived by W.H. Bragg and W.L. Bragg in 1913 to explain why the cleavage faces of crystals appear to reflect X-ray beams at certain angles of incidence (θ). The variable *d* is the distance between atomic layers in a

crystal, and the variable λ is the wavelength of the incident X-ray beam. This observation is an example of X-ray wave interference commonly known as X-ray diffraction (Figure 2.14).

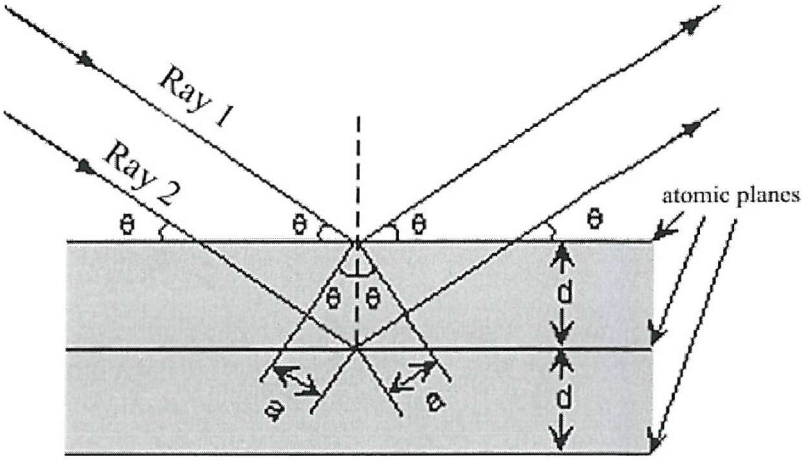


Figure 2.14 A figure showing diffraction from Bragg planes separated by a spacing, d.
Constructive interference occurs only if the difference in path length (2a) is a multiple of the radiation wavelength.

If the difference in path length for rays reflected from successive planes is equal to an integral number of wavelengths of the impinging X-rays then rays reflected from successive planes reinforce each other and produce a strong reflection. Theta (θ) is the angle of diffraction and its sine is inversely related to interplanar spacing d_{hkl}. Each set of parallel planes in the crystal produces one reflection.

The data gained through crystallography is given by the diffraction of the X-rays by the electron clouds that surround the atoms in the structure that has been crystallised.

2.18.3 Diffraction theory

Upon collection of diffraction data, the next step is the interpretation of the images collected so as to obtain a 3D picture of the electron density and therefore the atoms in the crystal. The process by which this is undertaken is made up of three steps.

The first is the determination of the structure factor amplitudes, F_{hkl}, of all of the X-ray beams that have given rise to the spots observed on the diffraction plates.

The second is the solving of the phase for each of the diffracted X-ray beams so as to determine the positions of the waves with respect to each other.

The final step is the production of an electron density map and refinement of the initial phases determined in the second stage. This will improve the electron density map and yield the three-dimensional structure of the protein. The overall process can be summed up in a single equation:

Equation 1: $$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}\ e^{-2\pi i(hx+ky+lz)}$$

From the equation, electron density at any particular point (xyz) can be calculated by constructing a Fourier series using structure factors F_{hkl}. F_{hkl} is a periodic function describing a complicated wave and therefore consists of a frequency, amplitude and phase. The frequency is equal to that of the X-ray source since it is describing a diffracted X-ray. The amplitude can be calculated directly as it is directly proportional to the square root of the intensity of the diffraction spot (I_{hkl}). However, the phase is unknown but is essential to calculate the electron density map and therefore the position of the atoms in the protein molecule. This is called the 'phase problem' and can be overcome by various methods including isomorphous replacement, molecular replacement and multiple-wavelength anomalous dispersion (MAD).

2.19 Computational methods for the analysis of diffraction data

All data processing was conducted using the CCP4 suite of programs (Collaborative Computational Project, 1994).

2.19.1 MOSFLM

The first set of data processing routines are autoindexing, mosaicity estimation, spot prediction, cell parameter refinement and integration of images (Collaborative Computational Project, 1994; Leslie, 1999).

- *Autoindexing of images*

Through this process, MOSFLM assigns a three digit number comprised of the integers h, k and l to each reflection. These are known as Miller integers and represent coordinates that describe the position of each reflection recorded in a 3D reciprocal-space lattice. The term reciprocal is used because there is an inverse relationship between the lattice spacing in the crystalline lattice (real unit cell) and the spacing in the lattice of the diffraction pattern. Autoindexing gives a predicted spot position (hkl) and intensity (I_{hkl}) for each spot in the reciprocal-space lattice.

Single images throughout a data set are opened for analysis. Spots with a certain, preferably higher, signal-to-noise ratio are selected by the programme through use of the `find spots` command. This process can be performed using spots from a single image or from multiple images.

Autoindexing then uses the selected spots to estimate potential cell parameters, crystal orientations and also gives a putative Bravais lattice. To each potential solution, the program assigns a penalty which is based upon the fit of the 14 lattice characters to the primitive triclinic solution and is normalised to give a maximum penalty of 999. The program shows all solutions with a penalty less than 200. In general, better solutions can be obtained by altering the signal-to-noise ratio range in which spots are analysed and therefore picking and indexing more, or less, spots to be autoindexed. Solutions with the highest symmetry after a large drop in penalty are generally chosen as candidates. Once a solution is picked, MOSFLM refines the solution and applies a user-input space group, while also refining the main beam position.

- *Mosaicity estimation*

Mosaicity is the spreading of spots recorded, and can be caused by imperfections in the crystal lattice and generally increase over the course of an experiment due to crystal damage from exposure to the X-ray beam. MOSFLM has a built-in function for mosaicity estimation. It is best to check the mosaicity for a few images spread throughout the data set. The average value is normally taken for further processing steps as an initial estimate.

- *Prediction of spots*

When all of the parameters are available to MOSFLM, the program is able to predict where spots will be found given the entered parameters. MOSFLM will display a series of prediction boxes over an image which should overlap observed spots. If these predictions do not match the observed data fairly accurately it is possible that incorrect parameters have been chosen. The mosaicity estimate can also be checked at this point, although it can be expected to increase over the data set.

- *Cell refinement*

This process is run to obtain the best possible fit of calculated to observed spot position before integration. In other words, to gain accurate cell parameters before integrating spot intensities. Post refinement uses the phi centroids of reflections which are spread across adjacent images to refine the cell parameters, crystal mis-setting angles and mosaic spread independently of camera parameters (e.g. crystal to detector distance, Y scale (the relative scale factor in the detector Y direction) and tilt and twist of the detector).

A minimum of two adjacent images are required, with a sufficiently large number of completely recorded partials for good statistics, depending on the mosaicity of the images across the data set. If higher mosaicity is observed, then more images will be required. Also, it is best to use at least two segments of data separated by ~90°.

All of the above tasks are generally run interactively through the MOSFLM graphical user interface.

- *Integration of images*

The following procedure is run as a batch job using a purpose-built input file, containing all of the relevant information about the integration parameters required. This fully integrates all of the images and outputs a reflection (.mtz) file, containing spot positions and intensities.

After the batch job has completed there are three files that must be checked for various values:

1. The `SUMMARY` file can be used in conjunction with `loggraph` to view graphically the parameters throughout the data set that have varied over the course of the data processing, thus far. Each graph should be checked to make sure there are no sharp discontinuities and that all graphs vary smoothly through the processing run. It is especially important that `Y scale` and detector distance do not vary greatly over the processing run as a large random variation in either indicates that post-refinement was not performed correctly and, therefore, one or more parameters could be unreliable.

The residuals (RESID) and weighted residuals (WRESID) for each image should remain constant. These values describe the quality of the spot prediction. Values for residuals should remain at ~0.1mm or below, and values for weighted residuals should remain at ~1.0 or below.

Also I/sigma(I) graphs should be checked for crystal degradation over time and therefore the highest achievable resolution for the later images of the data set.

2. The end of the `Mosflm.lp` file should be checked for `warning`, `error` and `severe error` messages. `Warnings` generally show that MOSFLM has detected a parameter that is not optimal and has made an attempt to correct it, and these can generally be ignored. `Errors` indicate where something has gone wrong and MOSFLM has been unable to correct it. In most cases the problem can be solved simply be following the recommendations in the message associated with the error. `Severe errors` or `fatal errors` indicate that a problem has occurred so as to cause termination of the integration run. Again this can sometimes be fixed by following the recommendations of the associated message. However, in extreme cases these errors may only be fixed by contacting the program authors.

It should also be checked that there were not excessive numbers of bad spots determined. The final 'rms residual' for each integrated image should also be checked. As in the 'SUMMARY' file it should remain at ~1.0. Again I/sigma(I) should be checked and should be at least 1.0 in the highest resolution shell of data.

If any of these factors do not fall within the boundaries set, integration must be rerun until their values are within recognised ranges.

3. The final `.mtz` output file can be checked and altered using `MTZDUMP`. The main thing to check is that it contains the correct number of batches (images) and also to check that the experimental details are correct i.e. wavelength, resolution limits, space group etc.

2.19.2 SORTMTZ

Sortmtz is now used to organise the symmetry related reflection list in the `.mtz` file output from MOSFLM on the basis of h, k and l (Collaborative Computational Project, 1994).

2.19.3 SCALA

Scaling and merging is the next step in data processing. This attempts to put all symmetry related reflections (known as multiplicity) on a common scale and also provides the main diagnostics of data quality. Scaling is essential as it smoothes out fluctuations between one image and the next. Various physical factors lead to observed intensities being on a different scale. Among these factors are radiation damage to the crystal, beam intensity fluctuations and variations in the crystal thickness during the data collection. Scales are determined by comparison of symmetry-related reflections i.e. by adjusting scale factors to get the best possible internal consistency of intensities (Collaborative Computational Project, 1994).

Once the programme has been run, the following files should be checked:

1. `ROGUES` should be checked for those reflections that have been rejected because the symmetry equivalents differ considerably and also those reflections where Emin (the normalised structure factor amplitude) is too large.

2. In the `Scala.log` file the values of Nbias, Rmeas, I/sigma(I), multiplicity and completeness should all be checked for consistency and should all adhere to certain values (Table 2.6). Rmeas gives a good, stable measurement describing the internal consistency of the input data set and gives higher and more accurate values than R-merge, which has classically been used as an indicator of dataset consistency.

Factor	Recognised value
Bias	~0.001 or below
Rmeas	~0.1 or below
I/sigma(I)	~2.0 or above
Multiplicity	~2.0 or above
Completeness	~80% or above

Table 2.6 Recognised values for acceptable data processing of crystallographic data

Also within the input file it has been suggested that manipulation of the SDadd factor above 0.03 in order to balance standard deviation values to ~1.0 for full and partial reflections could show incorrect data processing, however this is not definitive.

2.19.4 TRUNCATE

This program generates structure factor amplitudes from spot intensities.

Each reflection that hits the detector can be described as a complex wave, decribed by a Fourier series composed of Fourier terms. Each term is a simple sine or cosine function whose wavelength is an integral fraction of the wavelength. The Fourier series that describes a diffracted X-ray can be described by a structure factor equation and the computed sum of the series for reflection *hkl* is called the structure factor amplitude (F_{hkl}). Each term within F_{hkl} describes the contribution of one atom to the reflection hkl and can be expressed as:

Equation 2: $F_{hkl} = \sum f_j \; e^{2\pi i(hxj+kyj+lzj)}$

The contribution of atom j to F_{hkl} depends on the element (i.e. its atomic structure factor (f_j)) and on its position within the unit cell (x_j, y_j, z_j). which establishes the phase of its contribution.

In the log file the Wilson plot can be checked to roughly assess the success of the scaling. In addition, an estimate of solvent/protein content of the asymmetric unit can be obtained.

Systematic absences

Certain symmetry related elements in the unit cell can cause specific reflections to be missing, giving recorded intensities equal to zero. For example, a two fold screw axis (2_1) along the c edge causes all $00l$ reflections having odd values of l to be missing. These systematic absences can give a good insight as to whether the space group chosen for processing of the data has been chosen correctly, by checking the processed data against the expected systematic absences associated with a certain space group. These can be checked utilising the CCP4 program HKLVIEW (Phil Evans, MRC LMB, Cambridge) that can actively display certain zones of reciprocal space ascertained from the processed data.

2.19.5 Molecular replacement

So far in the processing, amplitudes have been found from the diffraction spots. However, to develop an electron density map, phases for the diffraction spots are also needed. These can be found either by using molecular replacement, direct methods, isomorphous replacement or anomalous dispersion. In the case of this study two molecular replacement methods were used.

Since the structures of several ALAD enzymes were known, molecular replacement was attempted. Four programs associated with or incorporated into the CCP4 suite were used: AMORE (Navaza, 1994), BEAST (Read, 2001), MOLREP (Vagin, 1997) and XPLOR (Brunger, 1990). Each of these programs essentially does the same job; finding the phases from the experimental data. The programs differ in the methods and therefore algorithms they use to determine the phases.

Molecular replacement determines the orientation and positions of a molecule in the unit cell using a previously solved structure as a search model (Figure 2.15). The search and target molecules must have reasonable sequence identity (>25%) for there to be a good chance of success. Similarly, having a data set with good completeness can be crucial. There are two factors each program determines through molecular replacement, known as the rotation and translation functions.

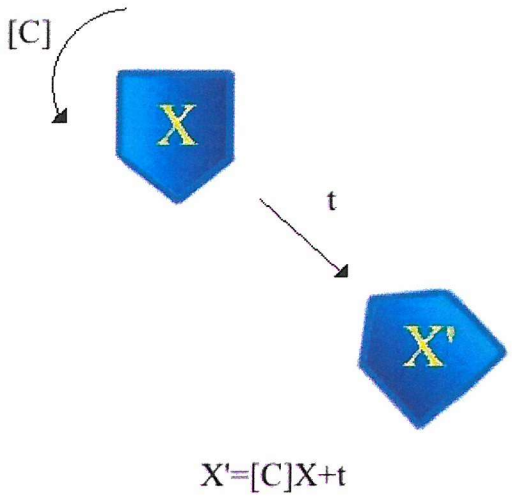


Figure 2.15 An illustration of molecular replacement methods.

Molecular replacement is dependant upon superimposing the known structure (X) on the target structure (X') in its unit cell. The transformation of X to X' is described by the equation X' = [C]X+t where C is the rotational matrix that rotates the co-ordinates of X into a new orientation and t is the translational factor defining the new position. The success of molecular replacement is dependant upon the accuracy and indeed the ability to determine 3 orientational parameters (rotation function) and 3 translational parameters (translation function).

The rotation function is defined by three parameters called Eulerian angles (α, β and γ) and the translation function by a further three translational parameters (t_x, t_y and t_z). This makes a total of six parameters to search for to complete a molecular replacement. In theory it is possible to search for all six parameters in parallel, but in practice, this would take an incredible amount of time, even with powerful computers. An alternative method was devised by Rossman and Blow (1962). They demonstrated that the search for the orientation of similar subunits in a unit cell could be divided into two parts by using the Patterson function, namely the rotation and translation functions.

The Patterson function

The Patterson function is the Fourier transform of the experimental intensities and, as such, requires no phase information. The Patterson map can be derived from the equation:

Equation 3: $$P_{(u,v,w)} = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \; e^{-2\pi i(hu+kv+lw)}$$

The Patterson map (for an example see Figure 2.16) calculated using this equation consists of peaks corresponding to all of the interatomic vectors. Each pair of atoms in the real unit cell gives rise to two peaks in the Patterson cell (i.e. for N atoms in the real unit cell there would be N^2 in the Patterson cell). The Patterson cell also contains an origin peak which describes vectors of zero length, corresponding to vectors between the atom and itself. Therefore, the Patterson unit cell consists of N^2-N non-origin peaks. It is plain from this that the higher the number of atoms in the real cell, the higher the number of vectors to be assigned and therefore the more densely packed the Patterson map becomes and the harder interpretation becomes.

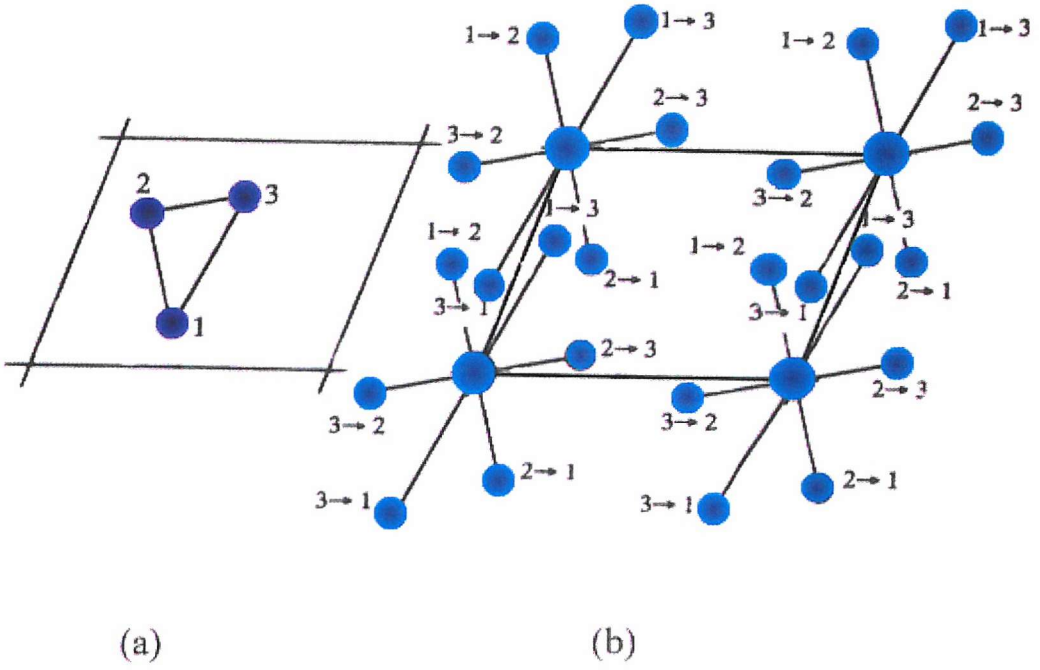


Figure 2.16 An example of a Patterson map (b) derived from a simple two-dimensional real cell containing three atoms (a).

The Rotation and Translation functions

The rotation function is evaluated either in real (Patterson) or reciprocal space with the difference between the programs used being the type of algorithms employed. With the Patterson map from a known structure the intramolecular vectors are independent of the position of the structure in the unit cell as long as no rotation occurs. When rotation is applied to the model, the arrangement of the intramolecular vectors changes and this is how the model is utilised so as to match the orientation of the target model. The purpose of the

rotation function is to search for matches between the Patterson map, or its reciprocal-space analogue, for the observed intensities of the target model. The results of rotation searches are normally evaluated in a search grid in Eulerian or polar angles covering the entire region of rotational interest. Grid points that coincide with the correct angular parameters are generally expected to give high rotation function values. The rotation function at other grid points can be used to estimate background noise.

Once an acceptable rotation function has been found, the determination of the translation function can begin. The translation function is used to determine the position of the search model with respect to the undetermined target structure. This search relies on the cross-vectors in the Patterson function and, in effect, the search model is moved around the unit cell and the cross-vectors are calculated and compared with the native Patterson map.

The programs used, generally output a large volume of peaks that correspond to predicted rotation vectors and these are evaluated by their correlation coefficient (an indication of the matches between the models which is independent of phase) and the free R factor (an indication of the level of disagreement between the models).

2.19.6 Refinement and Model Building

Refinement is the process by which the improvement of the agreement between the calculated and the observed structure factor amplitudes is undertaken. This method is generally referred to as reciprocal space refinement and requires the adjustment of atomic positions (x, y, z) to values that coincide with the observed structure factor amplitudes. Least square refinement is undertaken by selecting atom positions that minimise the squares of difference between corresponding calculated and observed structure factor amplitudes and this process is expressed by equation 4:

Equation 4: $\phi = \sum W_{hkl}(|F_o| - |F_c|)^2$

Where φ = sum of the squared differences between observed and calculated structure factor amplitudes. W_{hkl} = weighting of the difference which is dependant upon the reliability of the measured intensity.

The parameters refined throughout the process include atomic positions, temperature factors (B-factors) and occupancy. A major problem associated with refining protein molecules is the large numbers of atoms present in the structure resulting in poor observation-to-parameter ratios. Therefore, constraints or restraints are used in the refinement to overcome this problem. With constrained refinement, a fixed value is used for a certain parameter, e.g. to fix bond lengths and angles, allowing only dihedral angles to vary and therefore decreasing the number of parameters to be refined. Restrained refinement, in contrast, imposes elastic conditions upon the parameters, allowing a variation of a certain amount for bond length and angle. There is a variety of programs capable of refinement and in this study SHELX (Sheldrick, 1998) and RESTRAIN (Driessen *et al.*, 1989) were used.

Electron density maps and model building

As well as computational refinement, manual model building is required during refinement. To be able to carry out manual model building, there is a requirement for electron density maps so that the model can be rebuilt into density where needed. Two types of electron density maps are normally calculated during model building. Maps are calculated using F_o-F_c as the coefficient and these provide information about the new structure. Maps are calculated using $2F_o$-F_c and these give an approximation of the electron density of the model and a difference map can be obtained. The phases that are initially rough estimates and the corresponding electron density maps can be relatively uninformative. The map can be improved through the use of any discernable unit cell contents to calculate new structure factors during the course of refinement and the phases of these structure factors are used along with the original native amplitudes to recalculate the maps. This can be a cyclic process by which the model is manually rebuilt. The model is then submitted for computational refinement and new maps can be built until the model becomes more accurate. The computational refinement program conducts minute adjustment to the atomic positions to improve the agreement between the amplitudes calculated from the current model and the original amplitudes measured in the native data set. In order to visualise the maps and model for model building the program TURBO-FRODO (Jones, 1978) was used.

Chapter Three: Characterisation of a recombinant *Pisum sativum* 5-aminolaevulinic acid dehydratase (ALAD)

3.1 Introduction

Previous work within the Shoolingin-Jordan group (Southampton University) has yielded purified ALAD from yeast, *E. coli*, human erythrocytes and *P. sativum*. Previous studies on *P. sativum* ALAD (Cheung *et al.*, 1997) were conducted on a recombinant form of the enzyme which included a predicted chloroplast targeting sequence of ~70 amino acids in length attached to the *N*-terminal end of the protein (Figure 3.1). This form of the enzyme had been shown to have a high specific activity of 280 μmol mg^{-1} hr^{-1} and was also shown to require Mg^{2+} for activity, although activity could be demonstrated using Mn^{2+} or Co^{2+}. Activity, however, was not supported by Zn^{2+}.

```
         10         20         30         40         50         60
          |          |          |          |          |          |
HTFVDLKSPF TLSNYLSFSS SKRRQPPSLF TVRASDSDFE AAVVAGKVPE APPVPPTPAS

         70         80         90        100        110        120
          |          |          |          |          |          |
PAGTPVVPSL PIQRRPRRNR RSPALRSAFQ ETTLSPANFV YPLFIHEGEE DTPIGAMPGC

        130        140        150        160        170        180
          |          |          |          |          |          |
YRLGWRHGLL EEVAKARDVG VNSVVLFPKI PDALKTPTGD EAYNEDGLVP RSIRLLKDKY

        190        200        210        220        230        240
          |          |          |          |          |          |
PDLIIYTDVA LDPYSSDGHD GIVREDGVIM NDETVHQLCK QAVAQARAGA DVVSPSDMMD

        250        260        270        280        290        300
          |          |          |          |          |          |
GRVGAMRVAL DAEGFQHVSI MSYTAKYASS FYGPFREALD SNPRFGDKKT YQMNPANYRE

        310        320        330        340        350        360
          |          |          |          |          |          |
ALTEMREDES EGADILLVKP GLPYLDIIRL LRDNSPLPIA AYQVSGEYSM IKAGGALKMI

        370        380        390
          |          |          |
DEEKVMMESL LCLRRAGADI ILTYFALQAA RTLCGEKR
```

Figure 3.1 The protein sequence encoded by the full chloroplast precursor ALAD gene. The proposed chloroplast localisation sequence is shown in green, with the proposed mature enzyme N-terminus originating directly after the residues in green. The metal binding domain is shown in blue and the active site lysines in pink.

As with previous studies on the human and *E. coli* enzymes, it was demonstrated that the substrate ALA forms a Schiff base with the P-site of the plant enzyme, indicating that catalysis occurs in much the same way as in ALADs that utilise Zn^{2+}. It was also established that the plant enzyme forms an octamer, as expected from studies with the human and *E. coli* enzymes.

This chapter describes the synthesis of a recombinant *P. sativum* ALAD without the predicted chloroplast targetting sequence (Figure 3.1). An effective and reproducible purification procedure was developed to provide sufficient quantities of recombinant *P. sativum* ALAD for characterisation and crystallisation. The purification typically yielded 7.25mg of purified enzyme per litre of growth media.

Studies on the optimal temperature of recombinant *P. sativum* ALAD and *E. coli* ALAD revealed temperature optima of 25°C and 48°C, respectively - a marked difference considering the high similarity (66%) between the two species of ALAD. Studies of the pH optima of *P. sativum* have shown the enzyme to be most active at pH 9.3 and the assay conditions were reflected this accordingly.

In addition, the chemical dimerisation of the substrate, ALA, to form a dihydropyrazine has been investigated under a variety of pH values in a range of biological buffers.

Stability studies on the Mg^{2+}-dependent ALADs have revealed a susceptibility to oxidation of thiol groups or protein aggregation via inter-octameric disulphide bridge formation, in contrast to what has previously been reported for Mg^{2+} dependant ALADs. Monitoring of the protein activity loss by native PAGE has also implied possible breakdown in quaternary structure to form tetramers or dimers during the course of protein degradation.

3.2 Synthesis of a recombinant *P. sativum* ALAD cDNA

A recombinant *P. sativum* ALAD cDNA was synthesised by Dr M Sarwar (Southampton University). The double stranded coding region of 987 base pairs, encoding the 329 amino acids of pea ALAD was constructed from a total of 18 oligonucleotides, 9 for each strand. The primers had an overlap of approximately 10 base pairs. Only the codon specifying Pro

3 was changed, from CCC to CCG, as CCC is known to be a weakly expressing proline codon in *E. coli*. The 5'-end of the cDNA was engineered with an *Nde*I restriction site and the 3'-end with a *Bam*HI restriction site. These restriction sites enabled insertion of the cDNA into the required plasmid vector for expression in a bacterial host. Once synthesised, the recombinant pea ALAD cDNA was cloned into the pT7-7 vector (Tabor, 1985). The pT7-7 vector contains the T7 RNA polymerase promoter. T7 RNA polymerase is a very active enzyme that synthesises its complimentary RNA several times faster than that of *E. coli* RNA polymerase and is incorporated into the genome of genetically engineered bacteria used for protein overexpression.

After cloning, the cDNA insert was sent for sequencing by Oswell (Southampton, UK) (Figure 3.2). The sequence of the plasmid DNA showed the correct insert had been synthesised in the correct position (Figure 3.1).

```
  1  AACGANCGNA GTTACTCTNN TGAAAAACCC AAAAAANGTT TTACTATAAN AAGGAGATAT ACATATGCCG ATCCAGCGTC GTCCGCGCCG TANCCGCCGC  100
101  TCTCCCCGCAC TTAGATCGGC GTTTCAGGAA ACAACTCTAT CACCTGCGAA TTTTGTATAT CCACTGTTTA TACATGAAGG TGAGGAGGAT ACCCCTATTG  200
201  GGGCTATGCC TGGATGCTAC AGGCTTGGTT GGAGACACGG ACTTTTGGAA GAGGTTGCAA AAGCACGGGA TGTTGGTGTT AACAGTGTTG TGCTCTTCCC  300
301  CAAAATTCCA GATGCTTTGA AGACCCCCAC AGGAGATGAA GCATATAATG AGGATGGTTT AGTGCCTCGG TCTATACGAT TGCTCAAGGA TAAGTACCCT  400
401  GATCTTATTA TTTACACAGA TGTTGCATTA GATCCTTATT CGTCAGATGG ACATGATGGC ATAGTCAGAG AAGATGGAGT TATTATGAAT GATGAGACTG  500
501  TTCATCAGCT TTGTAAACAA GCTGTAGCCC AGGCACGGGC TGGAGCAGAT GTTGTCAGCC CCAGCGATAT GATGGATGGT CGGGTAGGAG CGATGCGAGT  600
601  AGCTCTCGAT GCTTNAAGGC TNTCAGCATG TTTCTATTAT GTCATATACT GCAAAGTATG CAAGCTCATT TTATGGCCCA TTTAGAGAAG CATTGGATTC  700
701  AAACCCCCGG TTTGNAGACA AGAAGACCTA CCAGATGAAA TCANCAAATN CAGAGAGGNN TTTACTGANA NGANGGAAGA TGAGTCTGAN GGAGCNTGAN  800
801  ATNTGTTGGT AAAAACCCGC CTTCCTATTT GGANNTNATA AGGGTACTTA NGAGATAATT CTCCACTGCN AATGCGCATA CCAGGTTTTT GGNGAANATN  900
901  NANTNATAAA AACCNGTGGG GCTNTAAAAA TGANTGACAA AANNAAGGTA TAA                                                    1000
```

Figure 3.2 The sequence derived from the sequencing map of the plasmid DNA analysed by Oswell (Southampton, UK). The start of the reading frame is indicated by a green bar above the start codon. The termination codon has a red bar above the codon. Bases not accurately determined by the computer software (N) were double checked with the sequencing map (Figure 3.3). The entire coding region consisted of 987 bases corresponding to a protein of 329AA in length.

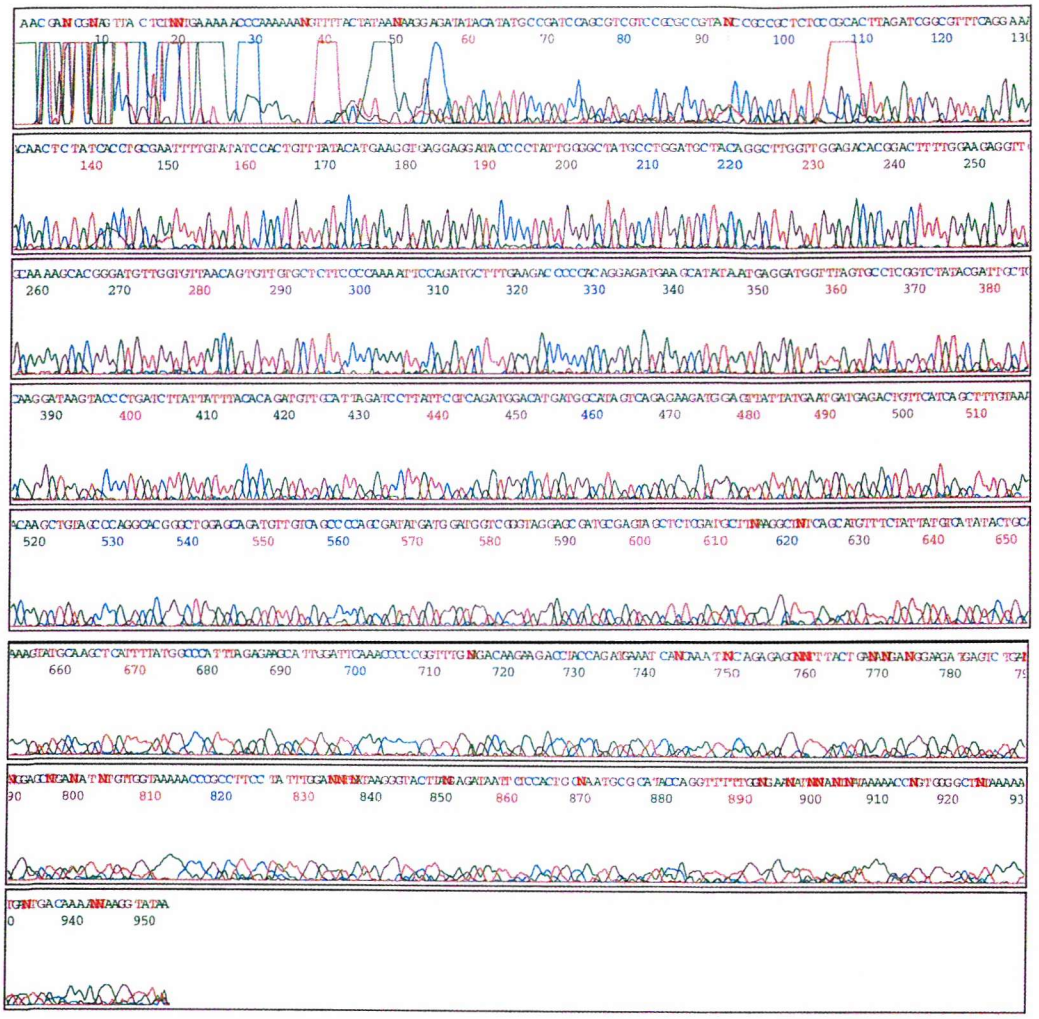


Figure 3.3 The sequencing map showing analysis of plasmid DNA specifying the shortened recombinant *P. sativum* cDNA derived from the *hem2* gene. The computer-generated printout (Figure 3.2) was double checked against this sequencing map for errors and to resolve undetermined bases (Bases are predicted from the peak colours present: red = T, green = A, blue = C and black = G).

3.3 Purification of shortened recombinant *P. sativum* ALAD

3.3.1 Initial protein expression studies

A number of colonies harbouring the recombinant plasmids were selected for characterisation by SDS PAGE and activity assays. Cultures were grown for each clone, which was sonicated and centrifuged to give supernatant and pellet fractions. SDS PAGE analysis indicated that the required protein was soluble and that sufficient protein could be obtained to proceed with the enzyme purification. Large scale (800mL) cultures were

therefore grown following the confirmation of soluble protein expression. Bacterial pellets from this large-scale growth could be stored at -20°C without any significant loss of enzyme activity. In order to characterise and crystallise the recombinant pea ALAD, an effective and reproducible purification procedure was developed.

3.3.2 Cell growth

Sample cultures (20mL) of the shortened pea ALAD gene construct were grown overnight at 37°C. These sample cultures were then used to inoculate 800mL flasks of Luria broth media and grown to an OD of 1. The cells were then induced using IPTG and incubated at 15°C for a further 48 hours. The cells were then centrifuged at 6,000 x g for 12 mins and the pellets stored at –20°C.

3.3.3 Sonication and heat treatment

Frozen cells were thawed and resuspended in 50mM Tris/HCl buffer, pH 8.0, containing 10mM $MgCl_2$, 10mM β-mercaptoethanol and 10mM PMSF. The cells were then sonicated and a heat treatment was performed (50°C for 3mins). The resultant sample was then centrifuged for 30 mins at 10,000 x g.

3.3.4 DEAE ion exchange chromatography

The above supernatant was then loaded directly onto a DEAE ion exchange column, equilibrated with 50mM Tris/HCl buffer, pH 8.2, and the fractions were eluted in a linear gradient from 0-500mM NaCl dissolved in the aforementioned buffer in a total volume 500mL.

3.3.5 Sephacryl S-300 gel filtration chromatography

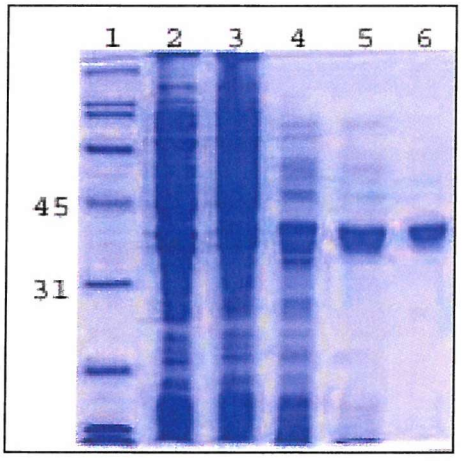
The fractions with the highest specific activity were pooled and then loaded onto an Sephacryl S-300 gel filtration column, equilibrated in 50mM Tris HCl buffer, pH 8.4, and the fractions were collected. The samples with the highest specific activity were pooled.

3.3.6 Protein purification overview

Stage of Purification	Volume (mL)	Total Protein (mg)	Total Activity	Specific Activity (µmol/mg/ml)	% Yield
Sonication	60	812.3	1973.4	2.4	100
Heat treatment	60	104.6	1214.2	11.6	62
DEAE	58	20.6	464.7	22.5	23
S-300	45	12.5	397.9	31.8	20

Table 3.1: Analysis of purification of pea ALAD through specific activity and yield at different stage of the purification.

The specific activity of the pure recombinant *P. sativum* enzyme is a great deal lower than that determined for the pure wildtype *P. sativum* ALAD with the chloroplast targeting sequence present (~240µmol mg^{-1} hr^{-1}). The maximum specific activity recorded for the purest fractions was ~70µmol mg^{-1} hr^{-1}. In comparison to the human (Mills-Davies, 2000; Butler, 2003) and *E. coli* (Norton, 1999) this figure is still relatively high, where the maximum specific activities recorded were ~45µmol mg^{-1} hr^{-1} and ~35µmol mg^{-1} hr^{-1} respectively.



Overall purification:

1. BioRad marker
2. Cells
3. Sonication
4. Heat treatment
5. DEAE column
6. Gel filtration

Figure 3.4 Gel showing purification stages of recombinant *P. sativum* ALAD

3.4 Mass spectrometry of the purified recombinant *P. sativum* ALAD.

Samples of *P. sativum* ALAD were analysed on a Micromass VG Quattro II mass spectrometer. Samples of protein were prepared by diluting the enzyme to 0.1 mg ml^{-1} in 50:50:1 acetonitrile:water:formic acid solution.

The predominant peak exhibited a molecular weight of 36711 with another species of molecular weight 36638 observed. Neither molecular weight matches the predicted molecular weight from the amino acid sequence (36611). The peak corresponding to a molecular weight of 36711 would possibly correspond to a molecule of 2-mercaptoethanol bound to a free cysteine and one magnesium ion bound to the subunit (36611 + 78 +24 = 36713). The peak of molecular weight 36638 could well correspond to a subunit with one magnesium ions bound (36611 + 24 = 36635), which is interesting, as the magnesium dependent ALADs are reported to have a low binding affinity for the magnesium ions.

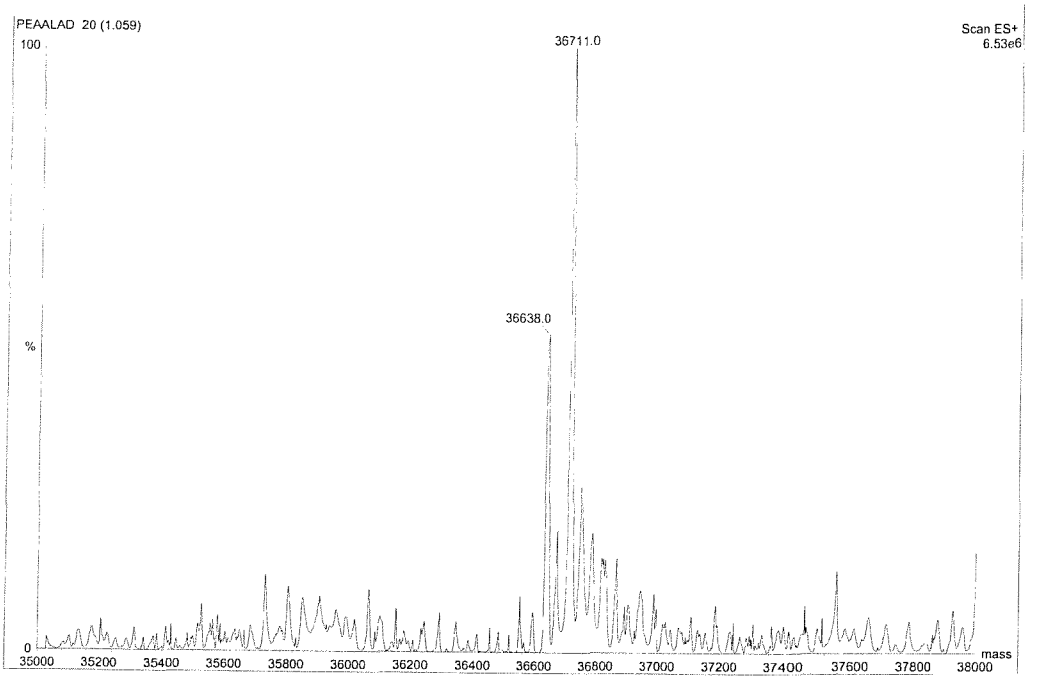


Figure 3.5 A mass spectra of the purified *P. sativum* ALAD. Two predominant species are observed at molecular weights of 36638 and 36711.

The suggestion that a surface cysteine (Cys 324) in one subunit may make a disulphide bond with the equivalent cysteine from an adjacent subunit within the dimer has been

suggested (Kervinen *et al.*, 2000). The location of these cysteines between neighbouring subunits in the dimers of the octameric structure are shown (Figure 3.6).

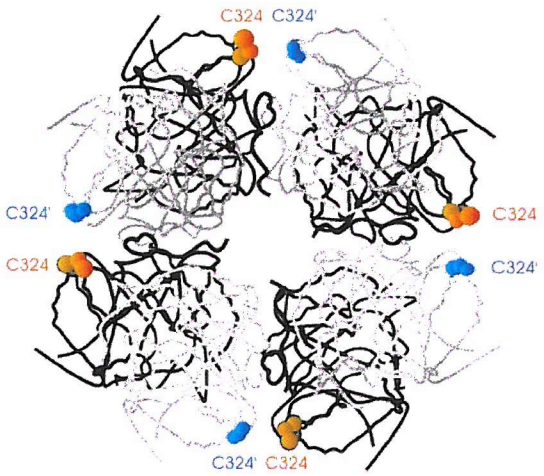


Figure 3.6 A model of the *P. sativum* ALAD based on the crystal structure of *P. aeruginosa*. Cys324 of one subunit is indicated in orange and cys324` of the neighbouring subunit is indicated in blue. (Reproduced from Kervinen *et al.* (2000)).

3.5 Optimal temperature studies

Enzyme assays were conducted using 10µg of pure pea enzyme at temperatures from 12°C to 43°C, following the protocol previously described in the methods section. The temperature optimum for the pea ALAD was found to be ~25°C (Figure 3.7) and similar treatment of the *E. coli* ALAD showed a temperature optimum of ~48°C (Figure 3.8). As can be seen from these graphs, the temperature optima for the two enzyme species vary substantially, given that they share 66% amino acid similarity.

The variation in the graphical data could well be explained by dissociation of the octameric structure into either dimers or monomers at certain temperatures, or, alternatively, dissociation of aggregates into octamers at higher temperature. It would appear to be more likely that octamers dissociate into tetramers or dimers, rather than monomers, since the strength of the interactions between monomers has been noted by crystallographers as being the strongest.

However, one other possible reason for the radical changes in temperature affects could result from the changes in pH of buffers as a function of temperature (up to 0.5 pH units) and this could influence the enzyme activity observed. This is thought unlikely, however, due to the broad pH range over which both the recombinant *P. sativum* and *E. coli* ALADs are optimally active.

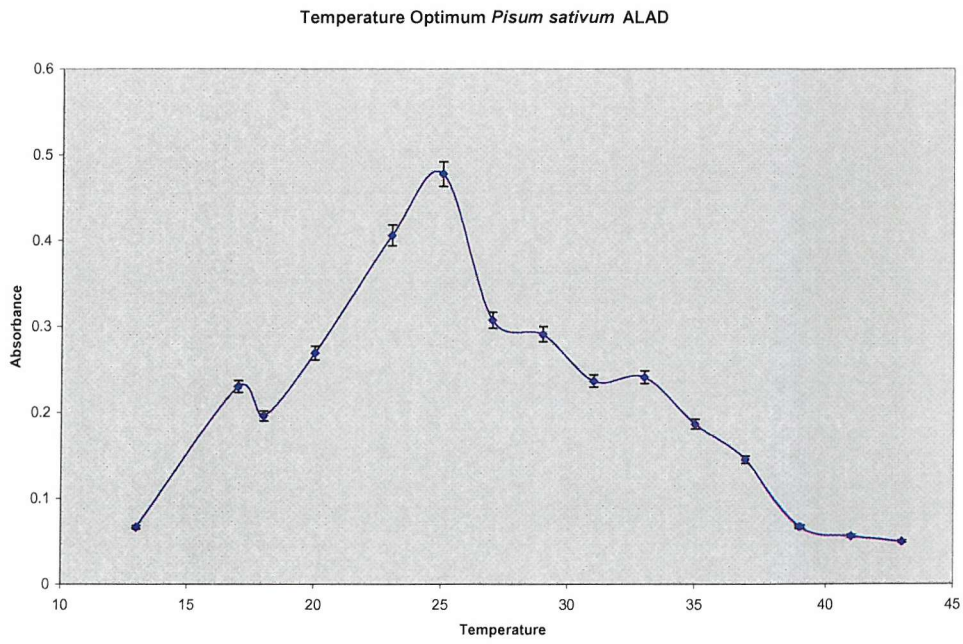


Figure 3.7 The effects of temperature on recombinant pea ALAD over a 30°C temperature range, with a temperature optimum at ~25°C.

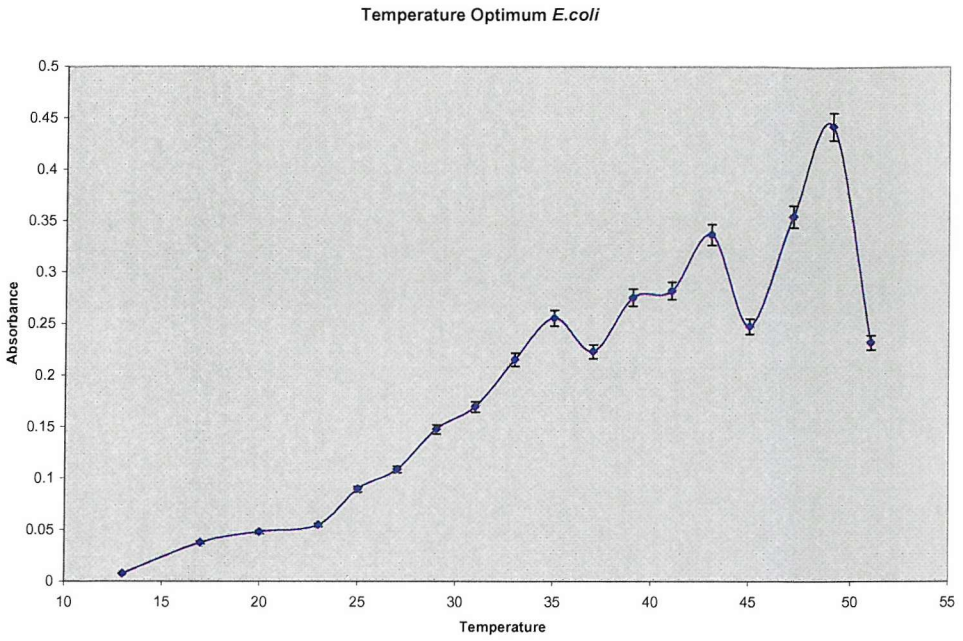


Figure 3.8 The effects of temperature on *E. coli* ALAD over a temperature range of 40°C, with a temperature optimum of ~48°C.

3.6 Optimal pH studies

Enzyme assays were conducted using 10µg of pure enzyme in a variety of assay buffers over a wide pH range. The buffers were chosen because of their buffering capacity over the desired pH. The buffers used were Tris, *bis*-Tris propane and CHES. Each assay buffer was made up at the required pH with the addition of 10mM $MgCl_2$ and 1mM 2-mercaptoethanol. The pH optimum tests were also conducted with buffered ALA (10mM Tris/HCl buffer, pH 7.0) to determine any effect of buffering the substrate. The assay was then conducted as described in Chapter Two. The results are shown in Figures 3.9 to 3.11:

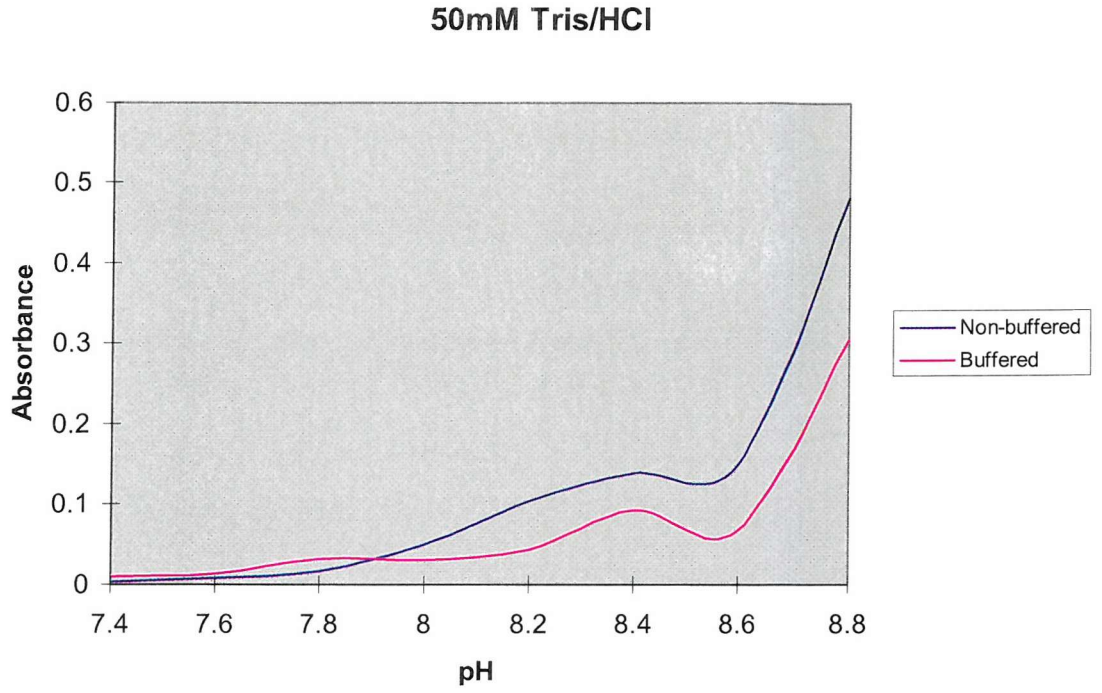


Figure 3.9 The effect of pH on enzyme activity over the 7.4-8.8 pH range using 50mM Tris/HCl buffer and also buffered and non-buffered substrate at a concentration of 10mM.

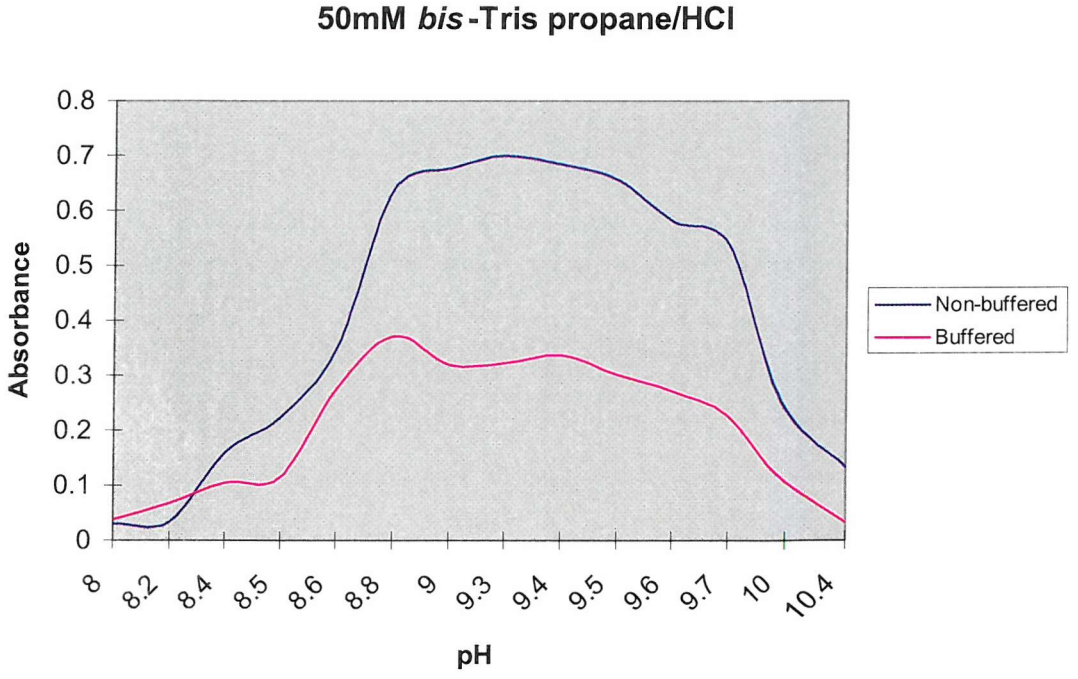


Figure 3.10 The effect of pH on enzyme activity over the 8.0-10.4 pH range using 50mM *bis*-Tris propane/HCl buffer and also buffered and non-buffered substrate at a concentration of 10mM.

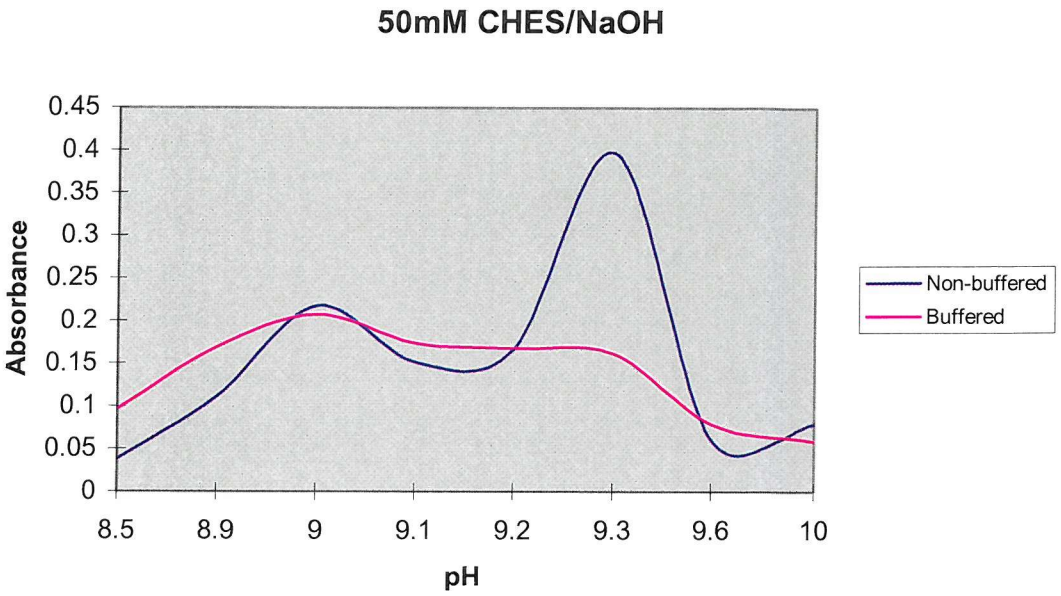


Figure 3.11 The effect of pH on enzyme activity over the 8.5-10.0 pH range using 50mM CHES/NaOH buffer and also buffered and non-buffered substrate at a concentration of 10mM.

The highest observed activity can be seen using 50mM *bis*-Tris propane/HCl buffer at pH 9.3. Unfortunately buffering of the ALA resulted in a reduction in activity, in some cases by over 50%, and this was observed to have a greater effect at higher pHs suggesting that the effective ALA concentration was less.

3.7 Investigating the dimerisation of ALA into a pyrazine

From the results gained from characterising the pH optimum of the recombinant ALAD enzymes, buffering of ALA can be seen to reduce the observed activity of the enzyme over the given pH ranges. This can possibly be explained by the chemical instability of ALA in base causing dimerisation to dihydropyrazine.

Previous work (Butler, 1992; Novo, 1996) investigating the dimerisation of ALA has shown the chemical dimerisation to be an irreversible process at alkaline pHs, with production of two different products, a dihydropyrazine and pyrazine (Figure 3.12).

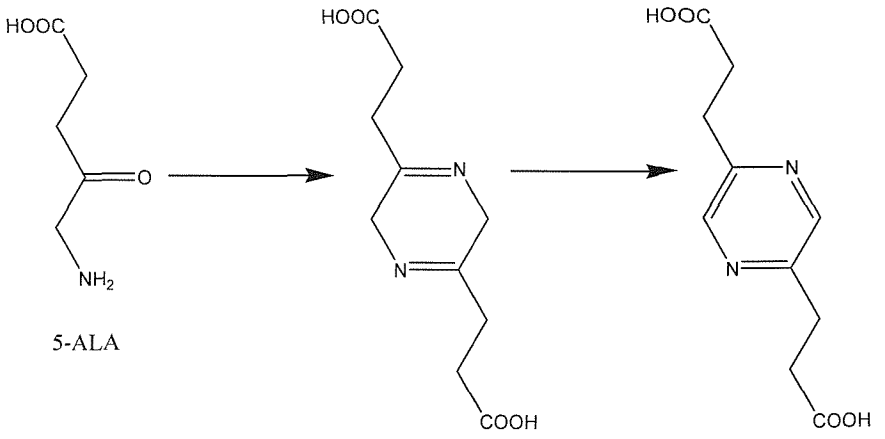


Figure 3.12 The chemical dimerisation of the substrate, ALA. The substrate may dimersise to form 2,5-(β-carboxyethyl)-dihydropyrazine that is readily oxidised to 2,5-(ß-carboxyethyl)-pyrazine under aerobic conditions.

^{1}H NMR studies were conducted to investigate the rate of formation of the pyrazine and therefore the potential reduction in available substrate at the pH values and temperatures investigated above for the *P. sativum, E. coli* and for the human recombinant ALAD enzyme.

Assay buffers were prepared, as described in the methods section, omitting any metal ions. ^{1}H NMR spectra were recorded for the assay buffers before the addition of ALA, to a final concentration of 100mM. Following addition of ALA to the buffer, ^{1}H NMR spectra (Figures 3.13 and 3.14) were recorded every 20 minutes to monitor any dimerisation of the ALA. The experiment was repeated for each assay buffer at both 25°C and 37°C. A control was also run to monitor the dimerisation of 10mM ALA in AnalaR water over a 24hr period.

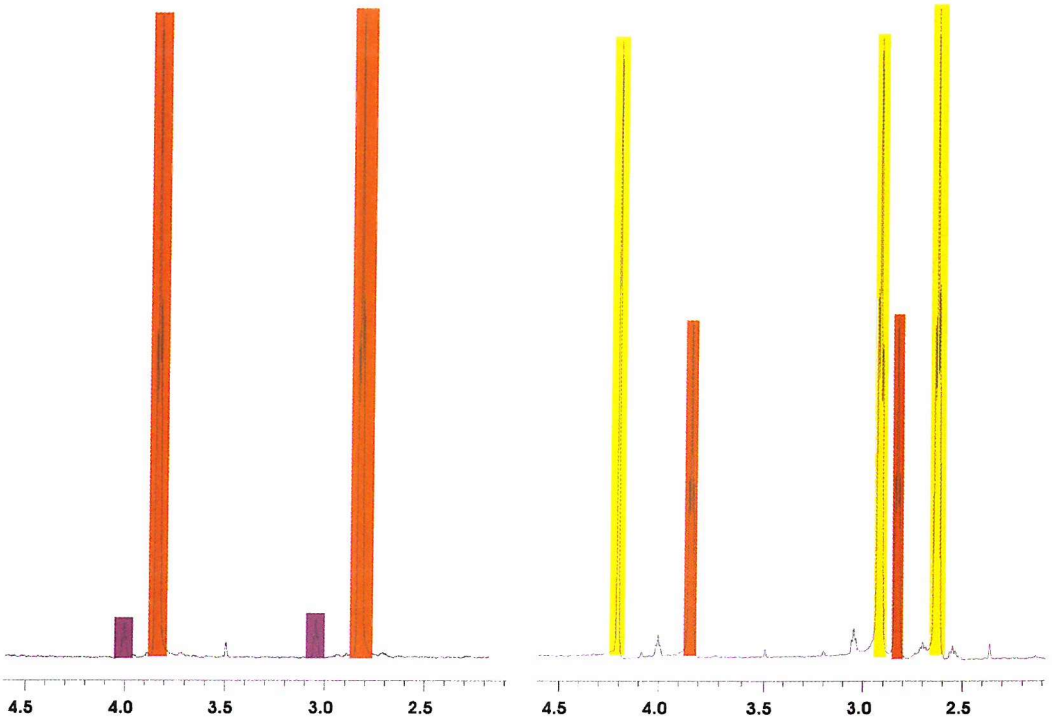


Figure 3.13 ^{1}H NMR spectra of the *E. coli* assay buffer without ALA, and then a further spectrum after ALA addition at 37°C. Peaks labelled in yellow (δ=2.60, 2.90 and 4.20) correspond to ALA. Peaks labelled in orange (δ=2.80 and 3.80) and purple (δ=3.05 and 4.00) show the active and the disulphide form of 2-mercaptoethanol, an integral part of the assay buffers.

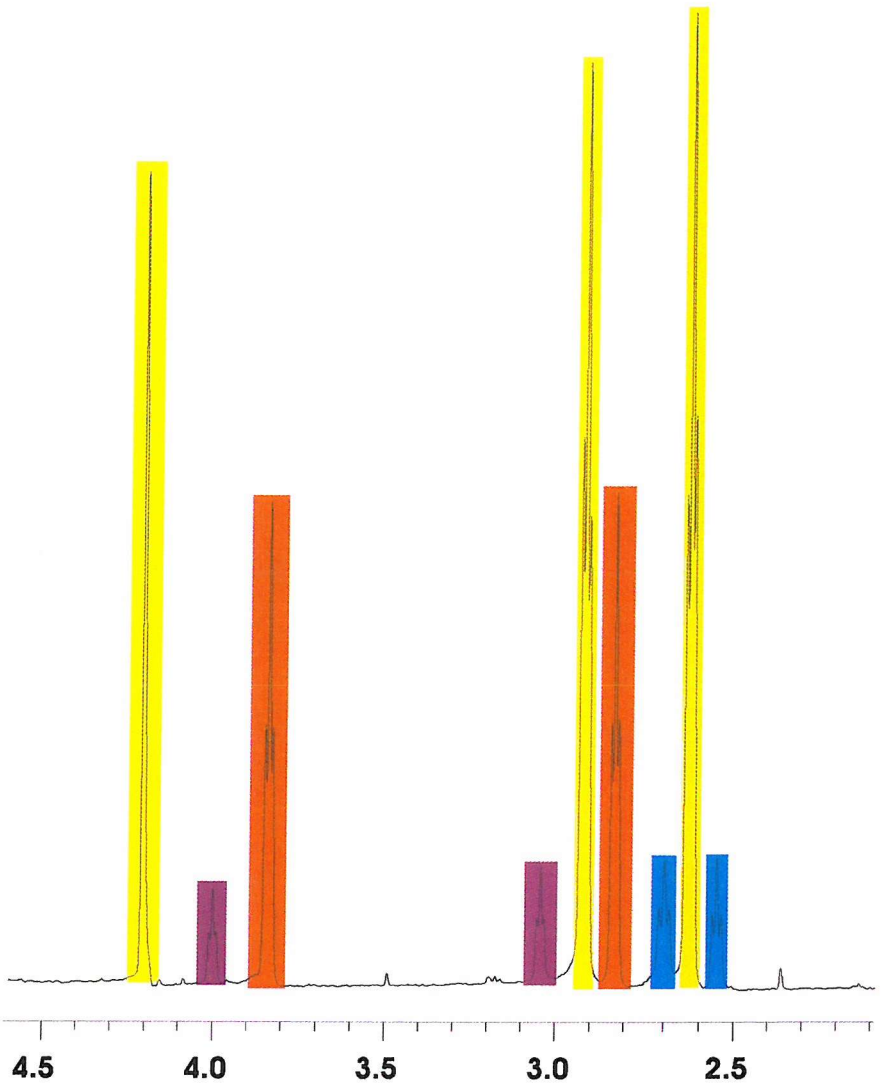


Figure 3.14 The final ^{1}H NMR spectra for the experiment following the dimerisation of ALA under the *E. coli* assay conditions at 37°C. The additional peaks are labelled in blue, and these coincide with the dihydropyrazine product formed by the chemical dimerisation of ALA.

^{1}H NMR assignments for the different compounds were conducted and the peaks for ALA and the product integrated were used to calculate the percentage conversion of ALA to pyrazine (Figure 3.15).

Based upon the ^{1}H NMR observed it was assumed that the major product formed from the dimerisation was most likely to be the pyrazine rather than the dihydropyrazine. This was based upon the fact that the dihydropyrazine would be predicted to give an additional peak at ~3.0ppm, and no major peak could be observed corresponding to this. It was therefore

assumed that under aerobic conditions that the dihydropyrazine product of the dimerisation was most likely to be rapidly oxidised to the pyrazine product.

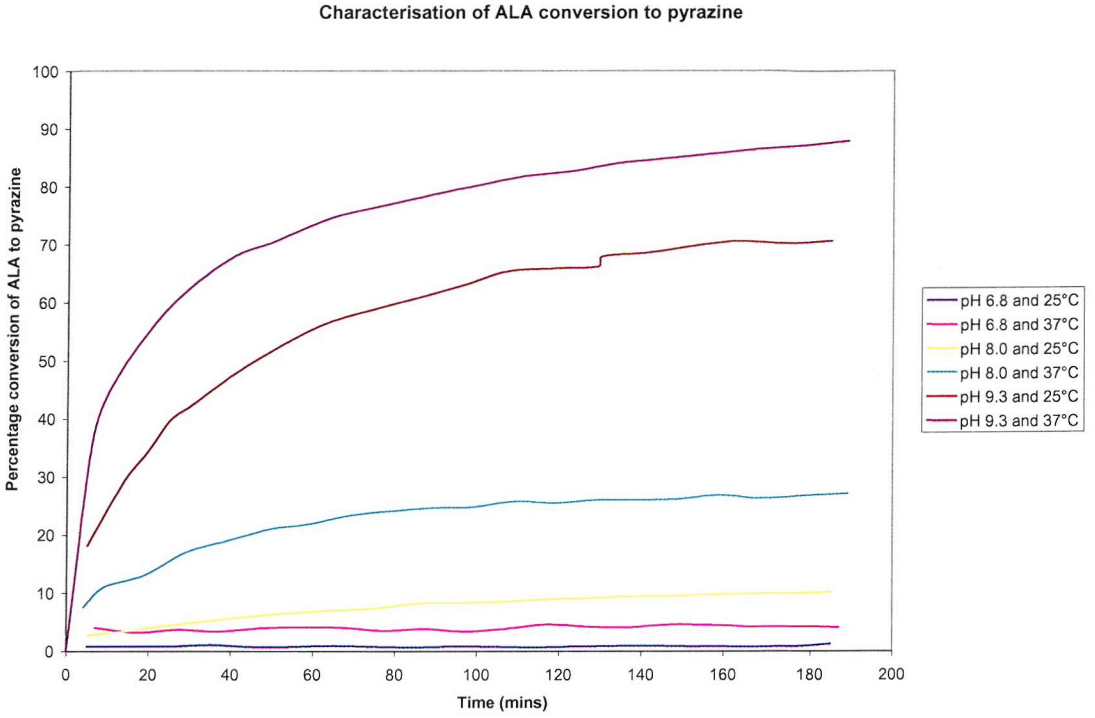


Figure 3.15 Formation of ALA pyrazine over time. Experiments were conducted at pH 6.8 (potassium phosphate buffer), pH 8.0 (potassium phosphate buffer) and pH 9.3 (*bis*-Tris propane/HCl buffer) at temperatures of 25°C and 37°C.

It can be seen that the increase in pH, and also temperature, increases the speed at which the pyrazine forms. Given that the average *P. sativum* ALAD assay is conducted over 10 minutes at 25°C or 37°C, the importance of these results, particularly with regard to enzymes assayed at high pHs, can be seen – at least 20% of the substrate is lost to form pyrazine under these conditions over a 10 minute time period. Therefore the recorded activity of the enzyme could be somewhat underestimated at elevated pHs and temperature. This effect could be magnified if the pyrazine was an ALAD inhibitor.

3.8 Stability of recombinant *P. sativum* ALAD

Stability of recombinant pea ALAD was tested at room temperature (18°C), at -20°C and at -80°C, with and without a nitrogen environment. Samples were assayed every 24 hours

over a 12 day period. Samples were also taken at each 24 hour interval and analysed by SDS PAGE and native PAGE. Results are shown in Figures 3.16 - 3.18 for each condition:

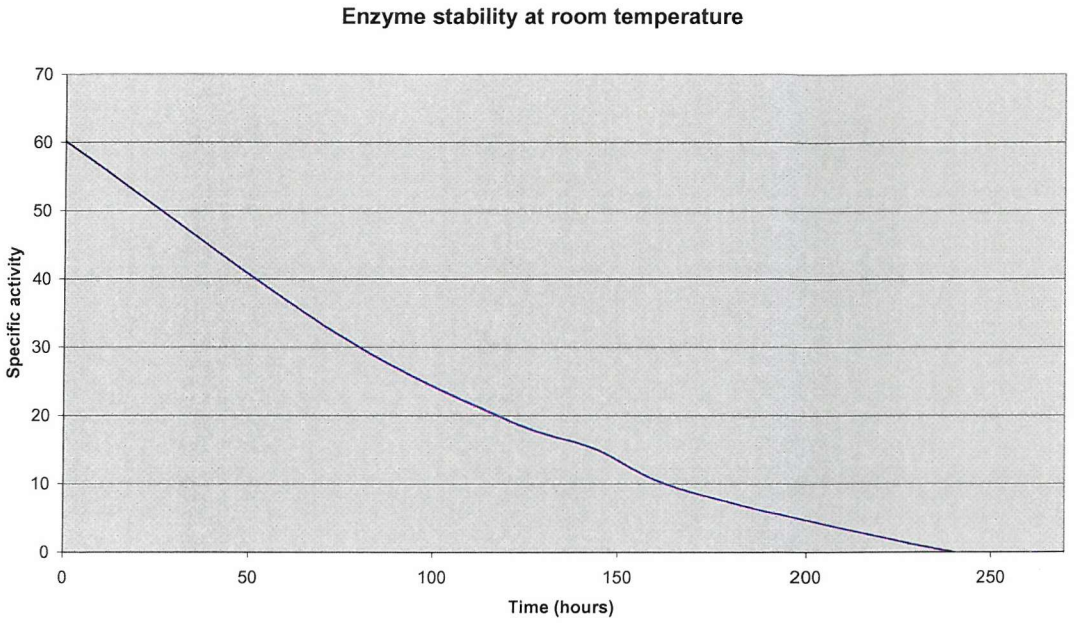


Figure 3.16 A graph showing a time course of enzyme stability at 18°C over 12 days.

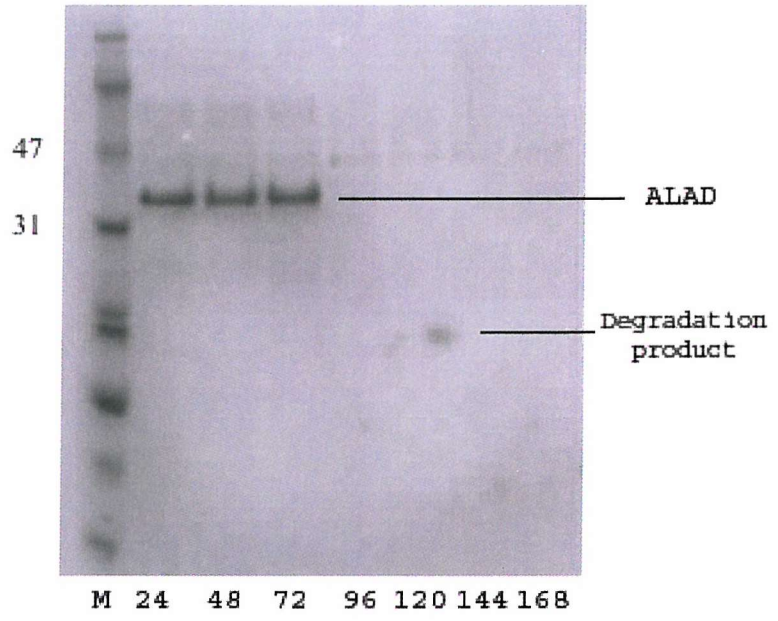


Figure 3.17 SDS PAGE gel of samples taken over the enzyme stability time course (24 hrs to 168 hrs).

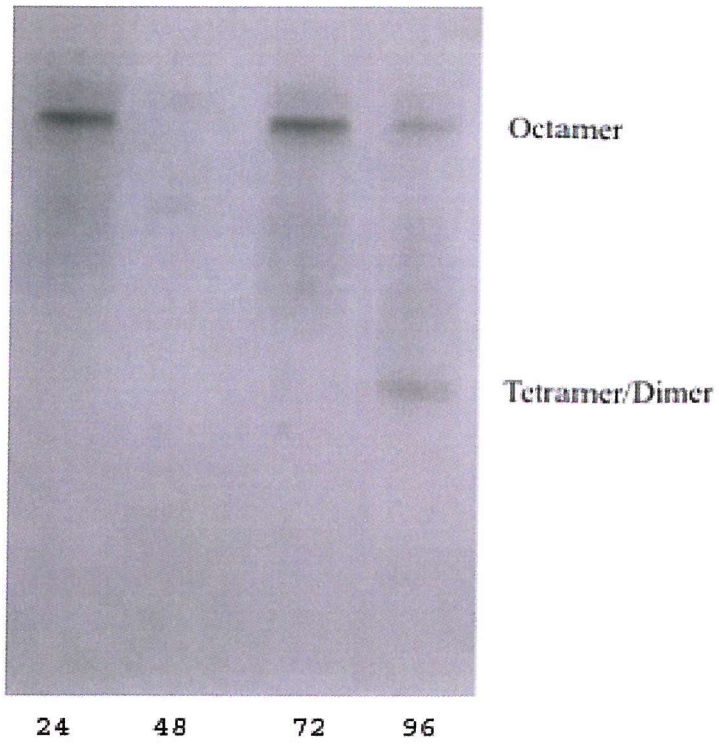


Figure 3.18 Native PAGE gel of samples taken over the enzyme stability time course (24hrs – 96hrs).

The major cause of instability appears to be due to aggregation and precipitation of the protein over time to form a white precipitate. This would appear to be due primarily to oxidation, as attempts to keep the enzyme under a nitrogen environment increased the lifetime. The loss of the enzyme from solution can be seen by SDS PAGE gel. Also the native PAGE gel shows dissociation of the subunits of the octamer into either tetramers or dimers over the same time period.

The apparent aggregation of subunits could have been due to the availability of a free cysteine residue as mentioned previously (Cys324) (Kervinen *et al.*, 2000) which would allow the formation of disulphide bridges between different octamers (Figure 3.6). In light of this, a single point mutation was generated by Dr M. Sarwar (Southampton University) in order to investigate this possibility.

3.9 Recombinant *P. sativum* ALAD with a mutation of cysteine 326 to alanine

The mutant Cys326Ala was purified by the same method as that for the recombinant *P. sativum* and yielded similar results and a similar specific activity. The stability of the pure protein was investigated under similar conditions to those discussed above for investigating the stability of the native enzyme. However, no significant differences in stability in solution were observed.

3.10 Discussion

In conclusion, a shortened recombinant form of the *P. sativum hem2* gene has been synthesised and shown to overexpress in *E. coli* BL21(DE3). A new growth and overexpression protocol has been optimised and, along with a newly devised purification technique, has allowed consistent production of milligram amounts of the *P. sativum* ALAD protein. The new purification technique has allowed production of protein in three days in comparison to previous protocols, which on average took five days. Additionally, the specific activity of the shortened *P. sativum* ALAD has been observed to be much lower (70μmol mg^{-1} ml^{-1}) than that observed for the full length protein (280 μmol mg^{-1} ml^{-1}) (Cheung *et al.*, 1997). The protein has been characterised by mass spectroscopy, SDS-PAGE electrophoresis and non-denaturing gel electrophoresis yielding results indicating an octameric/hexameric protein with a monomeric subunit M_r of 36,611, consistent with the predicted size derived from the nucleotide sequence.

Investigations into the optimal pH and optimal temperatures for the *P. sativum* and *E. coli* ALAD enzymes have yielded interesting differences between the two enzymes, especially considering the high overall similarity shown between their two primary amino acid sequences. Additionally, this has allowed the optimisation of the conditions required for assays of the *P. sativum* ALAD (50mM *bis*-Tris propane, pH 9.3) that, interestingly, differ from those previously reported for the full length *P. sativum* ALAD (50mM Tris-HCl, pH 8.5) (Cheung *et al.*, 1997). The data observed have also indicated the possibility of dissociation of the quaternary structure of both the *P. sativum* and *E. coli* ALADs over the

temperature range shown. The observation of what appear to be two distinct pH optimums for the *P. sativum* enzyme could be indicative of an octameric-hexameric equilibrium, with corresponding differences in characteristics for each quaternary form.

Quantitative NMR investigations into the chemical dimerisation of ALA into the pyrazine and dihydropyrazine under aerobic conditions have been undertaken and have highlighted the importance of pH and temperature when working with this reactive substrate under assay conditions of ALAD. Further investigations could be undertaken to determine whether the dihydropyrazine and/or pyrazine, formed from this dimerisation reaction, can inhibit the enzyme.

Stability of the enzyme in solution under various conditions has been monitored. Results indicate that oxidation appears to play an important role in the stability of the enzyme, along with aggregation of the subunits through interaction between octamers. By the use of a Cys324Ala mutant, the formation of disulphide bridges between octamers, through Cys323 residues, has been investigated as a possible cause of aggregation. Results however did not support this theory.

Additionally, native gel electrophoresis experiments monitoring molecular changes to the enzyme have indicated that, as the protein precipitates, the quaternary structure of the enzyme alters and an equilibrium appears to form between two species - potentially an octameric form and also a smaller molecular species. It is possible, since during the degradation process the concentration of the protein decreases in solution, that this indicates a concentration dependence for formation of the octamer in solution. This has been indicated in previous work on ribonucleotide reductase where the concentration of the enzyme determines the quaternary structure in association with ATP concentrations (Kashlan, 2003). The observations may indicate that the formation of different quaternary structures of ALAD could be mediated not only by enzyme concentration, but also by the concentration of Mg^{2+}, temperature and pH.

Recent work on the Phe12Leu mutant of human erythrocyte ALAD has implied that this variant, with a mutation in the *N*-terminal arm of the monomer, forms a stable hexamer. As has been stated previously, the *N*-terminal arm is essential for contacts between monomers within the dimer and the rearrangement of the arm due to the mutation, results in profound changes in kinetic behaviour for the enzyme. The hexamer has been shown to be composed from three dimers each rotated 120° around a central axis. This is in contrast to the structures characterised so far, where the dimers are described to be "hugging", where the *N*-terminal arm from one subunit of each dimer wraps around the barrel region of its neighbouring subunit. The new hexameric structure describes the dimer formation as being "detached", where the dimer maintains the barrel-to-barrel contacts but the *N*-terminal arms are detached rather than "hugging", which allows the formation of the hexamer (Breinig, 2003). This could be applied to the shortened *P. sativum* construct where there appears to be several distinct peaks when assaying at different temperatures and pH values, which could indicate different oligomeric confirmations, dependant upon the *N*-teminal arms confirmation at a given pH or temperature. It would be very interesting to conduct research into the conditions at which the dissociation from octamers into lower quaternary structure forms occurs and also whether the concentration of Mg^{2+} ions and substrate also affect quaternary structure.

Further work needs to be conducted to characterise the stability of the recombinant shortened construct of *P. sativum* ALAD. This could possibly be achieved by modification and development of a better storage buffer and possibly by the addition of a higher salt so as to minimise interactions between octamers.

Chapter Four: Crystallisation of *Pisum sativum* 5-aminolaevulinic acid dehydratase

4.1 Introduction

Up until now, crystal structures of ALADs from yeast(Erskine *et al.*, 1997), *E. coli* (Erskine *et al.*, 1999), human (Mills-Davies, 2000), and *P. aeruginosa* (Frankenberg *et al.*, 1999) have been solved. These structures have given great insight into mechanism of the Zn^{2+} dependant mammalian and bacterial ALADs. The structure of the *P. aeruginosa* enzyme, however, has not provided the expected information about the nature of the Mg^{2+} metal site. This, together with the fact that no ALAD structure has been solved for a Mg^{2+} dependant enzyme from a higher plant species, encouraged attempts to investigate the structure of the pea enzyme.

This chapter describes the attempts to crystallize and gain a phase solution for the recombinant *P. sativum* ALAD datasets. Unfortunately, no structure was forthcoming, due to the inability to solve the phases associated with the data sets obtained. However, due to the complexity of acquiring good quality datasets, insight into the methodology of crystallization has been gained, that could not only benefit further work on the crystallization of this ALAD, but may also assist others working on the crystallization of different proteins. Similarly, with regard to the attempts to solve the phase problem, a modified protocol for the growth and expression of seleno-L-methionine-labelled pea ALAD has been devised and shown to be a valid method for the obtaining labelled protein for MAD experiments.

4.2 Preliminary crystallisation condition screening

Molecular dimensions `Magic 100` screen and 'Clear Strategy' screens were used to ascertain conditions under which pure recombinant pea ALAD enzyme would crystallise. The hanging drop technique was used (see Chapter Two).

Each screen was conducted using pure protein at a concentration from 2.5-15 mg/mL in 2.5 mg/mL increments. Crystal forms that grew were analysed and assessed for desirable characteristics such as size (a minimum of 0.05mm in the shortest crystal edge), sharp

edges and overall perfection (Tables 4.1 and 4.2). Two of the conditions were selected using these discriminating factors:

Condition one: 1.6M Ammonium sulphate, 0.1M Mes pH 6.5, 10% w/v dioxane
Condition two: 2% w/v Dioxane, 0.1M bicine pH 9.0, 10% w/v PEG 20,000

Crystals yielded by these preliminary conditions were frozen, with the addition of 30% w/v glycerol as a cryoprotectant, and added directly to a drop of mother liquor, and mounted on an X-ray beam. However, neither set of crystals gave consistent diffraction data and both sets of crystals disintegrated rapidly on the beamline at the ESRF, Grenoble. It was decided that in order to obtain better crystals, crystal screen variations would be required.

Condition	Shape	Size (mm)	Condition
3	Rectangular plate	>0.03	Thin, sharp edges
27	Cube	<0.03	Clean, sharp edges
32	Cube	<0.01	Microcrystal, twinned
45	Convex oval plate	<0.03	Thin, cracked surface
53	Rod	<0.03	Thin, sharp edges
57	Cube	<0.03	Microcrystal, twinned
73	Rectangular plate	>0.03	Thick, slight surface cracks

Table 4.1 'Magic 100' crystallization screen hits of pea ALAD

Condition	Size (mm)	Condition
12	<0.03	Soft edged, unsuitable
17	<0.03	Soft edged, unsuitable
18	<0.03	Soft edged, unsuitable
24	<0.03	Soft edged, unsuitable

Table 4.2 'Clear Strategy' screen one hits

Condition	Size	Condition
18	<0.03	Soft edged, very fragile, unsuitable

Table 4.3 'Clear Strategy' screen two hit

4.3 Secondary crystallisation condition screening

Subsequently, crystals screens were devised around the preliminary conditions in an attempt to gain larger and more stable crystals (Tables 4.4 and 4.5). As hoped, these screens did yield larger crystals, but these were only marginally more stable. Again, crystals from both condition screens were frozen and mounted. Crystals (Figure 4.1) from the condition one screen diffracted weakly to ~10Å, but the crystals from the condition two screen again disintegrated on the beamline. At this point, condition one seemed more promising, so condition two was discarded as a line of investigation.

A	B	C	D	E	F	
0.1M Mes pH 6.50 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.50 5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 6.75 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.75 5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 7.00 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 7.00 5% dioxane 2.4M Ammonium sulphate	1
0.1M Mes pH 6.50 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.50 5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 6.75 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.75 5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 7.00 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 7.00 5% dioxane 2.4M Ammonium sulphate	2
0.1M Mes pH 6.50 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.50 7.5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 6.75 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.75 7.5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 7.00 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 7.00 7.5% dioxane 2.4M Ammonium sulphate	3
0.1M Mes pH 6.50 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.50 7.5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 6.75 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.75 7.5% dioxane 2.4M Ammonium sulphate	0.1M Mes pH 7.00 7.5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 7.00 7.5% dioxane 2.4M Ammonium sulphate	4

Table 4.4 Secondary crystallisation screen for condition one

A	B	C	D	E	F	
0.1M Bicine pH 8.50 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.50 3% Dioxane 15% PEG 20,000	0.1M Bicine pH 8.75 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.75 3% Dioxane 15% PEG 20,000	0.1M Bicine pH 9.00 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 9.00 3% Dioxane 15% PEG 20,000	1
0.1M Bicine pH 8.50 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.50 3% Dioxane 15% PEG 20,000	0.1M Bicine pH 8.75 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.75 3% Dioxane 15% PEG 20,000	0.1M Bicine pH 9.00 3% Dioxane 10% PEG 20,000	0.1M Bicine pH 9.00 3% Dioxane 15% PEG 20,000	2
0.1M Bicine pH 8.50 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.50 5% Dioxane 15% PEG 20,000	0.1M Bicine pH 8.75 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.75 5% Dioxane 15% PEG 20,000	0.1M Bicine pH 9.00 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 9.00 5% Dioxane 15% PEG 20,000	3
0.1M Bicine pH 8.50 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.50 5% Dioxane 15% PEG 20,000	0.1M Bicine pH 8.75 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 8.75 5% Dioxane 15% PEG 20,000	0.1M Bicine pH 9.00 5% Dioxane 10% PEG 20,000	0.1M Bicine pH 9.00 5% Dioxane 15% PEG 20,000	4

Table 4.5 Secondary crystallisation screen for condition two

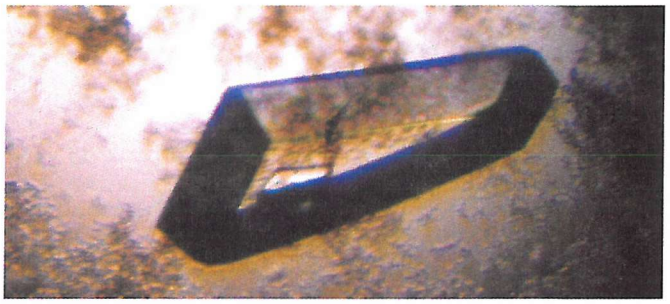


Figure 4.1 Example of a crystal yielded by secondary crystallisation screening.

The crystal was ~1.0mm in length and was frozen under liquid ether and mounted on the in-house rotating anode X-ray generator. Diffraction data was gained to ~10Å was gained.

4.4 Stability of crystals

It was considered that a major factor in the poor resolution of the diffraction data may well have been related to the stability of the crystals (and therefore the protein). It was noted that a slight change of solvent conditions for the crystal did not change the overall crystal but did cause what seemed like slight damage to the surface of the crystal. This damage could also be caused by transferring the crystal to a cryo-loop and exposing the crystal to physical stresses. Therefore, further work was conducted; to improve the stability of the crystals by improving both the crystal packing and the crystallisation process; by the use of additional ligands; improving the crystal mounting procedure through the use of alternative cryoprotectants or wet mounting of the crystals.

4.5 Improvement of crystal packing

Various factors are thought to influence the packing of crystals and, to investigate this, several approaches were undertaken.

4.5.1 Temperature screening

The temperatures at which the crystallisation screens are performed determine the rate at which vapour diffusion takes place. Attempts were therefore made to grow the crystals at 4°C in a cold room, as opposed to 25°C at which all other previous screens had been conducted. Unfortunately these screens yielded no crystals, as has been noted for human ALAD.

4.5.2 Oil interface

The use of oils layered on the mother liquor in the wells is also thought to slow the rate of vapour diffusion between the hanging drop and the main buffer. Screens using paraffin oil, Dow Corning oil and a 50:50 mix of paraffin and Dow Corning oil were set up. Strangely, these screens yielded crystals faster than the previous conditions. The crystals obtained, however, showed no increase in diffraction quality.

4.5.3 Ligand binding

Co-crystallisation screens were set up using 7mM concentrations of 5-aminolaevulinic acid (ALA), laevulinic acid (LA), 4,7-dioxosebacic acid (DOSA) and porphobilinogen (PBG). The screens with added ALA and LA yielded crystals. Again, diffraction data was not improved significantly. Also, ligand soaking was employed as a way of introducing selected ligands into the protein structure, but with little effect to the data observed.

4.5.4 Site-directed mutagenesis

Three mutants were designed that were thought to be good candidates for investigating metal binding at the proposed metal binding regions of the pea ALAD. It was thought that perturbing the binding of the Mg2+ ions could result in a trapped intermediate that would, in turn, result in a more stable and ordered crystal form.

Three mutants were generated by Dr M. Sarwar (Southampton University) each with a mutation in which aspartic acid was substituted by asparagine at the key positions 123, 131 and 168 (Figure 4.2). These constructs were then grown, expressed and purified in a similar fashion to the native *P. sativum* enzyme.

```
          10         20         30         40         50         60
          |          |          |          |          |          |
MPIQRRPRRN RRSPALRSAF QETTLSPANF VYPLFIHEGE EDTPIGAMPG CYRLGWRHGL

          70         80         90         100        110        120
          |          |          |          |          |          |
LEEVAKARDV GVNSVVLFPK IPDALKTPTG DEAYNEDGLV PRSIRLLKDK YPDLIIYTDV

          130        140        150        160        170        180
          |          |          |          |          |          |
ALDPYSSDGH DGIVREDGVI MNDETVHQLC KQAVAQARAG ADVVSPSDMM DGRVGAMRVA

          190        200        210        220        230        240
          |          |          |          |          |          |
LDAEGFQHVS IMSYTAKYAS SFYGPFREAL DSNPRFGDKK TYQMNPANYR EALTEMREDE

          250        260        270        280        290        300
          |          |          |          |          |          |
SEGADILLVK PGLPYLDIIR LLRDNSPLPI AAYQVSGEYS MIKAGGALKM IDEEKVMMES

          310        320
          |          |
LLCLRRAGAD IILTYFALQA ARTLCGEKR
```

Figure 4.2 The protein sequence encoded by the recombinant *P. sativum* cDNA derived from the *hem2* gene. The positions of the putative metal binding aspartates mutated are shown in red and the active site lysines in pink.

These mutant recombinant *P. sativum* ALADs were all devoid of activity, and could not be activated by incubation with Mg^{2+}. Pure protein was used for crystallographic trials using similar secondary crystallisation conditions as those used for the recombinant native enzyme. Unfortunately, although similar crystals were formed, no better quality diffraction data was obtained and further work was continued with the native recombinant *P. sativum* ALAD.

4.5.5 Further pH dependant screening

During initial screening, different crystal forms were occasionally and briefly observed at the higher and lower pH boundaries of the conditions. These observations suggested that across the small pH range over which the screens were conducted; two separate conditions

of crystal formation arose depending solely on the pH. Therefore, following this idea, a much wider pH screen was attempted over the pH range of 5 - 9.3 in 0.25 pH unit increments in order to see if a greater range of pH could yield additional pH-dependant crystallisation conditions. Over this range, three different crystal forms (see below) were observed. The photographs of these screens can be seen in Figures 4.4 - 4.7.

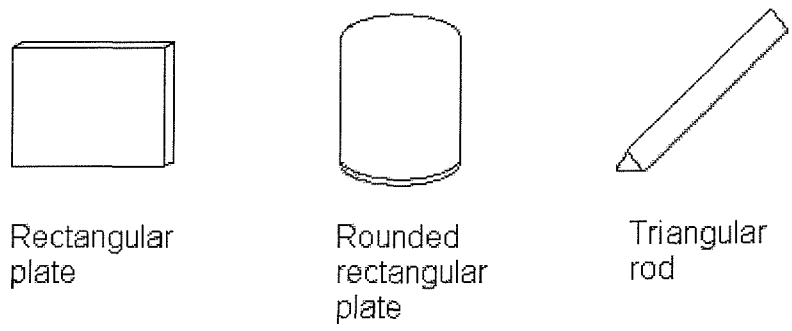


Figure 4.3 Three different crystal forms were observed over the wider pH range from pH 6.0 - 9.25. Although the physical appearance of the crystals varied over the entire pH range, preliminary data collection and interpretation showed that all three variations of crystal form had similar crystal packing and unit cell dimensions.

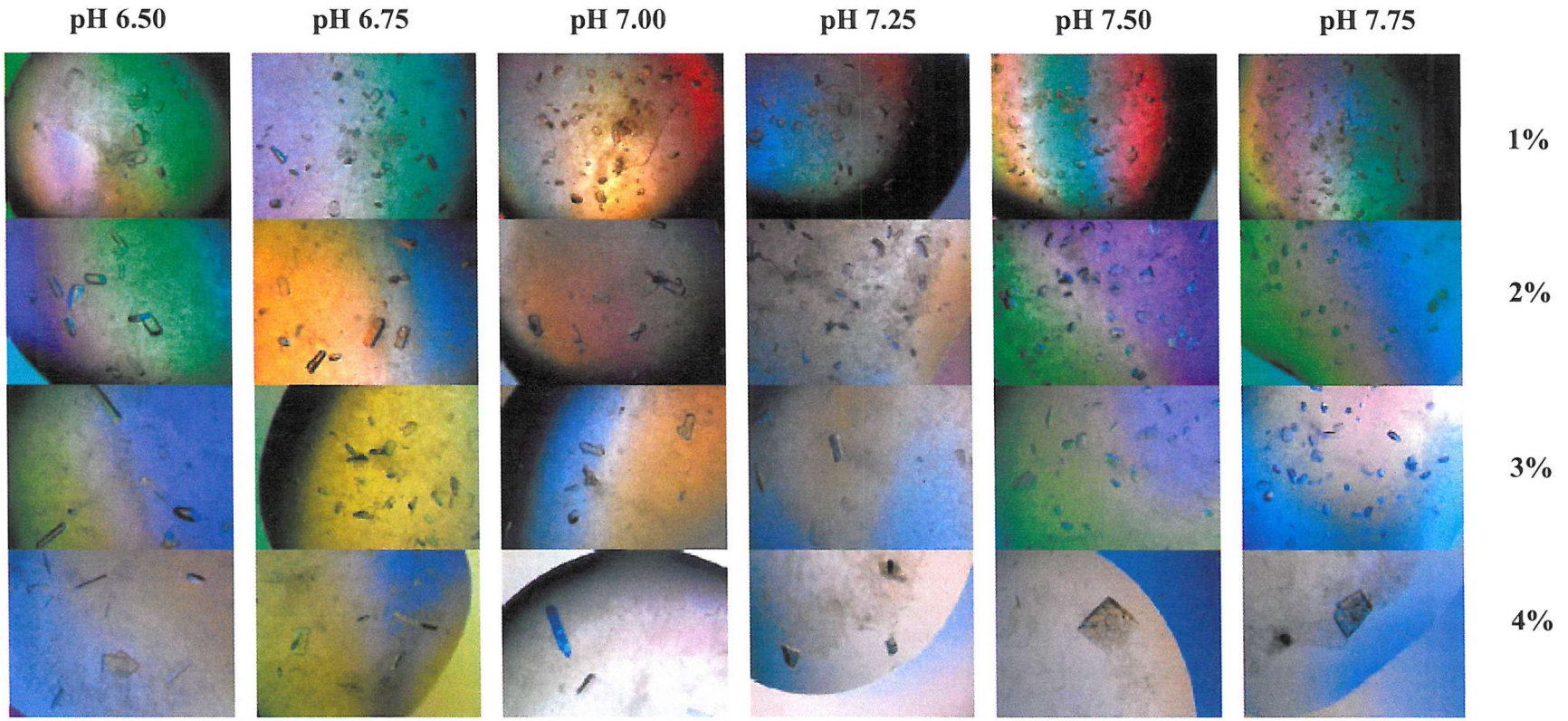


Figure 4.4 Adapted crystal screen showing variation of crystal forms across a range of pH from 6.50-7.75 (rows), and dioxane concentrations from 1-4% (columns). Each well contained a standard concentration of 2.0M ammonium sulphate and 7mM laevulinic acid. The crystals were grown within a drop composed of 5µl of 10mg mL^{-1} protein solution and 5µl of mother liquor.

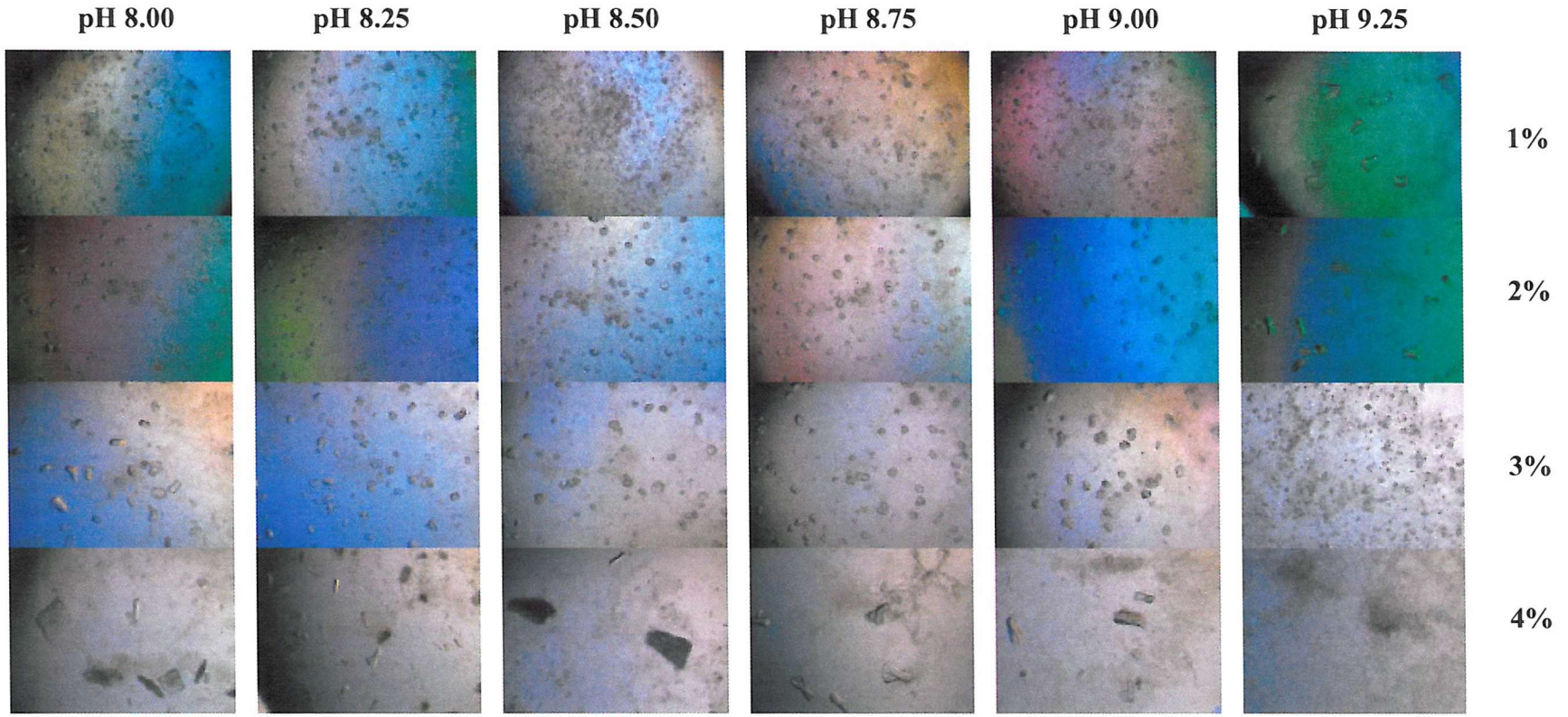


Figure 4.5 Adapted crystal screen showing variation of crystal forms across a range of pH from 7.75-9.25 (rows), and dioxane concentrations from 1-4% (columns). Each well contained a standard concentration of 2.0M ammonium sulphate and 7mM laevulinic acid. The crystals were grown within a drop composed of 5µl of 10mg mL^{-1} protein solution and 5µl of mother liquor.

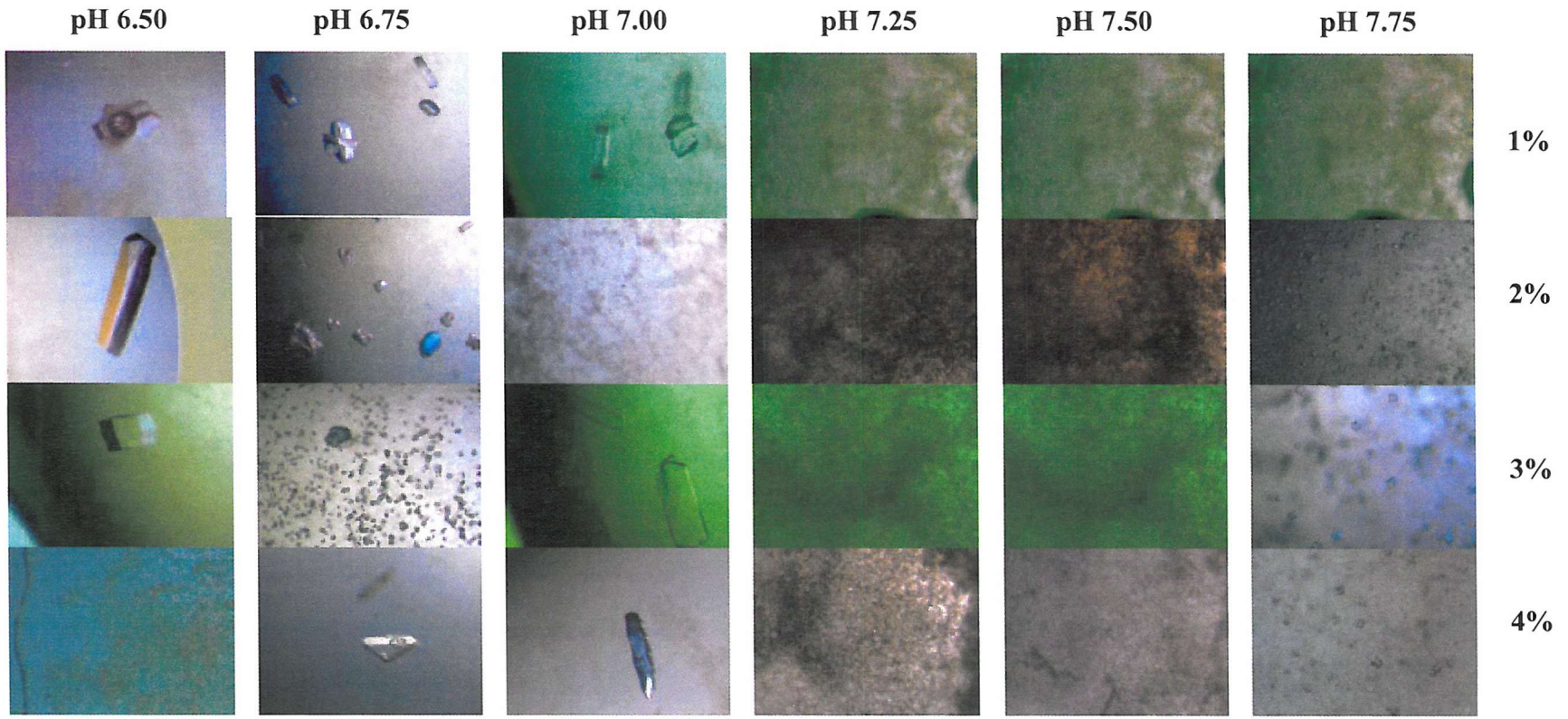


Figure 4.6 Adapted crystal screen showing variation of crystal forms across a range of pH from 6.50-7.75 (rows), and dioxane concentrations from 1-4% (columns). Each well contained a standard concentration of 2.0M ammonium sulphate, 7mM laevulinic acid and an excess of EDTA. The crystals were grown within a drop composed of 5µl of 10mg mL^{-1} protein solution and 5µl of mother liquor.

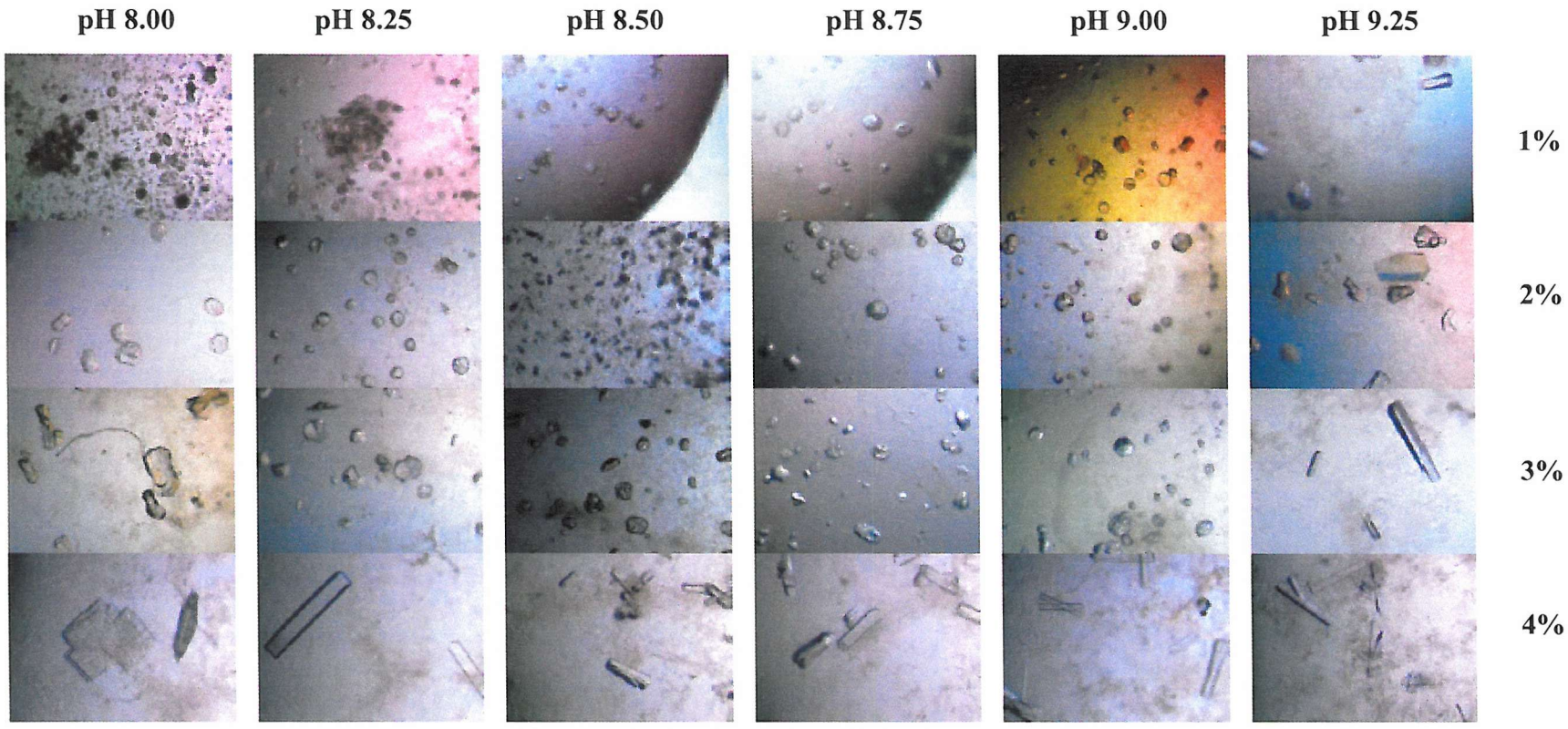


Figure 4.7 Adapted crystal screen showing variation of crystal forms across a range of pH from 7.75-9.25 (rows), and dioxane concentrations from 1-4% (columns). Each well contained a standard concentration of 2.0M ammonium sulphate, 7mM laevulinic acid and an excess of EDTA. The crystals were grown within a drop composed of 5µl of 10mg mL^{-1} protein solution and 5µl of mother liquor.

Each of the three crystal forms was wet mounted and diffraction data were collected to ~6Å. Each crystal was auto-indexed using MOSFLM (Collaborative Computational Project, 1994) to gain information about cell dimensions and Bravais lattice. The results showed that although the crystal forms appeared different, the cell and lattice parameters had not changed, despite the variation in pH.

4.5.6 Constitutive metal ion screening

Screens were also conducted with pure enzyme that had had the metal ions removed from the solution (using EDTA). Screens were conducted over the pH range 6.0-9.25, but no improvement in diffraction quality was observed. Interestingly, the removal of the metal ions appeared to affect the pI of the enzyme.

4.5.7 Crystal dehydration techniques

The dehydration of crystals to improve crystal packing has been documented in several cases (Schick, 1994; Fu, 1999). Existing crystals were immersed in drops of 3.0M solutions of sodium chloride and ammonium sulphate, and also 1.5M:1.5M mixes of the two solutions and the drops were inverted over wells of the same solutions and left for 7 days. Crystals were observed to reduce in size, by at least 25% of their former size. Unfortunately no improvement in diffraction data was observed.

4.6 Improvement of the crystal mounting procedure

Because of the poor stability of the crystals, attempts were made to improve the crystal handling and mounting procedures.

4.6.1 Wet mounting of crystals

As an alternative to freezing, wet mounting of the crystals was attempted (see methods) to avoid the stresses to the crystal imposed by the freezing process. This improved diffraction quality to around 6 - 8Å, depending on the crystal.

4.6.2 Screening and use of cryoprotectants

Following the preliminary problems of freezing the crystals without damaging the crystal

surface, several known and suggested cryoprotectants were tested (as described in the Chapter Two) and maximum concentrations were determined for the specific conditions required:

Cryoprotectant	**% v/v or w/v required**
(*2R, 3R*)-(-)-2,3-Butandiol	15%
D-Glucose	15%
Glycerol	20%
Glycerol:3.0M ammonium sulphate	10%
50:50 Glycerol:isopropanol	18%
Xylitol	18%
Ethylene glycol	15%
Polyethylene glycol $_{400}$	Immiscible with mother liquor
MPD (2-methyl-2,4-pentanediol)	Immiscible with mother liquor

Table 4.6 The minimal required amounts of cryoprotectant required for vitrification of the frozen crystallisation buffer.

Incorporation of the cryoprotectant into the freezing protocol resulted in the improvement of the diffraction data to ~5Å for the majority of crystals. A separate secondary screen was designed (Table 4.7), based on the secondary screen conditions conducted for the native *P. sativum* enzyme (Table 4.1) and one resulting crystal allowed data collection to 3.8Å at the ESRF,Grenoble.

A	B	C	D	E	F	
15% Butandiol	15% Glucose	15% Butandiol	15% Glucose	15% Butandiol	15% Glucose	1
20% Glycerol	18% Xylitol	20% Glycerol	18% Xylitol	20% Glycerol	18% Xylitol	2
15% Butandiol	15% Glucose	15% Butandiol	15% Glucose	15% Butandiol	15% Glucose	3
20% Glycerol	18% Xylitol	20% Glycerol	18% Xylitol	20% Glycerol	18% Xylitol	4

Table 4.7 Additional cryoprotectant addition to the secondary crystallisation screen conditions devised for the native *P. sativum* ALAD. The condition corresponding to B1 (*0.1M* Mes pH *6.50*, 5% dioxane, 2.4*M* ammonium sulphate with 15% glucose) yielded one crystal that diffracted to 3.8Å at the ESRF, Grenoble.

4.6.3 Counter-diffusion crystallisation techniques

Due to the observed instability of the crystals formed, an alternative method of crystallisation was attempted (see Chapter Two). Counter-diffusion crystallisation techniques have the advantage over vapour diffusion techniques because, once crystals are formed, no direct manipulation is required so that the Lindemann tubes in which the crystals are grown can be used directly on the X-ray beam, frozen or unfrozen, as required (Gavira, 2002; Gavira *et al.*, 2002). The disadvantage of this process is that the initial 'set-up' is incredibly time-consuming and involves the usage of relatively large amounts of protein, crystallisation buffer and cryoprotectant. As yet (within three months) these trials have yielded no crystals.

4.7 Changes in crystal form over time

As has been noted in the initial crystal screens, rectangular plate crystals formed. These crystals appeared to degrade over time and eventually the protein composition of the crystal would precipitate out of solution in the shape of the crystals. However, six months later the same trays were observed to have grown crystals of a cubic form on the perimeter of the precipitated protein. Crystals were initially wet-mounted on the in-house X-ray source to yield data to 3.2Å and further crystals were frozen in 30% glycerol and transported to the ESRF, Grenoble, where they diffracted to 2.8Å. These crystals appeared far more stable than previous forms of the crystal and did not show any damage, either after manipulation with the cryo-loop, or after careful addition of cryoprotectant solution.

4.8 Datasets collected

The datasets were collected at the ESRF, Grenoble. The diffraction data was processed using the CCP4 (Collaborative Computational Project, 1994) suite of programs including MOSFLM (Leslie, 1999), SORTMTZ, SCALA and TRUNCATE (French, 1978) which are described in more detail in Chapter Two. The statistics for the two datasets are displayed below in Tables 4.8 and 4.9.

Statistics for entire dataset	
Resolution range (Å)	42.3-3.8
R_{merge} (%)	9.6
Completeness (%)	99.7
Multiplicity	7.3
Mean I/σ (I)	5.0
Statistics for outer shell	
Resolution range (Å)	4.01-3.80
R_{merge} (%)	37.1
Completeness (%)	99.7
Multiplicity	7.1
Mean I/σ (I)	2.0

Table 4.8 Statistics for the 3.8Å dataset.

Statistics for entire dataset	
Resolution range (Å)	42.3-2.9
R_{merge} (%)	8.0
Completeness (%)	82.8
Multiplicity	3.3
Mean I/σ (I)	7.0
Statistics for outer shell	
Resolution range (Å)	3.06-2.90
R_{merge} (%)	37.5
Completeness (%)	87.2
Multiplicity	3.3
Mean I/σ (I)	1.9

Table 4.9 Statistics for the 2.9Å dataset.

4.9 Molecular replacement

Molecular replacement was attempted using AMORE (Navaza, 1994), BEAST (Read, 2001), MOLREP (Vagin, 1997) and XPLOR (Brunger, 1990). All of the available structures that have been obtained so far were used as search models and, in addition, a similarity model built by alignment of the known structures was obtained from SWISS-MODEL (Guex, 1997), an internet-based tool for comparative protein modelling. Unfortunately, although possible rotation function hits were achieved, none of the programs were able to find translation function hits for the observed data.

Crystallographic screening

Pure selenomethionine-labelled protein (growth, expression and purification described in chapter 2) was used for crystallographic screens including the molecular dimensions screens I and II, and also the predetermined secondary screens used for the unlabelled

protein. Identical hits were observed using the molecular dimensions screens I and II as those found using the unlabelled protein, but viable crystals were more fragile than those previously obtained with unlabelled protein and twinning was observed (Figure 4.8a). Crystals obtained from the secondary screen conditions (Figure 4.8b) were frozen and mounted at the ESRF, Grenoble, and a wavelength absorption scan undertaken (Figure 4.9), which showed a textbook example of the wavelength spectrum of the absorbance-edge expected from selenomethionine labelled protein. Unfortunately, the crystals did not diffract below 7Å, and therefore no dataset was collected.

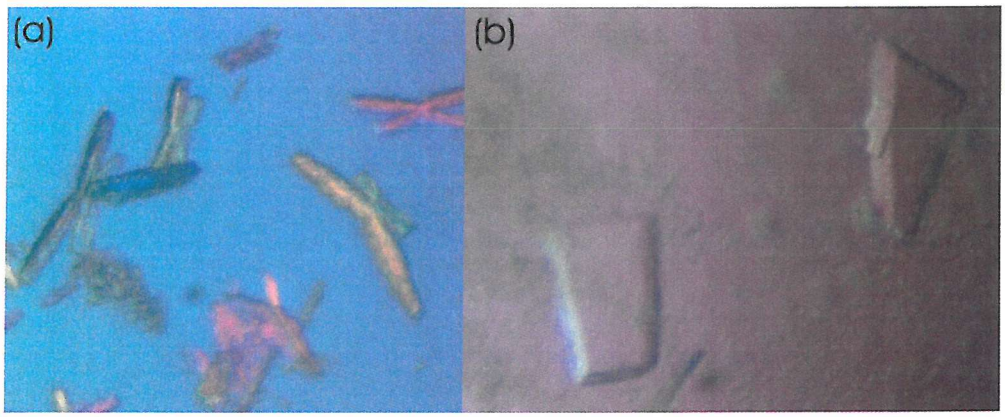


Figure 4.8a showing viable crystals obtained from molecular dimensions screens I and II and Figure 4.8b showing crystals obtained from secondary screen conditions for condition one.

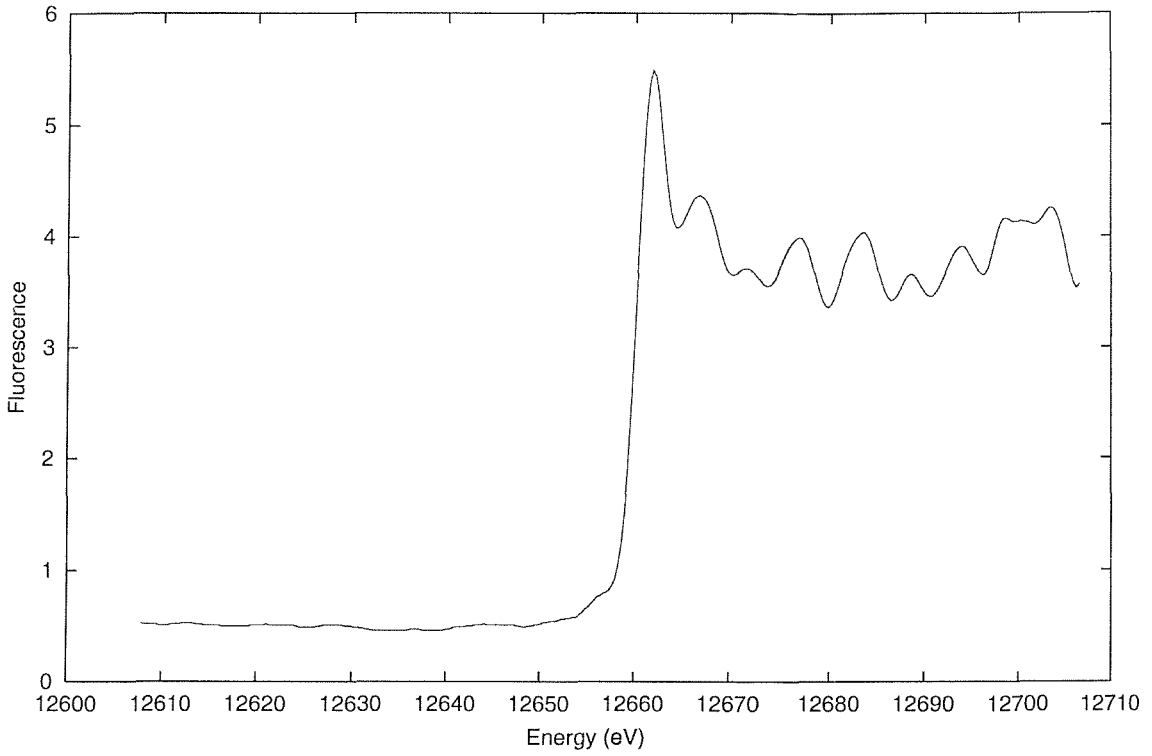


Figure 4.9 The absorption spectrum for crystals of the selenomethionine derivative of *P. sativum* ALAD. The graph shows the energy scan for the crystallised selenomethionine derivative of *P. sativum* ALAD with an absorption edge at 12660eV, as would be expected for a protein with selenomethionine incorporated.

4.10 Discussion

In conclusion, conditions for the crystallisation of *P. sativum* ALAD have been obtained and refined to allow the consistent production of protein crystals.

Attempts have been made to further refine these conditions and improve crystal packing through the use of such techniques such as temperature modifications, modifying the interface of the mother liquor in the hanging drop method of crystallisation, through addition of ligands to the crystal, through site directed mutagenesis studies and additional pH screening to obtain different crystal forms. However, none of these techniques appeared to improve the overall packing of the crystals and the diffraction data obtainable. Nevertheless, the extended pH condition screening from pH 6.00 to pH 9.25 has yielded some very interesting results. It has been proved that crystals can be grown over a large pH

range and that the crystal forms observed vary over that pH range. Three crystal forms were observed to grow under these conditions and these results may well be implemented in the crystallisation of other proteins in order to study the effects that different pH values have on protein structure. This may also allow better crystallisation of proteins from which it is difficult to obtain diffraction quality crystal data.

An additional method of protein crystallography has been utilised (counter-diffusion) so as to obtain crystals with minimal handling to avoid undue damage to crystals but unfortunately, at the time of writing, no crystals were forthcoming.

Finally, large "chunky" crystals were observed to form after a six-month period which yielded diffraction data to 2.9Å. Unfortunately phasing of the data was not possible utilising either of four of the available molecular replacement programs at present – AMORE (Navaza, 1994), BEAST (Read, 2001), MOLREP (Vagin, 1997) and XPLOR(Brunger 1990). During the studies, structures of ALADs from *E. coli*, yeast, *P. aeruginosa* and human were all used as models for the molecular replacement of the data in addition to a model predicted by the SWISS-MODEL on-line structure prediction programs. As more structures become available, the phasing problem for *P. sativum* ALAD can, possibly, in time, be overcome.

To solve the phase problem, a selenomethionine-labelled protein growth and overexpression protocol was adapted for use with this enzyme and subsequent purification and crystallisation trials of the protein yielded promising crystals that were frozen and taken to the ESRF, Grenoble for analysis. Despite showing an absorption wavelength spectrum of the absorbance-edge expected from selenomethionine labelled protein, the crystals did not diffract below 7Å.

It appears that the crystals may require an extended time scale (~6 months) so as to form crystals similar to those observed with the unlabelled protein. This could reflect slight differences in protein solubility or indicate a change in quaternary structure over the time of

the crystal studies. If the protein sample at the start of the crystallisation process is an equilibrium mixture between octameric and hexameric forms of the protein, then it is understandable that the poor data collected could result from poor protein packing, due to heterogeneous quaternary structures. The formation of the better crystals after six months could be a result of lowering protein concentration in the drop and therefore a shift in the equilibrium of the octamer resulting in crystals with better packing formed predominantly from a hexameric form of the enzyme. Additionally this would explain our inability to gain a molecular replacement for the datasets obtained.

It would therefore be interesting to investigate further whether the enzyme has concentration-dependence in this manner and, if so, determine ideal concentrations of the protein for homogeneity of a quaternary species before continuing crystal trials. If ideal conditions could be found, this could allow for better crystal packing and therefore greatly improve the datasets available.

Chapter Five: Design and testing of knowledge based inhibitors of ALAD

5.1 Kinetics of the inhibition of ALAD

As has been mentioned, ALAD catalyses the condensation of two identical substrates and therefore belongs to the class of two substrate enzymes. The kinetics of this class is obviously more complex than that of single substrate enzyme but, nonetheless, follows the laws of Michaelis-Menten kinetics. The first person to discuss this problem was Granick in (1958), proposing the following three conditions in order to explain this kinetic behaviour: the two substrates are binding the enzyme in a predetermined sequence; the dissociation constant of the first enzyme-substrate complex is considerably lower compared to the second and the final step, the release of PBG, is the rate determining step. Accepting these conditions, the formation of PBG does indeed follow Michaelis-Menten kinetics, but it is important to remember that following these rules we are only measuring kinetic data for the second substrate molecule and the subsequent transformation to the product.

5.2 Structural requirements for substrate recognition

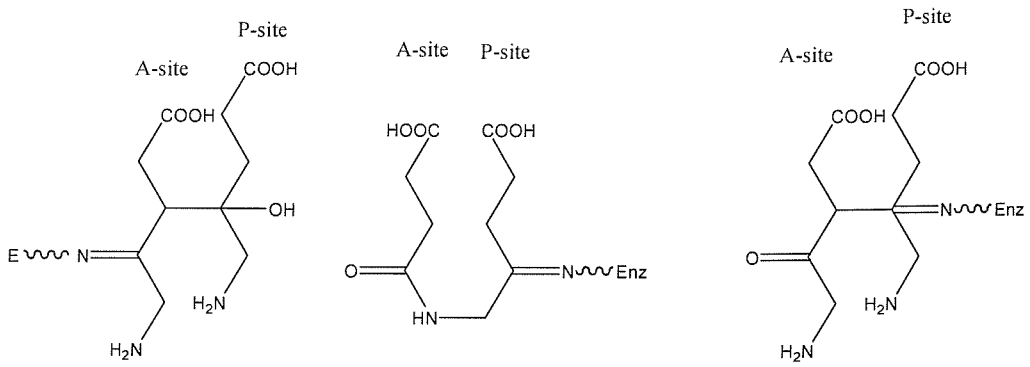
Previously conducted inhibition studies of substrate analogues have shown the carboxylate function at position 1 to be very important for recognition within the active site in conjunction with the keto group at postion 4.. Additional substituents at the position C2 and C3 are tolerated. Variations at position 5 are tolerated. The amino or ammonium group at C5 is not an essential element for recognition (Appleton *et al.*, 1998). This is in surprising contrast to other enzyme-catalysed reactions, where the presence of the charged ammonium group is crucial for the recognition process. Neutral and polar substituents can be used to replace this amino group, but switching to a negatively charged carboxylate or to a sterically demanding group like a monophthalimide results in inhibitors with low affinity (Stauffer *et al.*, 2001; Frere *et al.*, 2002)

The important difference between species of ALAD is the type of metal ion bound at the active site, either Zn^{2+} or Mg^{2+}, and it is thought that the metal ion bound can determine the

susceptibility of different species to different inhibitors. This could be very important especially for agrochemical and pharmaceutical trials where selective inhibition is imperative for the viability of a product.

5.3 Mechanisms of ALAD catalysis

Different mechanisms have been proposed for the Knorr type reaction that occurs between the two identical substrate molecules to form porphobilinogen. The mechanism must explain the dichotomy between the chemical condensation and the biochemical transformation. Several mechanisms have been proposed, Nandi and Shemin (1968) proposed a C-C bond formation, via an aldol type reaction as the central step for the biosynthetic mechanism. Shoolingin-Jordan has proposed two mechanisms: one whereby the first step of the biosynthetic mechanism involves a C-N bond forming, again via an aldol-type reaction (Jordan, 1991) and another whereby a C-C bond is formed in a similar fashion to that proposed by Shemin (Jordan and Seehra, 1980). These three mechanisms would produce three significantly different intermediates as shown in Figure 5.1.

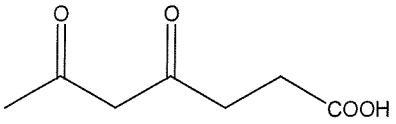


Shemin *Shoolingin-Jordan I* *Shoolingin-Jordan II*

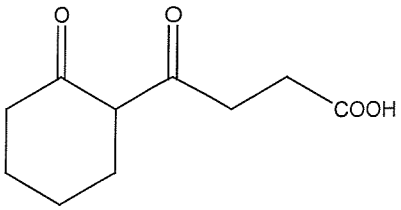
Figure 5.1 Postulated intermediates for three different mechanisms for ALAD. The intermediate proposed by Shemin, whereby the enzyme initially binds the A-site substrate followed by the formation of the C-C bond, and the intermediates proposed by Shoolingin-Jordan, whereby the enzyme binds the P-site substrate and the reaction occurs either by the formation of the C-C or C-N bonds.

5.4 Previous work on inhibition of ALAD

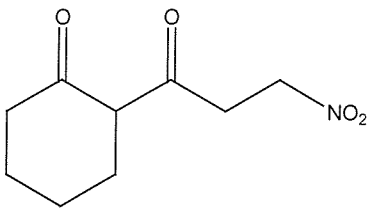
Up until now work has mainly concentrated on either substrate or product analogues which in general show competitive inhibition. This competitive inhibition has been attributed to direct competition between the substrate and inhibitor. The best known inhibitor observed is succinyl acetone which shows a non-competitive or mixed inhibition (K_i=1.4mM) (Neier, 1996). Almost all compounds with a similar structural element to succinyl acetone, containing a 4-keto acid, show a non-competitive inhibition (Jarret *et al.*, 1998). Only two of these compounds, with a similar structural element to succinyl acetone, have shown competitive inhibition (see Figure 5.2).



(1) Succinyl acetone, K_i=1.4mM



(2) 4-Oxo-4-(2-oxocyclohexyl)-butyric acid K_i=318mM



(3) 2-(3-Nitropropionyl)-cyclohexanone, K_i=16.5mM

Figure 5.2 Derivatives of succinyl acetone forming kinetically different inhibitors. Introduction of one nitro group in place of the carboxylate group induces the switch between non-competitive and competitive inhibition(Jarret *et al.*, 1998).

5.4.1 Non-competitive inhibition

It is assumed that non-competitive inhibitors interact equally with the A-site and the P-site of the enzyme. In order that the reaction with the P-site is kinetically plausible, the Schiff base formed between the inhibitor and the P-site lysine has to be more stable than, or as stable as, the Schiff base formed between the substrate and the active site lysine. Under these conditions one would be expected to observe the mixed inhibition of the free enzyme:

competitive inhibition of the enzyme linked with substrate and competitive inhibition of the free enzyme itself. Inhibitors 2 and 3 shown in Figure 5.2 are on the structural threshold between competitive and non-competitive inhibition. The important modification being the replacement of the carboxylate with the nitro group, in accordance with the hypothesis that the nitro group is better recognised by the A-site of the active site (Staufferr *et al.*, 2001)

5.4.2 Postulated intermediates

Diacids containing a keto function have been synthesised with extensions of the chain length connecting the two diacids (Table 5.1).

It has been proposed that up until a certain chain length is reached that inhibition is competitive and that the compounds are recognised as analogues of the substrate in the A-site of the active site. After the chain length is exceeded, it is assumed that the length of the molecule is sufficient to reach into both sites simultaneously and therefore inhibit non-competitively, interact by hydrophobic interaction or form a dead-end complex as an intermediate. A chain length of 9 carbons indicates a perfect length for total inhibition.(Erskine *et al.*, 2001; Jaffe, *et al.*, 2002)

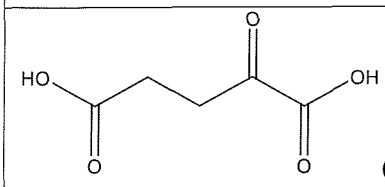
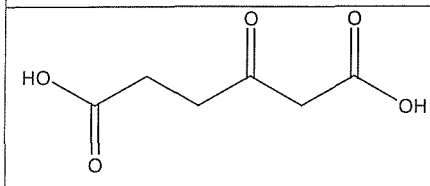
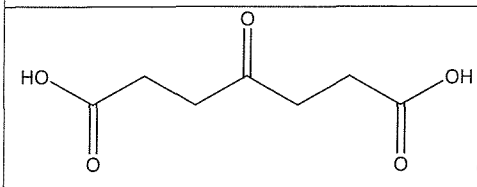
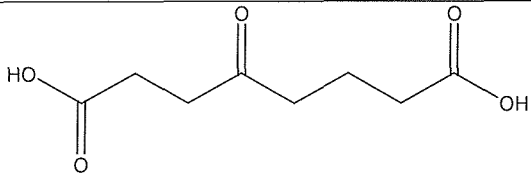
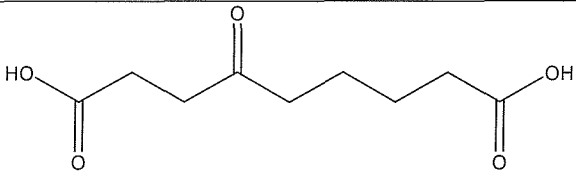
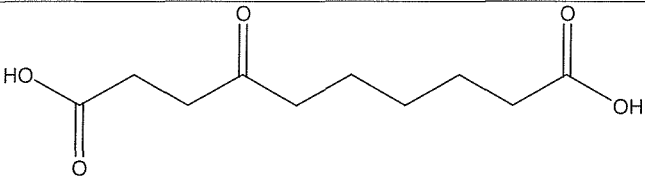
Structure	Ki	Type of inhibition
 (a)	8450mM	Competitive
 (b)	10400mM	Competitive
 (c)	8600mM	Competitive
 (d)	82mM	Non-competitive
 (e)	449mM	Non-competitive
 (f)	(-)	Inactivator

Table 5.1 Compounds with systematic increases in the chain length connecting the two diacids have been synthesised and tested.(Jarret *et al.*, 1998; Jarret *et al.*, 2000)

4-Oxosebacic acid (OSA)(**f**) and 4,7-oxosebacic acid (4,7-DOSA) (Figure 5.3) exhibit a time and concentration-dependant inhibition of ALAD. These compounds have also been tested for sensitivity with ALADs from *B. japonicum, P. aeruginosa, and P. sativum.* Reduced sensitivity to 4-OSA in comparison to 4,7-DOSA were noted for all of these

species. However, inactivation of *E. coli* ALAD occurs with 4-OSA whereas little or no inactivation occurs with these other species. Results have shown that 4-OSA exhibits a high selectivity for *E. coli* ALAD whereas 4,7-DOSA inhibits ALAD regardless of species.(Erskine *et al.*, 2001; Jaffe *et al.*, 2002)

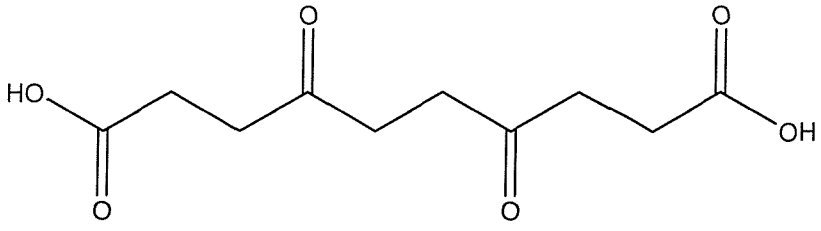


Figure 5.3 The structure of 4,7-oxosebacic acid (DOSA). DOSA is a derivative of 4-OSA with an additional keto group at position 7.

E. coli ALAD has been crystallised with 4-OSA and 4,7-DOSA. In both cases models show C1-C5 of the inhibitor to be bound with a Schiff base linkage between C4 and the amino group of Lys 247 and, similarly, to other structures with inhibitors bound in the P-site, the C1 carboxyl oxygen of the inhibitor makes hydrogen bonds with Ser 272 and Tyr 311.(Jaffe *et al.*, 2002)

In the X-ray structures the positions of the C2 and C3 of each inhibitor are not equivalent, mimicking the reported alternate confirmations of laevulinic acid bound at the P-site. The position determined of C1-C4 of 4-OSA closely resembles that of bound PBG. The bond between 4,7-DOSA C5 and C6 has a distorted *cis*-configuration in order to accommodate the second Schiff base linkage between C7 and the amino group of Lys 194. No such bond exists with 4-OSA. Both the configuration around C5 and C6 and the positions of C5-C10 are significantly different, suggesting that the A-site has more positional flexibility than that of the P-site.(Jaffe *et al.*, 2002)

This positional flexibility is essential to accommodate the multiple adaptations to the active site structure during the dimerisation of ALA to form PBG. In the case of both inhibitors the A-side ALA half, C8-C10, extends towards the lid region of the *E. coli* monomer, where extensive hydrogen bond linkages between the C10 carbonyl oxygens and arginine

residues occur. It has been proposed from this data that the observed multiple hydrogen bonds between the C10 carboxyl (equivalent to the A-side ALA) and Arg 204, Arg 215, Gln 219 and associated water molecules change confirmation to accommodate various positions of the A-side ALA during the course of the catalysed reaction.(Kervinen *et al.*, 2001; Jaffe *et al.*, 2002)

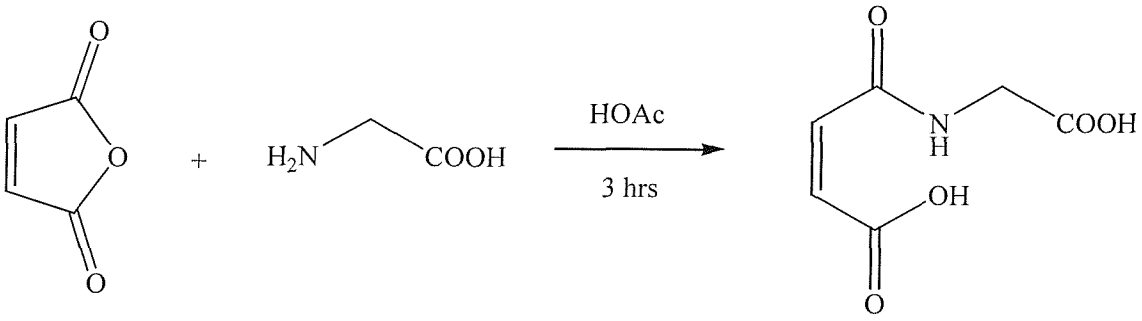
Although the basic features of bond making and breaking reactions are still unknown and may not be conserved between species, the structures incorporating 4-OSA and 4,7-DOSA have allowed a better understanding of the potential role of the two active site lysines and other groups in their vicinity through the course of the reaction. Also, it has been suggested that the active site water molecules may play an important role in the reaction scheme, and show mechanistic importance in fluctuations between multiple conformations of the active site lid.(Erskine *et al.*, 2001; Kervinen *et al.*, 2001; Jaffe *et al.*, 2002)

5.5 Synthesis of compounds as inhibitors of *P. sativum* ALAD

Compounds synthesised at Syngenta ((Jeallott`s Hill, Bracknell, UK).

Using the *E. coli* structure as a model, various compounds were modelled into the active site cleft of the enzyme using a standard modelling programme and were subsequently synthesised in collaboration with Syngenta (Jeallott`s Hill, Bracknell, UK). All of the compounds (described under headings 5.5.1-5.5.11) were synthesised by myself following the advice of colleagues in the labs at Syngenta and routine protocols.

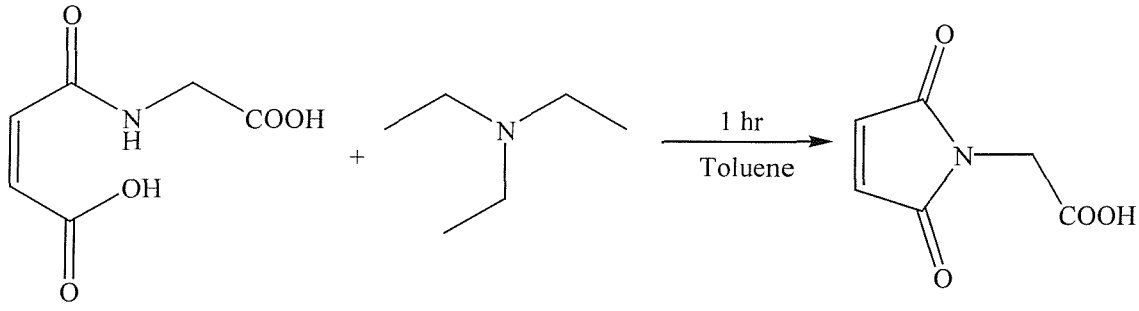
5.5.1 The synthesis of maleamic acid.



Method:

Maleic anhydride (4.17g, 0.0425mol) was dissolved in dry acetic acid (HOAc) (17.5mL) and mixed with glycine (3.19g, 0.0425mol) in HOAc (51mL). Reactants were mixed at room temperature for 3 hrs, over which time a white precipitate was observed to form. The white precipitate (maleamic acid) was filtered from the reaction mixture and washed with cold H_2O and dried.

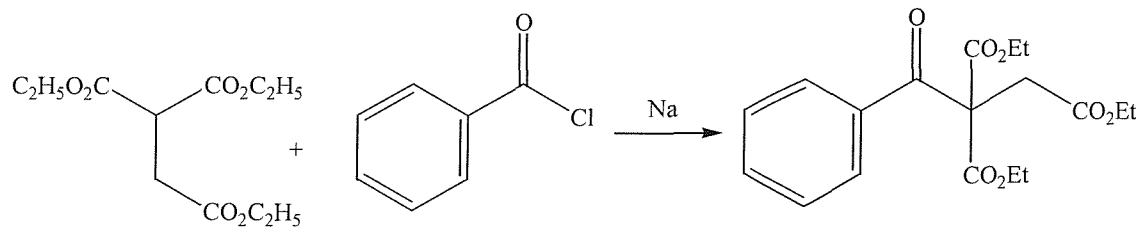
5.5.2 The synthesis of (2,5-dioxo-2,5-dihydropyrrol-1-yl)-acetic acid.



Method:

Maleamic acid (2.91g, 16.8mmol) was dissolved in dry toluene (500mL) and treated with triethylamine (3.55g, 35.1mmol). Reactants were refluxed with vigorous stirring for 1hr with removal of H_2O in a Dean-Stark apparatus. The reaction was then left to cool to room temperature. The toluene was then decanted from the resulting orange oil and the solution was concentrated under reduced pressure to give a white solid (the triethylammonium salt). The solid was then dissolved in 50mL H_2O and extracted with ethyl acetate (EtOAc) and dried over anhydrous $MgSO_4$. The resulting solution was concentrated under reduced pressure to give (2,5-dioxo-2,5-dihydropyrrol-1-yl)-acetic acid, in the form of a white solid. The product was then purified on a silica column by LC and the purity determined by NMR and MS. Final yield = 257mg, 1.6mmol, 10%.

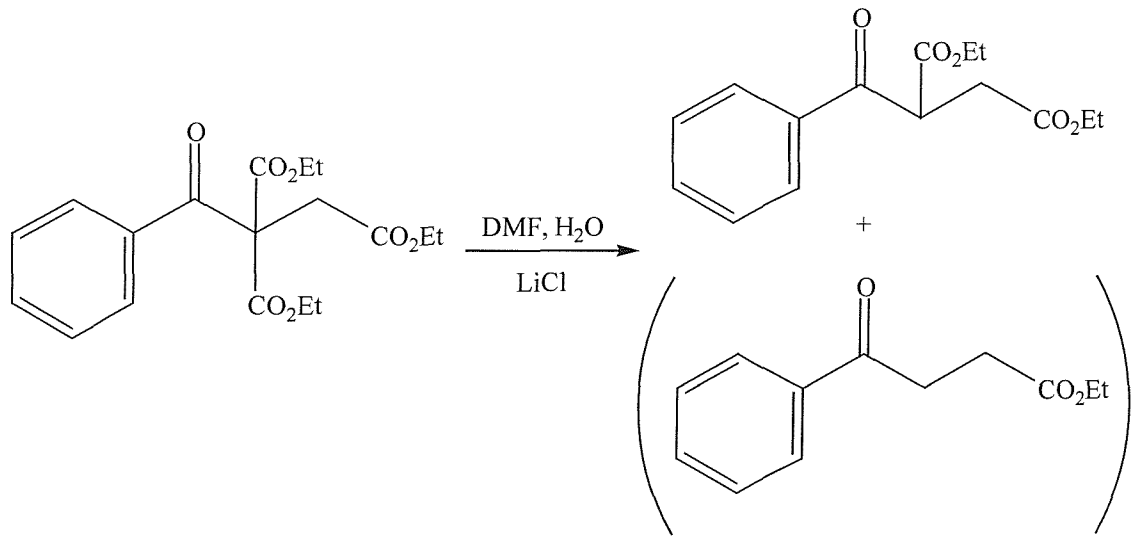
5.5.3 The synthesis of 2-benzoyl-2-ethoxycarbonylsuccinic acid diethyl ester.



Method:

Sodium metal slivers were added to triethyl 1,1,2-ethanetricarboxylate (1.45g, 10mmol) in dry tetrahydrofuran (THF) (50mL), which was then mixed with benzoyl chloride (2.46g, 10mmol) in dry THF (2mL). The reaction was mixed for two hours at room temperature, after which time the Na slivers were still visible. The mixture was therefore refluxed for 2hrs to dissolve the Na slivers fully, and the reaction mixture was then allowed to cool overnight. The resulting mixture was concentrated under reduced pressure and the products were partitioned between ether and H_2O. The layers were separated followed by a further partition of ether. The organic fractions were combined and backwashed with saturated brine solution before drying over anhydrous $MgSO_4$. The extract was then concentrated under pressure, to give a yellow oil (3.276g) in a yield of ~69%. Bulb to bulb distillation was conducted at 170°C, under reduced pressure, for 2 hours to yield 79% pure product (2-benzoyl-2-ethoxycarbonylsuccinic acid diethyl ester).

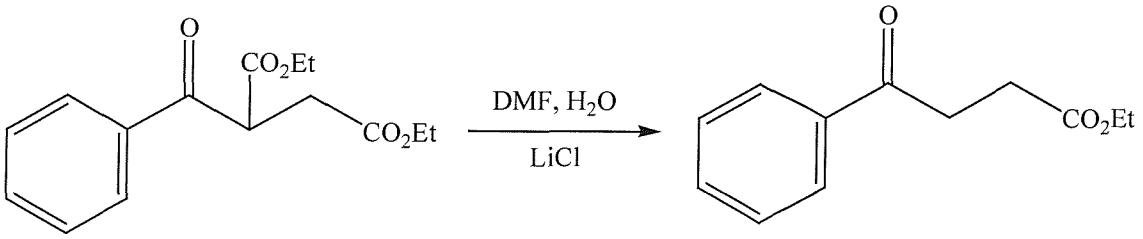
5.5.4 Reaction of 2-benzoyl-2-ethoxycarbonyl-succinic acid diethyl ester.



Method:

2-benzoyl-2-ethoxycarbonylsuccinic acid diethyl ester (2.6g, 8mmol) was mixed with *N,N*-dimethylformamide (DMF) (40mL), H_2O (576µl) and LiCl (680mg, 16mmol). The reactants were mixed and refluxed for 2hrs at 150°C and then analysed by GC. The product mixture was concentrated under low pressure, followed by washing it four times with H_2O, each time removing the aqueous phase and then backwashing with sodium bicarbonate and brine solution sequentially. The solvent layer was then dried over anhydrous $MgSO_4$ and concentrated under reduced pressure to yield a brown oil, which was analysed by TLC and NMR. There was an indication of a possible mixture of the carboxylate and ester intermediates. Therefore, the concentrated sample was separated on a silica column by LC, and the eluted fractions were analysed by TLC. Desired fractions were pooled and concentrated to yield 740mg of a product/intermediate mix.

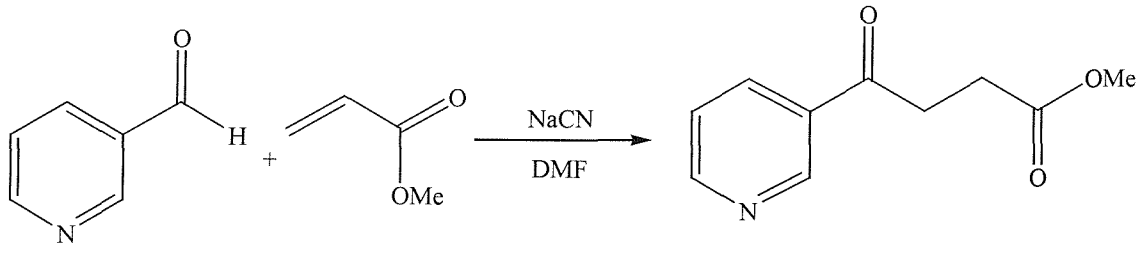
5.5.5 The synthesis of ethyl 4-oxo-4-phenylbutanoate.



Method:

The product from reaction 5.5.4 was, for the purposes of the reaction, assumed to be pure intermediate 2-benzoyl-succinic acid diethyl ester. 2-benzoyl-succinic acid diethyl ester (740mg, 2.7mmol) was dissolved in DMF (10mL), H_2O (93.6µl) and LiCl (110mg, 2.6mmol). The reactants were mixed and refluxed for 4 hrs at 150°C, after which time the reaction was left overnight to cool. The mixture was then concentrated under reduced pressure, and the concentrated mixture was treated with 30mL of ethyl acetate, washed three times with H_2O and once with brine, removing the aqueous layer after each wash. The solvent phase was then dried over anhydrous $MgSO_4$ and concentrated under reduced pressure to yield 343mg (9% yield) of pure product.

5.5.6 The synthesis of 4-oxo-4-pyridin-3-yl-butyric acid methyl ester.



Reference:

Patent WO 2000-2000JP8090, US 99-450245

"Preparation of 3-(2-aminoethylthio)methyl-4-oxo-4-(3-pyridyl)butanoic acid derivatives as neuroprotective agents." Bhaguat (2001)(Bhaguat 2001).

Method:

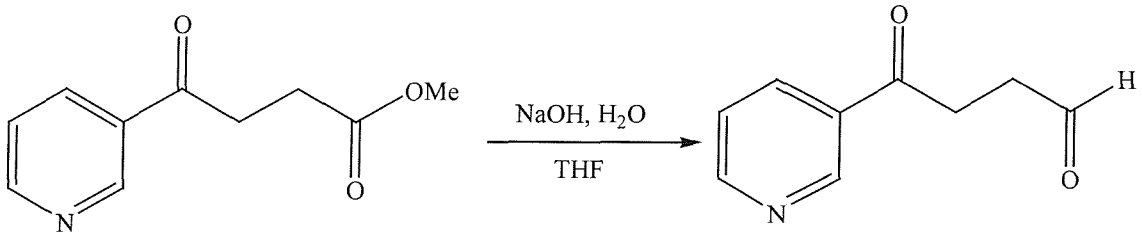
3-Pyridinecarboxyaldehyde (10.7g, 100mmol) was dissolved in DMF (20mL) and added over 10 minutes to a stirred solution of sodium cyanide (2.45g, 50mmol) in DMF (80mL) at room temperature. An exothermic reaction occurred with a colour change from yellow to red. The resulting solution was then stirred for 30 minutes, after which time methyl acrylate (8.6g, 200mmol) in DMF (80mL) was added dropwise over 50 minutes whilst continuing to stir the mixture. The reaction was then allowed to stir for a further 3 hours during which a colour change from red to brown occurred. After this time, acetic acid

(6.6mL) and water (30mL) were added after this time and stirring was continued for 5 minutes.

The mixture was then concentrated under reduced pressure, dissolved in water (360mL), and the pH was adjusted to 7.5 with solid sodium bicarbonate. The aqueous layer was extracted with chloroform (3 x 300mL). The chloroform layers were combined, washed with brine solution, dried over anhydrous $MgSO_4$, filtered and then concentrated under reduced pressure. This process yielded 16.6g of a brown liquid (crude extract).

The crude extract was then purified by LC on silica column using 50% ethyl acetate in hexane. Fractions were analysed by GC, and those over 90% pure were pooled and concentrated to yield 407mg, 2.1mmol, a total yield of 2.1% of 4-oxo-4-pyridin-3-yl-butyric acid methyl ester).

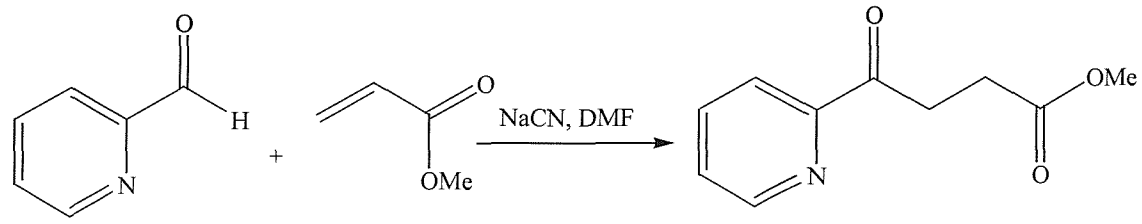
5.5.7 The demethylation of 4-oxo-4-pyridin-3-yl-butyric acid methyl ester



Method:

Product 3.1 (100mg, 0.5mmol) was dissolved in THF (5mL) and mixed with NaOH (83mg, 2mmol) in H_2O (1mL). The reactants were mixed for 30 minutes. The reaction mixture was acidified to pH5 using 5% citric acid solution and extracted with EtOAc (~20mL). The aqueous mixture was then further acidified to pH 3 and extracted with 2 x EtOAc (2 x ~20mL).

5.5.8 The synthesis of 4-oxo-4-pyridin-2-yl-butyric acid methyl ester.

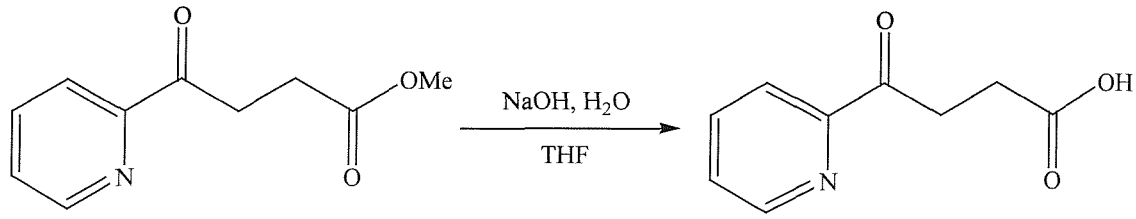


Method:

2-Pyridinecarboxyaldehyde (10.7g, 100mmol) in DMF (20mL) was added over 10 minutes to a stirred solution of sodium cyanide (2.45g, 50mmol) in DMF (80mL) at room temperature (an exothermic reaction with a yellow to red colour change). The solution was stirred for a further 30 minutes before addition of methyl acrylate (8.6g, 100mmol) in DMF (80mL) dropwise over 50 minutes. The reaction was stirred for a further 3 hours (with an observed red to brown colour change). Acetic acid (6.6mL) and water (30mL) were added to the mixture and stirring was continued or five minutes. The mixture was then concentrated under reduced pressure and allowed to cool overnight.

The concentrated mixture was dissolved in water (30mL) and the pH adjusted to pH 7.5 with solid sodium bicarbonate. The aqueous layer was extracted with chloroform (3X 250mL). The chloroform layers were combined, washed with sodium chloride solution and dried over anhydrous Na_2SO_4 and left to stand for 30 minutes. The dried chloroform layers were then filtered and concentrated under pressure to give a brown liquid (17.9g). Crystallisation was observed to occur in the crude mixture. Crystals were allowed to form and were filtered from the brown liquid and washed with ether. The crystals were analysed and deemed to be formed from an unwanted by-product of the reaction. The remaining crude extract was purified by LC on a silica column, and this yielded 168mg of 4-oxo-4-pyridin-2-yl-butyric acid methyl ester (0.9% yield).

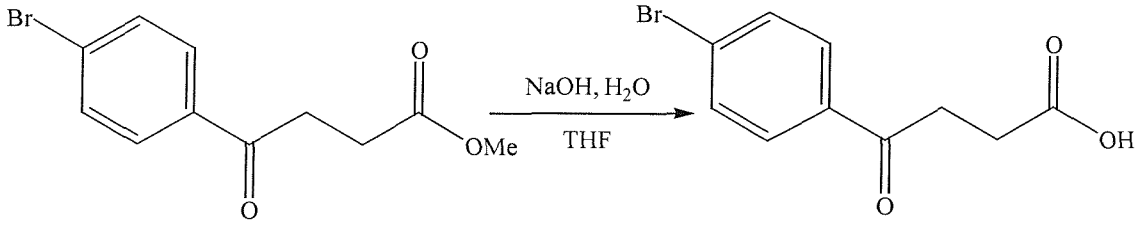
5.5.9 The demethylation of 4-oxo-4-pyridin-2-yl-butyric acid methyl ester



Method:

Product 4.1 (117mg, 0.6mmol) in THF (5mL) was mixed with NaOH (99.6mg, 2.4mmol) in 1mL H_2O and the mixture was stirred for 1½ hours. The reaction mixture was acidified to pH 5 with 5% citric acid and then extracted with EtOAc (~20mL). The aqueous mixture was then further acidified to pH 3 and extracted with 2 x EtOAc (2 x ~20mL). The fractions were then combined, washed with brine solution, dried under anhydrous sodium sulphate, filtered and then concentrated under reduced pressure to yield ~90% pure 4-oxo-4-(pyridin-2-yl)butanoic acid (66mg), a yield of 56%.

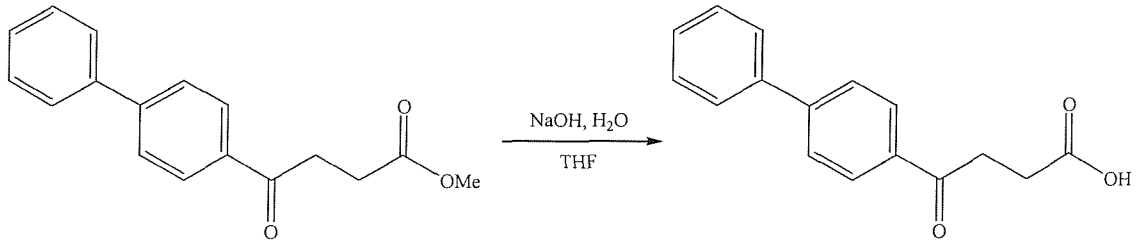
5.5.10 The demethylation of 4-(4-bromophenyl)-4-oxobutyric acid methyl ester



Method:

4-(4-Bromophenyl)-4-oxobutyric acid methyl ester (1.36g, 5mmol) was dissolved in THF (50mL) and mixed with NaOH (8.3mg, 20mM) in H_2O (10mL). The reactants were stirred for 2 hours. The reaction mixture was then concentrated under low pressure and H_2O was added to give a final reaction volume of 50mL, and then extracted with EtOAc (2 x ~50mL). The solvent layers were then combined, dried over anhydrous Na_2SO_4 and then concentrated under low pressure to yield 1.01g (4mmol) of 94% pure product 5.1 at a yield of 74%.

5.5.11 The demethylation of 4-biphenyl-4-yl-4-oxobutyric acid methyl ester

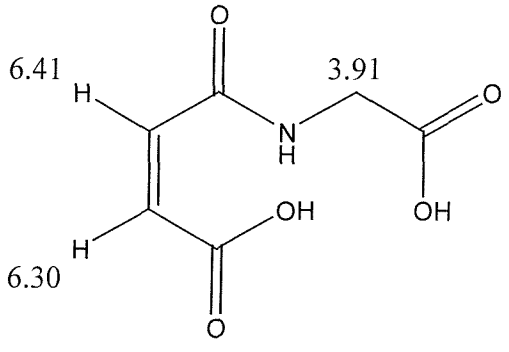


Method:

4-Biphenyl-4-yl-4-oxo-butyric acid methyl ester (1.41g, 5mmol) was dissolved in THF (50mL) and mixed with NaOH (830mg, 20mmol) in H_2O (10mL) for two hours. The reaction mixture was then concentrated under reduced pressure and H_2O added to give a final volume of 50mL, and then extracted with 2 x EtOAc (2 x 50mL). The solvent layers were the pooled, dried over anhydrous Na_2SO_4 and again concentrated under reduced pressure, to give 970mg (4mmol) of 96% pure product 6.1, a yield of 69%.

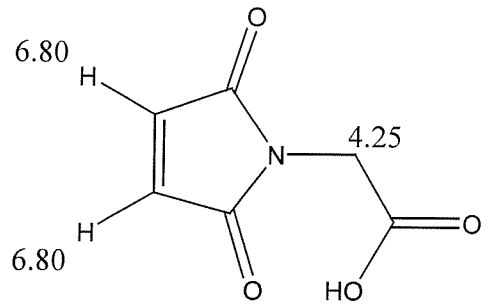
5.6 ^{1}H NMR assignments

^{1}H NMR shifts for maleamic acid



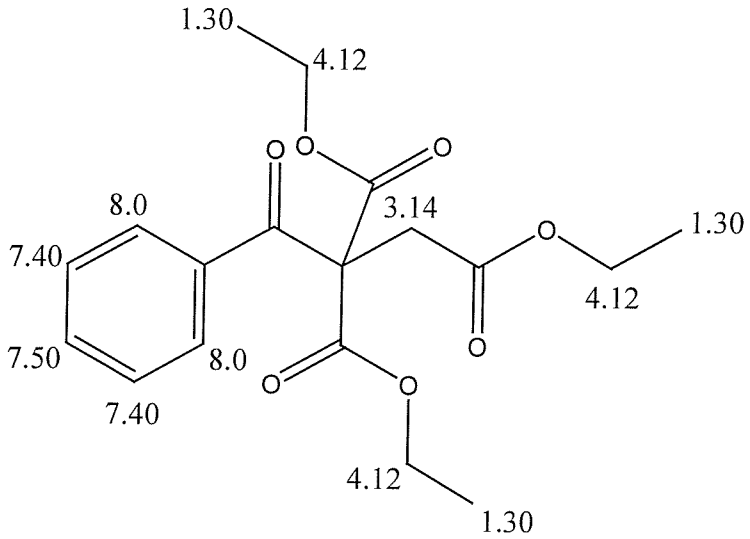
2H, m, δ3.91; 1H, d, δ6.30; 1H, d, δ6,41

^{1}H NMR shifts (2,5-dioxo-2,5-dihydropyrrol-1-yl)-acetic acid



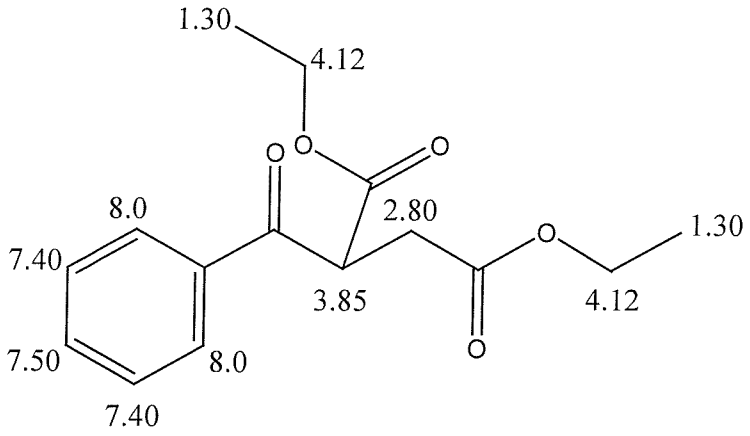
2H, s, δ4,25; 2H, s, δ6.80.

^{1}H NMR Shifts 2-benzoyl-2-ethoxycarbonylsuccinic acid diethyl ester

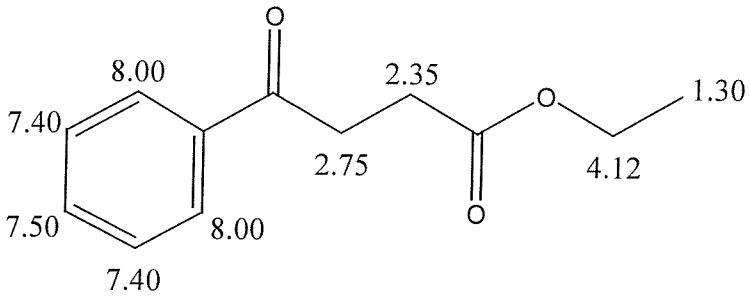


9H, m, δ1.30; 2H, s, δ3.25; 6H, m, δ4.12; 2H, m, δ7.40; 1H, t, δ7.5; 2H, m, δ8.0.

^{1}H NMR Shifts Intermediate and Product Mix from reaction 5.5.4

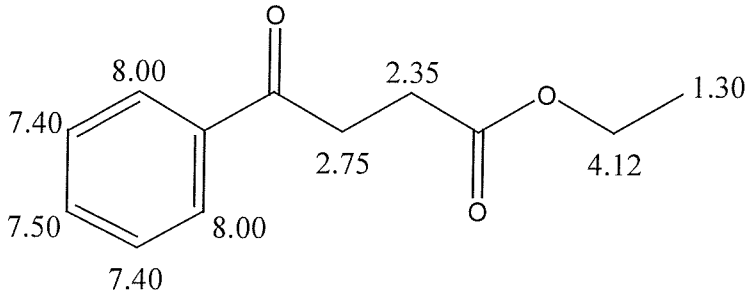


6H, m, δ1.30; 2H, d, δ2.80; 1H, t, δ3.85; 4H, m, δ4.12; 2H, m, δ7.40; 1H, t, δ7.50; 2H, m, δ8.0.



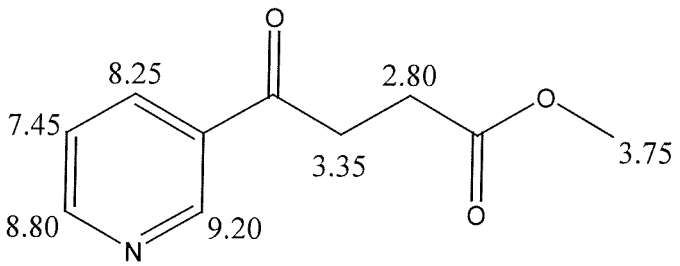
3H, t, 1.30; 2H, t, δ2.35; 2H, t, δ2.75; 2H, m, δ4.12; 2H, m, δ7.40; 1H, t, δ7.50; 2H, m, δ8.0.

^{1}H NMR Shifts ethyl 4-oxo-4-phenylbutanoate



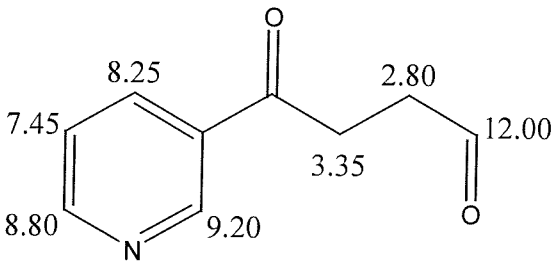
3H, t, 1.30; 2H, t, δ2.35; 2H, t, δ2.75; 2H, m, δ4.12; 2H, m, δ7.40; 1H, t, δ7.50; 2H, m, δ8.0.

^{1}H NMR Shifts 4-oxo-4-pyridin-3-ylbutyric acid methyl ester



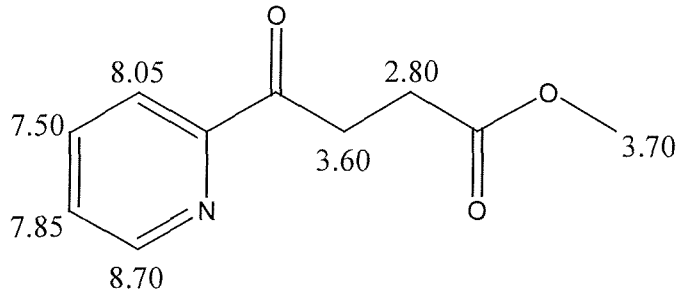
2H, t, δ2.80; 2H, t, δ3.35; 3H, s, δ3.75; 1H, m, δ7.45; 1H, d, δ8.25; 1H, d, δ8.8; 1H, s, δ9.2.

^{1}H NMR Shifts 4-oxo-4-pyridin-3-ylbutanoic acid



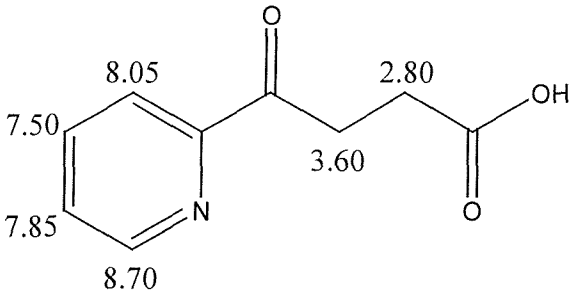
2H, t, δ2.80; 2H, t, δ3.35; 1H, m, δ7.45; 1H, d, δ8.25; 1H, d, δ8.8; 1H, s, δ9.2; 1H, s, δ12.00.

^{1}H NMR Shifts 4-oxo-4-pyridin-2-ylbutyric acid methyl ester



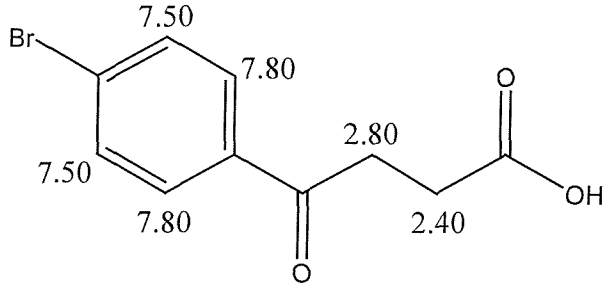
2H, t, δ2.80; 2H, t, δ3.60; 3H, s, δ3.70; 1H, m, δ7.50; 1H, m, δ7.85; 1H, d, δ8.05; 1H, d, δ8.70.

^{1}H NMR Shifts 4-oxo-4-pyridin-2-ylbutanoic acid



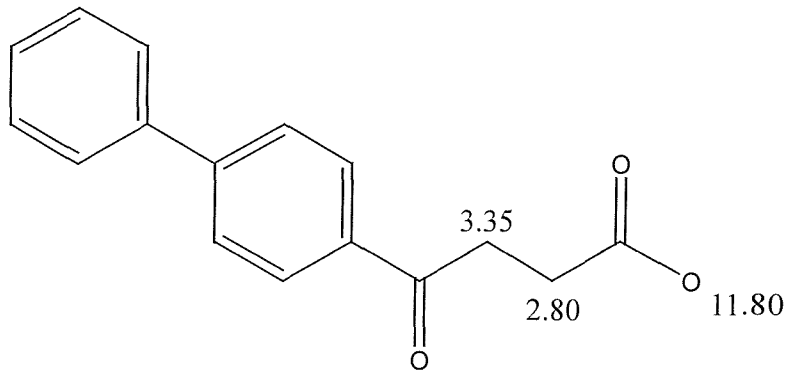
2H, t, δ2.80; 2H, t, δ3.60; 1H, m, δ7.50; 1H, m, δ7.85; 1H, d, δ8.05; 1H, d, δ8.70.

^{1}H NMR Shifts 4-(4-bromophenyl)-4-oxobutanoic acid



2H, t, δ2,40; 2H, t, δ2.80; 2H, d, δ7.80; 2H, d, δ7.50.

^{1}H NMR Shifts 4-biphenyl-4-yl-4-oxobutanoic acid



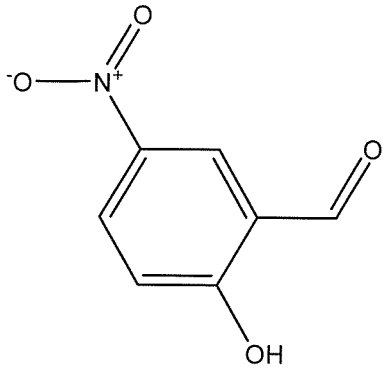
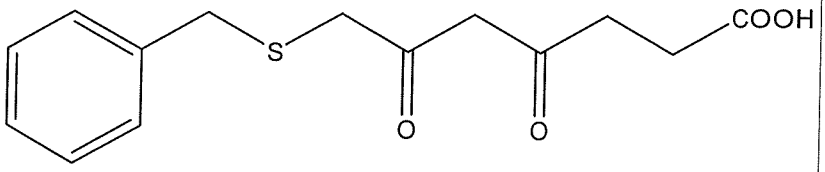
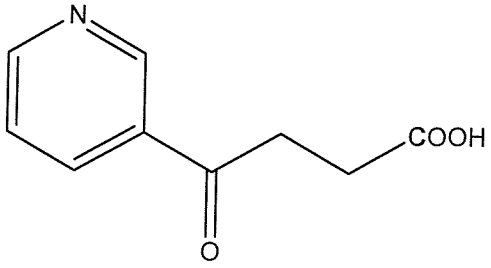
2H, t, δ2.80; 2H, t, δ3.35; 1H, s, δ11.80. At resonances between δ7.4-8.1 multiplets can be seen corresponding to aromatic rings.

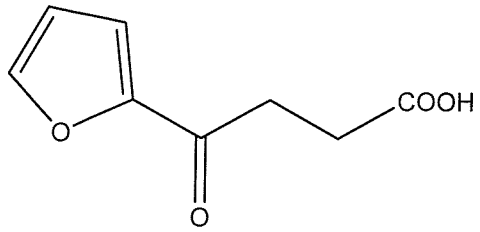
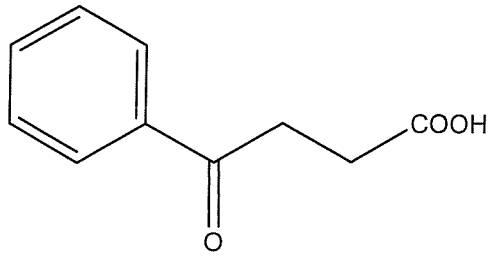
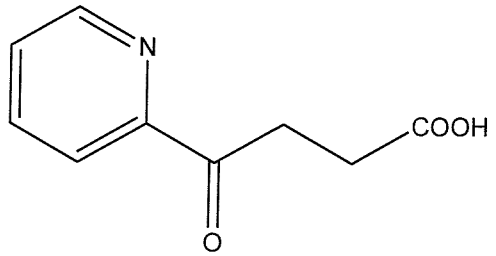
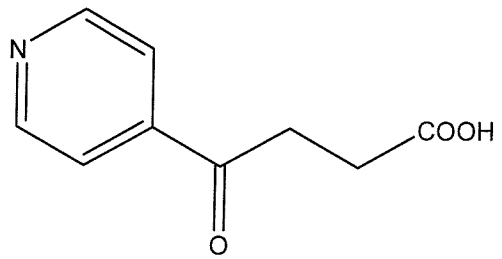
5.7 Compounds synthesised by Dr K-M. Cheung (Southampton UK).

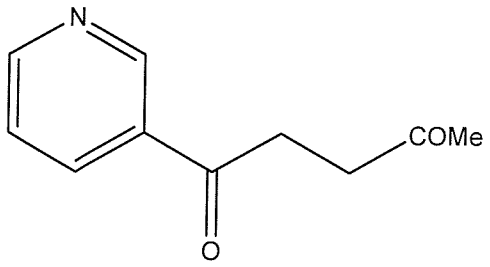
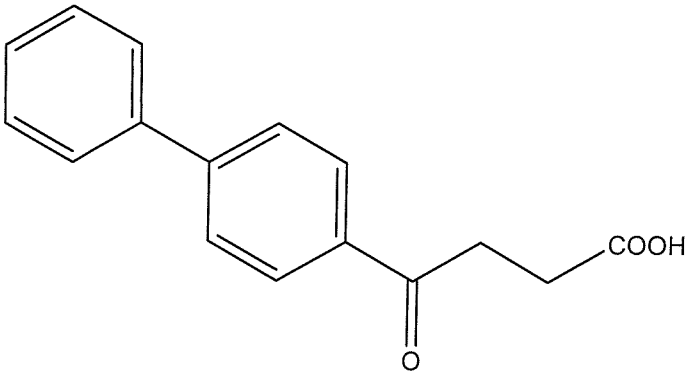
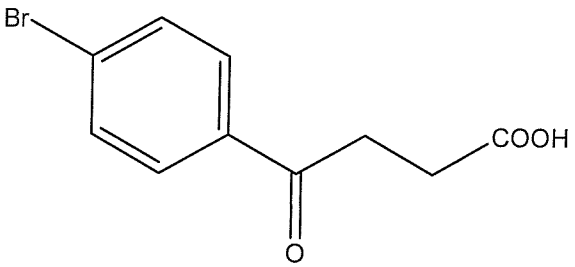
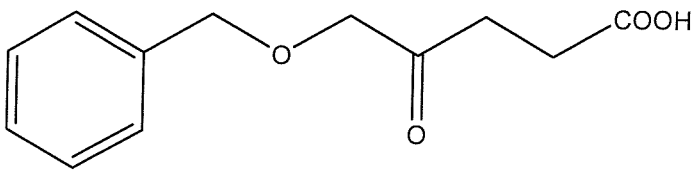
In addition to those synthesised in conjunction with Syngenta (Jeallott's Hill, Bracknell, UK) a library of compounds was also synthesised by Dr K-M. Cheung (Southampton UK).

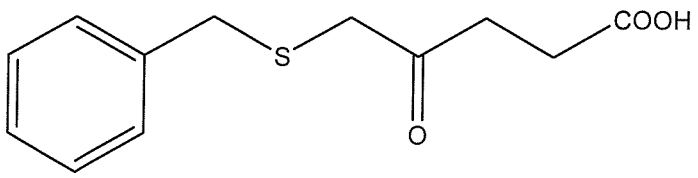
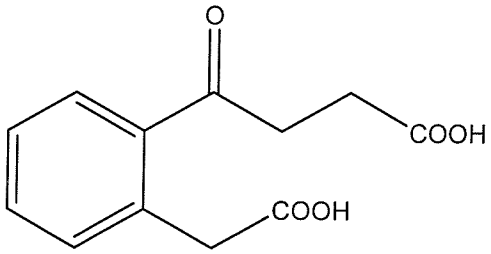
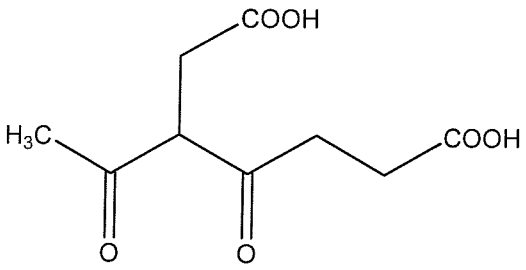
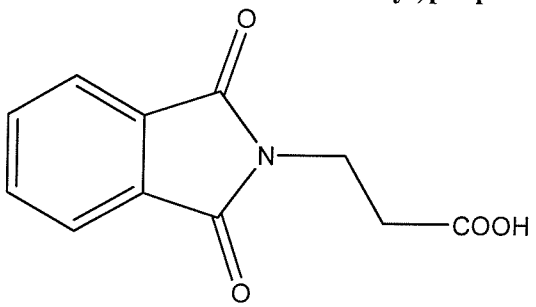
5.8 IC50 estimation

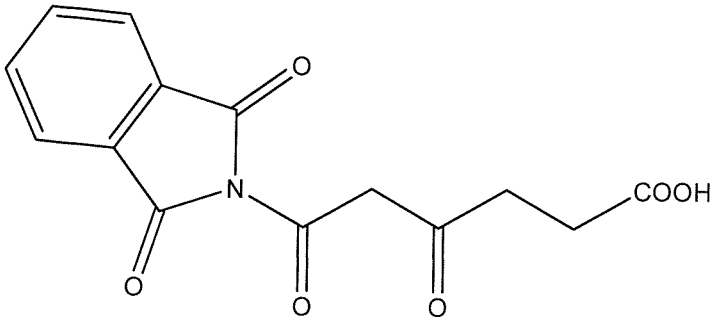
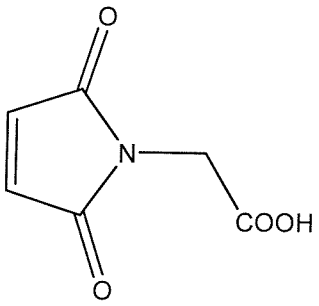
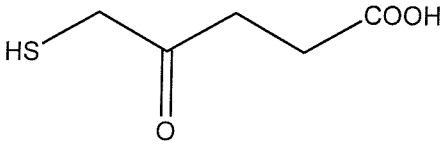
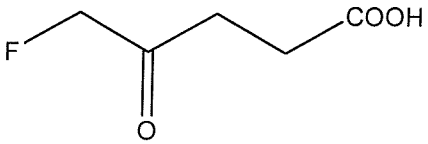
All of the synthesised compounds, including those synthesised by Dr K-M. Cheung (Southampton UK), were tested for inhibition against the *P. sativum* ALAD. Concentrations of 20mM, 2mM, 0.2mM and 0.02mM of the synthesised compounds were incubated at 37°C with the pure enzyme for one hour prior to assay as described in Chapter Two. The resulting data was then corrected against standard curves for the inhibition of the enzyme from the DMSO and methanol solvents used for preparation of the soluble compound solutions. Finally, bar charts were plotted for each compound showing the percentage inhibition of the enzyme by each compound, and the IC50 values calculated (results shown below). Results were qualitatively compared against the inhibitor 2-hydroxy-5-nitrobenzaldehyde (**1**) a known inhibitor of the *P. sativum* ALAD enzyme (Norton, 1999).

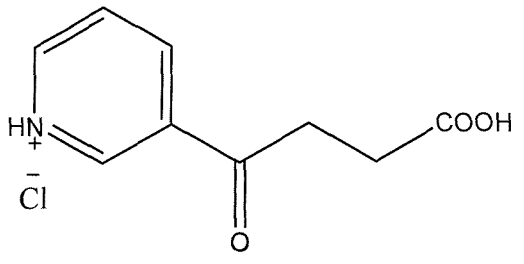
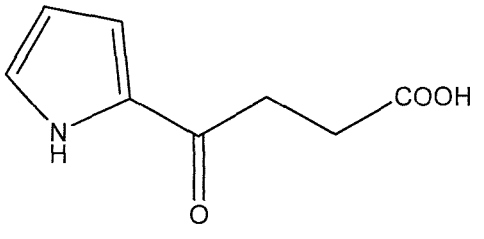
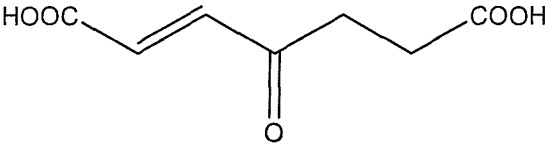
Compound	IC50 value	Source of Compound
(1) 2-hydroxy-5-nitrobenzaldehyde 	132μM	Sigma-Aldrich, UK
(2) 7-(benzylthio)-4,6-dioxoheptanoic acid 	184μM	Dr K-M. Cheung, Southampton, UK
(3) 4-oxo-4-(pyridin-3-yl)butanoic acid 	187μM	Syngenta, Jealott`s Hill, UK

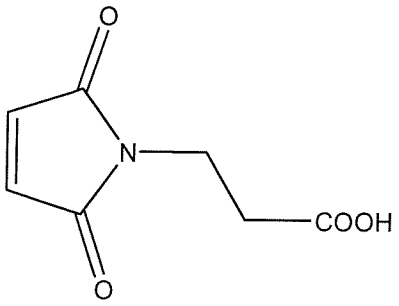
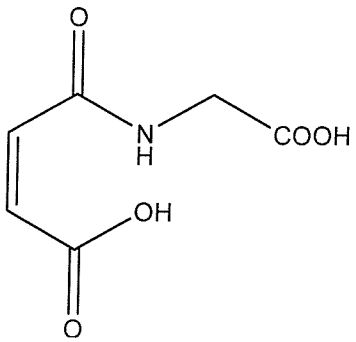
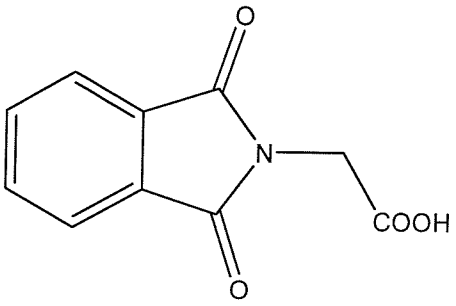
(4) 4-(furan-2-yl)-4-oxobutanoic acid 	253μM	Dr K-M. Cheung, Southampton, UK
(5) 4-oxo-4-phenylbutanoic acid 	283μM	Dr K-M. Cheung, Southampton, UK
(6) 4-oxo-4-(pyridin-2-yl)butanoic acid 	283μM	Syngenta, Jealott`s Hill, UK
(7) 4-oxo-4-(pyridin-4-yl)butanoic acid 	308μM	Syngenta, Jealott`s Hill, UK

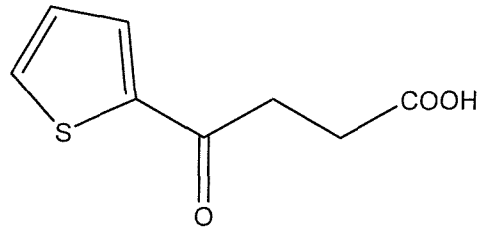
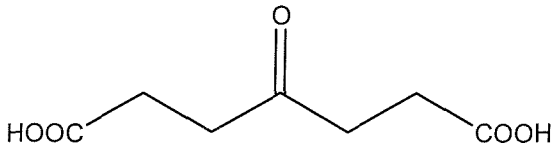
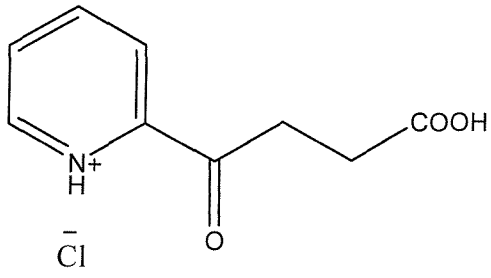
(8) 1-(pyridin-3-yl)pentane-1,4-dione 	331μM	Syngenta, Jealott`s Hill, UK
(9) 4-biphenyl-4-yl-4-oxobutanoic acid 	472μM	Syngenta, Jealott`s Hill, UK
(10) 4-(4-bromophenyl)-4-oxobutanoic acid 	1.6mM	Syngenta, Jealott`s Hill, UK
(11) 5-(benzyloxy)-4-oxopentanoic acid 	1.9mM	Dr K-M. Cheung, Southampton, UK

(12) 5-(benzylthio)-4-oxopentanoic acid 	2.0mM	Dr K-M. Cheung, Southampton, UK
(13) 4-(2-(carboxymethyl)phenyl)-4-oxobutanoic acid 	2.0mM	Dr K-M. Cheung, Southampton, UK
(14) 3-acetyl-4-oxoheptanedioic acid 	2.1mM	Dr K-M. Cheung, Southampton, UK
(15) 3-(1,3-dioxoisoindolin-2-yl)propanoic acid 	2.2mM	Dr K-M. Cheung, Southampton, UK

(16) 4,6-dioxo-6-(1,3-dioxoisoindolin-2-yl)hexanoic acid 	2.5mM	Dr K-M. Cheung, Southampton, UK
(17) (2,5-dioxo-2,5-dihydropyrrol-1-yl)-acetic acid 	3.4mM	Syngenta, Jealott`s Hill, UK
(18) 5-mercapto-4-oxopentanoic acid 	4.1mM	Dr K-M. Cheung, Southampton, UK
(19) 5-fluoro-4-oxopentanoic acid 	4.4mM	Dr K-M. Cheung, Southampton, UK

(20) 4-oxo-4-(pyridin-3-yl)butanoic acid 	4.6mM	Dr K-M. Cheung, Southampton, UK
(21) 4-oxo-4-(1H-pyrrol-2-yl)butanoic acid 	6.0mM	Dr K-M. Cheung, Southampton, UK
(22) (E)-4-oxohept-2-enedioic acid 	8.8mM	Dr K-M. Cheung, Southampton, UK

(23) 3-(2,5-dioxo-2H-pyrrol-1(5H)-yl)propanoic acid 	9.1mM	Syngenta, Jealott`s Hill, UK
(24) Maleamic acid 	11.7mM	Syngenta, Jealott`s Hill, UK
(25) 2-(1,3-dioxoisoindolin-2-yl)acetic acid 	11.8mM	Dr K-M. Cheung, Southampton, UK

(26) 4-oxo-4-(thiophen-2-yl)butanoic acid 	12.0mM	Dr K-M. Cheung, Southampton, UK
(27) 4-oxoheptanedioic acid 	89.0mM	Dr K-M. Cheung, Southampton, UK
(28) 4-oxo-4-(pyridin-2-yl)butanoic acid 	Not inhibitory	Dr K-M. Cheung, Southampton, UK

Although the exact method of binding was unknown, the predicted conformation of the three compounds that gave the best inhibition are shown in Figures 5.4, 5.5 and 5.6.

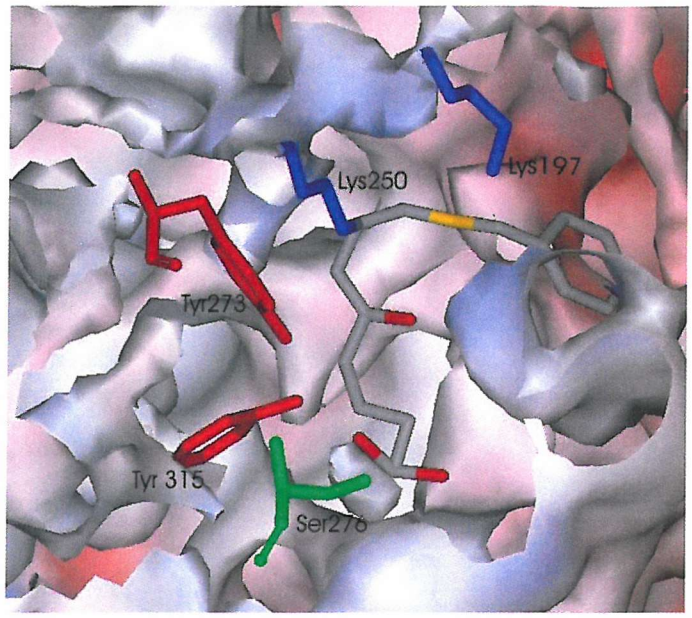


Figure 5.4 The predicted conformation of inhibitor 2 (7-(benzylthio)-4,6-dioxoheptanoic acid) to Lys 250 of the *P. sativum* ALAD subunits. The carboxylic group of the compound can be seen orientated towards Tyr 273, Tyr 315 and Ser 276 of the predicted structure.

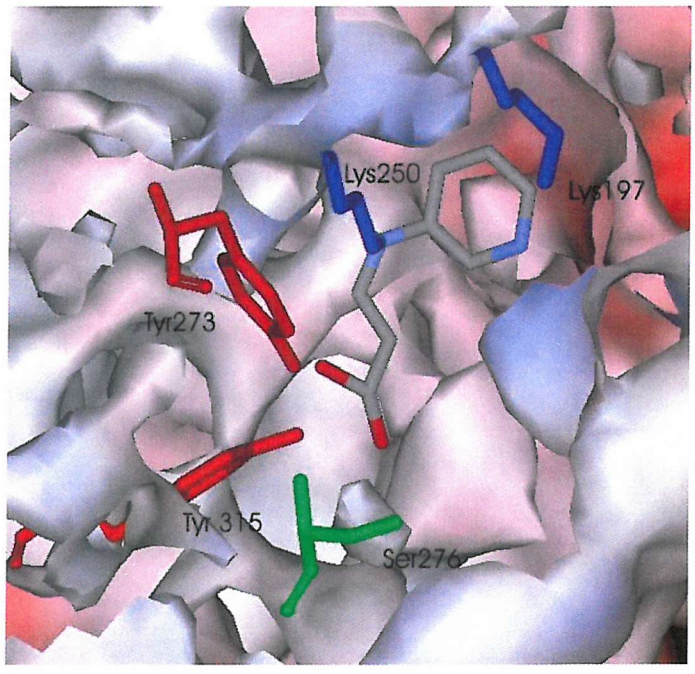


Figure 5.5 The predicted conformation of inhibitor 3 (4-oxo-4-(pyridin-3-yl)butanoic acid) to Lys 250 of the *P. sativum* ALAD subunits. The carboxylic group of the compound can be seen orientated towards Tyr 273, Tyr 315 and Ser 276 of the predicted structure.

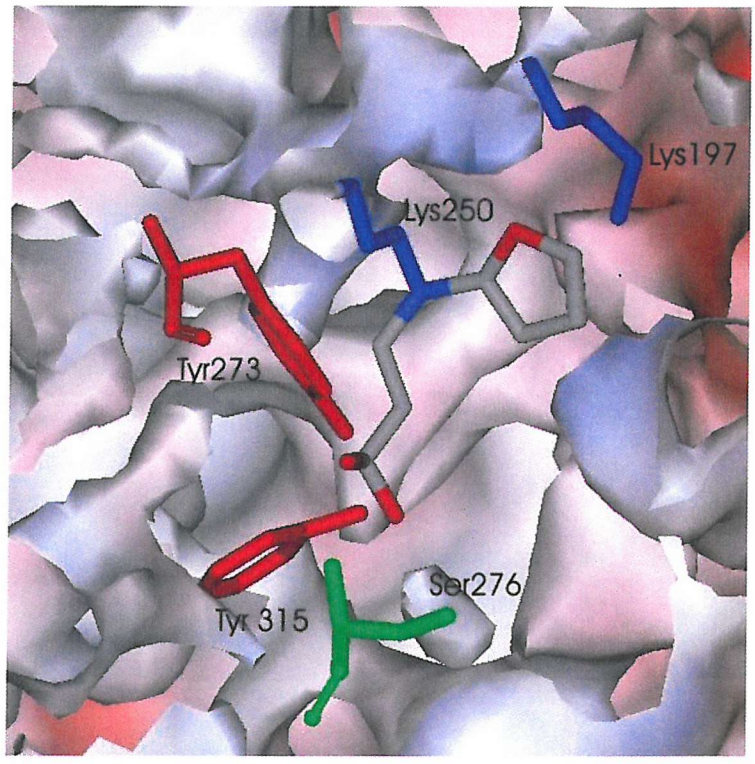


Figure 5.6 The predicted conformation of inhibitor 4 (4-(furan-2-yl)-4-oxobutanoic acid) to Lys 250 of the *P. sativum* ALAD subunits. The carboxylic group of the compound can be seen orientated towards Tyr 273, Tyr 315 and Ser 276 of the predicted structure.

5.9 Discussion

The intention of the inhibitor investigations was to attempt to discover candidates for agrochemical purposes. For compounds to be suitable for this purpose, IC50 values of <10μM and below would have to be observed. Unfortunately, no compounds met this criterion. From these studies the 4-ketoacid motif can be seen to be very important for substrate recognition and therefore competitive inhibition, as has been reported previously (Nandi and Shemin, 1968), due to the requirement of the enzyme for the formation of a Schiff base between the substrate and the amino group of the active site lysines. The *P. sativum* enzyme does not seem to require an amino group for compound recognition and the substitution of this amino group with a variety of pentane and hexane structures did not seem to affect the ability of the compounds to bind the enzyme. In fact these compounds were observed to be among the best inhibitors from those synthesised. Substitution at

various points of the rings with N or O, did not seem to improve the inhibition of the compounds except in the case of compound **3**, where introduction of a N at postion 3 of the benzene ring halved the IC50 in comparison to the primitive form (**5**). The fact that no compounds were synthesised with an amino group present at position 5 of the carbon chain could explain the lack of any compounds with IC50 values in the nanomolar range. No differences were observed when an additional keto group was introduced at C5.

Compounds based on a cyclopentane-1,3-dione showed a very low inhibition of the enzyme, with IC50 values ranging from 2.2mM (**15**) up to 11.8mM (**25**), indicating a potential lack of productive binding at the enzyme active site.

Since most of the compounds synthesised were based on the modification of the amine end of the substrate ALA and, in general, synthesis of longer chain compounds have been indicated as being more successful inhibitors of ALAD, further work should be undertaken to synthesise inhibitors of greater lengths so as to mimic intermediates of the catalytic process. In addition, compounds could be synthesised so as to investigate the dependence of the enzyme on the carboxylic group of the substrate for recognition.

Chapter Six: Investigations into the molecule found bound at the active site of human erythrocyte ALAD

6.1 Introduction

Until recently, the structures that have been gained for ALAD from yeast (Erskine *et al.*, 1999), *P. aeruginosa* (Frankenberg *et al.*, 1999) and *E. coli* (Erskine *et al.*, 1999) have all been gained through co-crystallisation with either substrate or inhibitor molecules, namely ALA, laevulinic acid, or 4,7-dioxosebacic acid. These structures have shown that the enzyme is a homo octamer comprised of a tetramer of dimers, with each monomer of the dimer containing an active site independent of the neighbouring subunits. It has been proposed that the subunit interactions between monomers in the dimer are essential for the catalytic activity of the enzyme, but the actual concrete evidence of the mechanism has been elusive.

Recently a structure for erythrocyte human ALAD determined in our laboratories by Mills-Davies (2000) showed electron density at the active site for a bound molecule, which at the time was assumed to be the enzyme catalysed product PBG (Figure 6.1). This assumption was based upon previous work by (Jaffe and Markham, 1988; Jaffe *et al.* 1990) who had conducted experimental work on the *E. coli* and bovine forms of ALAD, that indicated the presence of a PBG-like molecule from NMR studies.

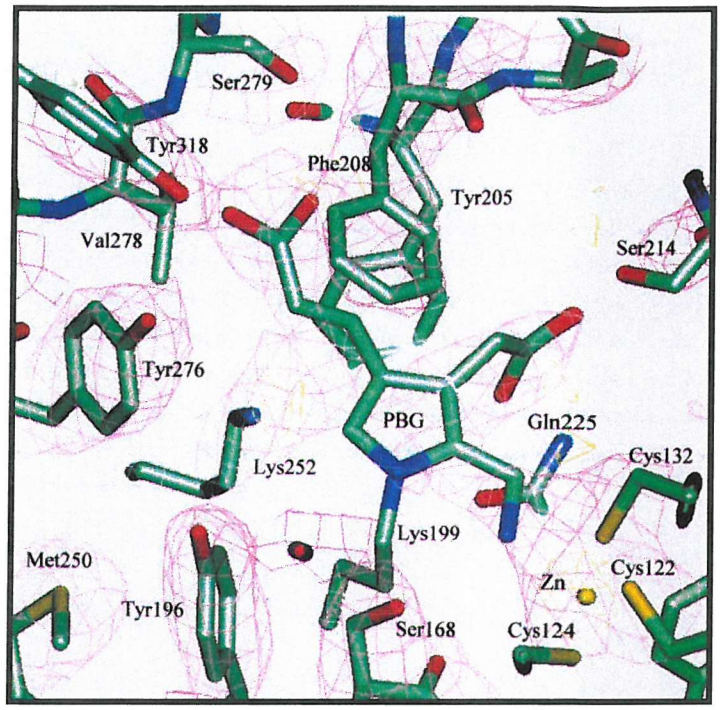


Figure 6.1 The active site of human ALAD from the crystal stucture of human erythrocyte ALAD . The molecule PBG was built into the available electron density observed at the active site. Proposed interactions are shown between Lys-199 and the PBG molecule (Mills-Davies, 2000).

The most recent breakthrough has been the publication of a high resolution native yeast structure (Erskine, 2003). Again, in this structure there is an area of electron density at one active site of each dimer showing the binding of an unknown molecule. This electron density has allowed the authors to build a proposed intermediate into the available electron density based upon the mechanism first proposed by Shoolingin-Jordan where the C-C bond between the substrate molecules is formed first in the catalytic mechanism (Figure 6.2). From these results the authors proposed a reaction scheme incorporating the proposed intermediate (Figure 6.3).

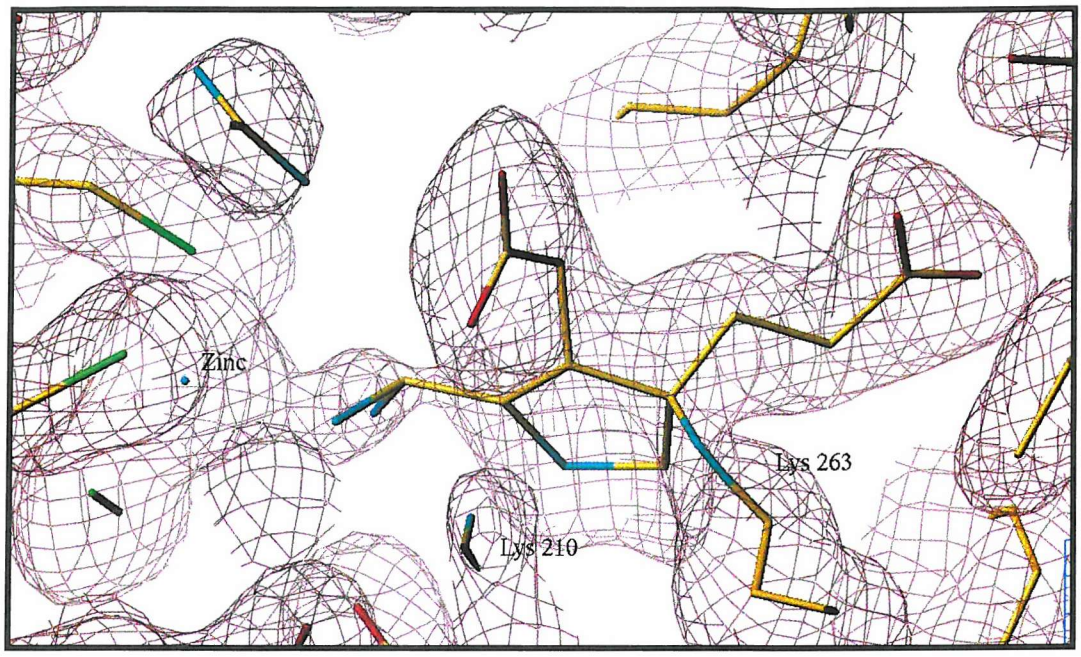


Figure 6.2 The electron density map of the putative intermediate covalently bound to the Lys263 of yeast ALAD at 1.6Å resolution.

The P-site is on the right-hand side and the A-site is formed by residues on the left-hand side, including the zinc ion. The tip of the other invariant lysine (Lys-210) can be seen in the foreground to the left of Lys-263 but has been partially omitted for clarity. Image provided by J. B. Cooper, Southampton University (Erskine *et al.*, 2003).

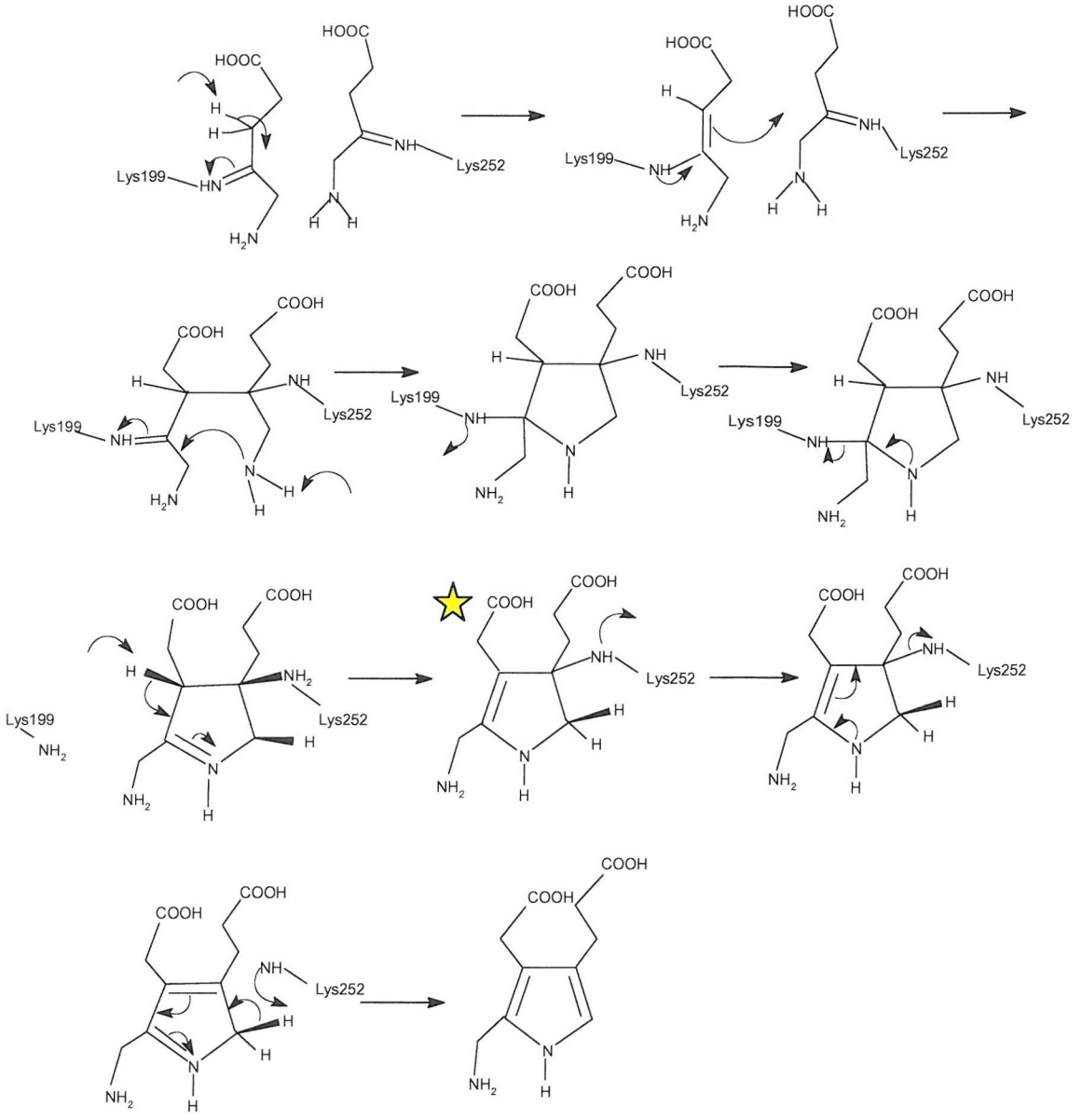
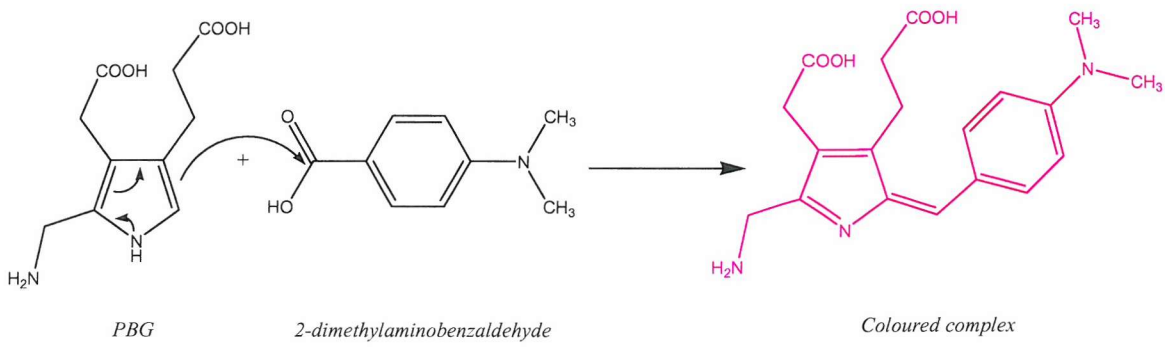


Figure 6.3 A possible mechanism for ALAD (Erskine *et al.*, 2003).

Two Schiff bases are proposed to form utilising Lys252 and Lys199. The Schiff base formed with Lys199 may allow proton abstraction at the C-3 position of A-side ALA, which could lead to nucleophilic attack on the P-side ALA forming the C-C bond linking the two substrates. The C-N bond would then form followed by the loss of the second deprotonation of the A-site substrate to yield the putative intermediate found in the crystallised enzymes. It is thought that the final steps of catalysis may require conformational changes of the enzyme. The putative intermediate is labelled with a star (☆).

The structure of human erythrocyte ALAD shows binding of the molecule at only one active site of each dimer, which has given rise to the proposal that ALAD, or at least the Zn^{2+} dependant form of the enzyme, possesses half-site reactivity.

In addition to structural studies, work by Butler (2003) has deduced that the molecule bound at the active site of the human ALAD was not a pyrrole, as it did not react with Erlich`s reagent. Therefore it could not be PBG, as proposed in the human erythrocyte ALAD structure, as it did not have the free α-position available for the reaction with the aldehyde of the Erlich`s reagent(Figure 6.4).



PBG *2-dimethylaminobenzaldehyde* *Coloured complex*

Figure 6.4 The reaction of PBG with 2-dimethylaminobenzaldehyde (core component of Erlich`s reagent) to form a coloured complex.

Native mass spectrophotometric analysis of the human recombinant ALAD (Butler, 2003) obtained a single species with a mass of 291579 Da, 1247 Da greater than the sequence predicted mass of the octamer (290332 Da). The difference was assigned to the presence of 4 molecules of PBG (4 x 226 = 904) associated with 8 Zn^{2+} ions (8 x 65 = 520) bound to the octamer, a total of 1424 Da. In light of the proposed molecule bound to four of the active sites of the octamer, it is proposed that it is more likely that the difference observed is due to the presence of 4 molecules of the proposed intermediate (4 x 242.25 = 969) (Figure 6.5) associated with 4 Zn^{2+} ions (4 x 65.1 = 261.6) and an associated NH_4^+ ion molecule derived from the buffers used for the technique (18Da), giving a total expected mass increase of 1248.6 Da, a figure very close to the observed mass difference.

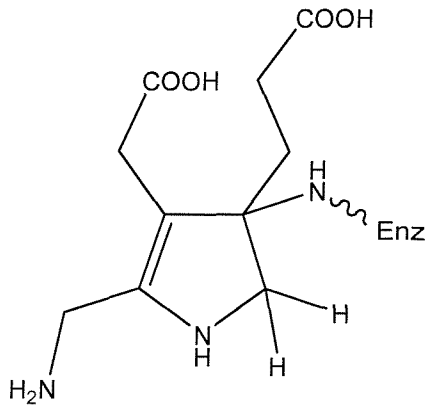


Figure 6.5 The proposed intermediate found bound at the active site of yeast ALAD

6.2 Experiments to determine the origin of the molecule observed at the active site of human recombinant ALAD

Since it had already been shown that the molecule bound at the active site of human recombinant ALAD was not PBG, but that the molecule closely resembles PBG (as first considered by Mill-Davies (2000)), it was assumed that the molecule bound must, in some form, be an intermediate of the enzyme catalysed reaction. In order to test this hypothesis, various experiments were designed and performed using 14C-radiolabelled ALA, so as to characterise the nature of the molecule bound at the active site. Human recombinant ALAD was purified as described in Chapter Two.

6.2.1 Production of the ^{14}C-radiolabelled molecule bound at the active site of human recombinant ALAD

To determine that the observed molecule was indeed derived from the substrate, ALA, experiments utilising ^{14}C-ALA with a specific activity of 51.5mCi/mmol were undertaken. Initially the tracer, ^{14}C-ALA (10nmoles), was mixed with unlabelled ALA (600nmoles), to act as a carrier, and the mixture was incubated with 1mg of human recombinant ALAD (30nmoles of subunit) in 100mM Tris/HCl, pH 7.0, containing 50μM $ZnCl_2$ and 5mM DTT. The reaction was incubated for 10mins at 37°C so as to allow turnover of substrate

by all available subunits. The reaction mixture was then loaded onto a Pharmacia PD10 column and eluted in the same buffer as used for the incubation. Fractions were eluted in 0.5ml samples in Eppendorf tubes. The fractions were then assayed for protein concentration and carbon-14 counts (Figure 6.6). Additionally, it was checked that the enzyme was still active by conducting enzyme assays (described in chapter 2).

It was immediately obvious that most of the carbon-14 label eluted with the later fractions from the PD10 column, indicative of smaller molecules. Since a relatively small amount of tracer was added to the carrier this was not altogether unexpected as through the course of the ten minute incubation period, the majority of the ALA, whether ^{14}C-labelled or unlabelled, would have been converted to PBG, and then separated on the PD10 column.

As expected the protein eluted in the early fractions due to its size, and was easily identifiable though the use of the activity and protein assays, confirming that the enzyme was still viable and that the substrate had not formed a dead-end complex. More importantly, it was also observed that a small amount of the ^{14}C-radiolabel remained associated with the enzyme after the PD10 column, confirming the theory that the observed molecule was indeed an intermediate of the enzyme catalysed reaction.

6.2.2 Stoichiometry of the ALA binding to human recombinant ALAD

10nmoles of ^{14}C-ALA was used as a radioactive tracer in 600nmoles of unlabelled ALA carrier. The 10nmoles of ^{14}C-ALA had a specific activity of 515.5nCi which should have given a total of 1133000dpm.

On elution from the PD10 column the human recombinant ALAD (1mg) had a count rate of on average 11250dpm. Bearing in mind that the labelled ^{14}C-ALA had been diluted 1:60 and that the enzyme was subjected to 10 turnovers (only 1 in 10 molecules of labelled ALA would remain bound, with the remainder being converted to PBG), it can be assumed that if all of the substrate used had been ^{14}C-radiolabelled, then 1mg (30nmoles of subunit) of

human recombinant ALAD would be associated with 6750000dpm or 59.57nmoles of ALA. Based upon this data, it is proposed that each subunit of enzyme binds 2 molecules of ALA.

6.2.3 Turnover of the putative intermediate

Enzyme with the associated labelled molecule from the previous experiment (300μg, 10nmoles) was incubated with excess unlabelled ALA (600nmoles) with the aim of causing turnover of the associated ^{14}C-labelled molecule and in doing so demonstrating that the molecule was subject to turnover and therefore a reaction intermediate. As before, after incubation, the sample was subjected to gel filtration on a PD10 column and 0.5ml fractions collected. The fractions were assayed for protein concentration and carbon-14 counts (Figure 6.6). Additionally, it was checked that the enzyme was still active by conducting enzyme assays (described in chapter 2).

Approximately 3.5% of the ^{14}C-label remained associated with the enzyme while the rest eluted in the later fractions, indicative of smaller molecules. This experiment confirmed that the molecule is displaced by the addition of ALA, and indeed is subject to turnover in the presence of ALA. Therefore the molecule can be assigned as an intermediate rather than a cofactor.

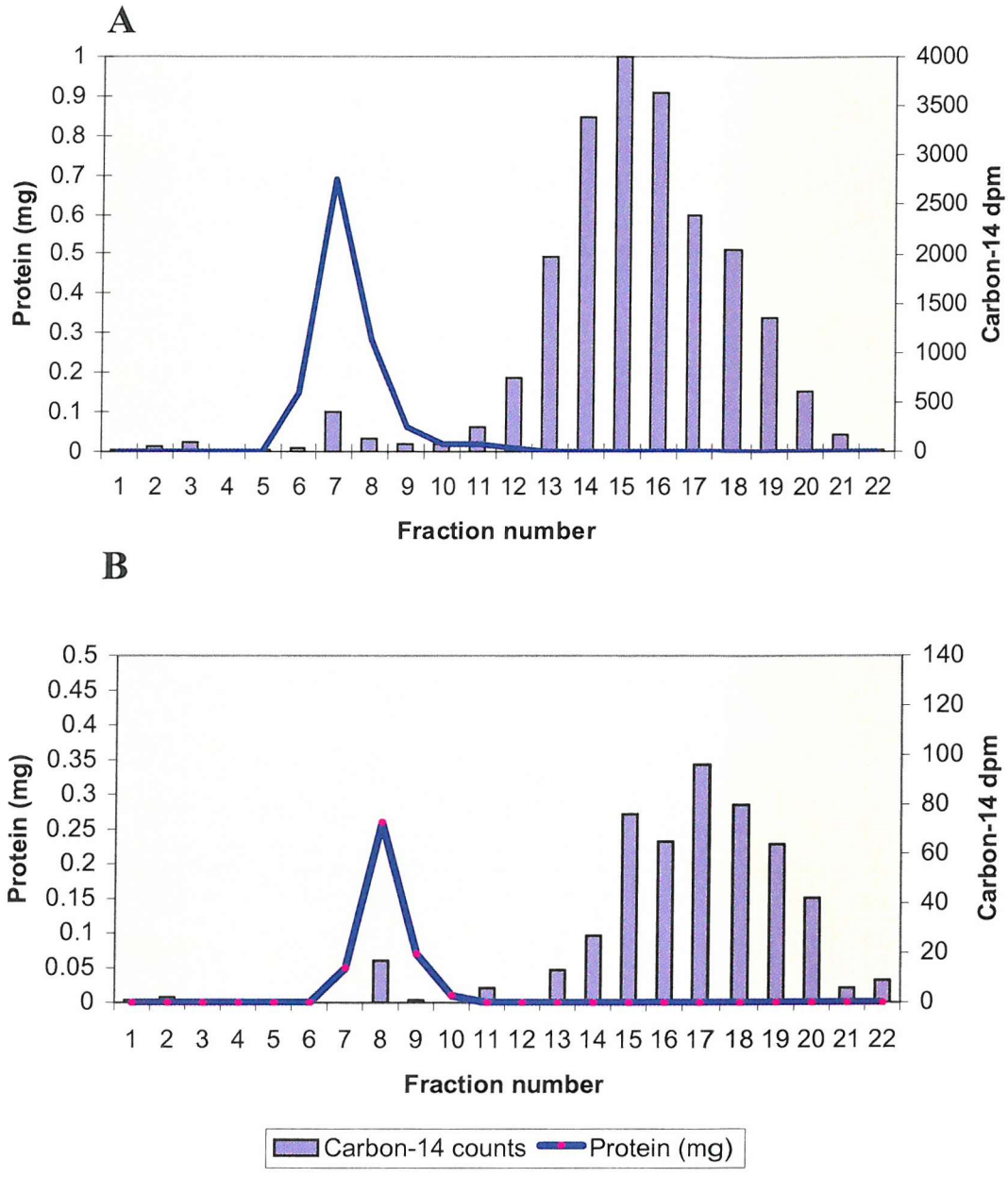


Figure 6.6 Graphs to show association of ^{14}C-ALA with human recombinant ALAD.

Graph A shows the results from incubation of recombinant human ALAD with ^{14}C-ALA. It is clear that, although the majority of the ^{14}C-radiolabel is converted into PBG (fractions 11-21), a percentage of it remains associated with the protein (fractions 5-10).

Graph B shows the results from incubation of the enzyme associated with ^{14}C-radiolabel with unlabelled ALA. The majority of the ^{14}C-radiolabel is found in the latter fractions (fractions 11-22), indicating that the ^{14}C-labelled intermediate has been displaced from the enzyme.

6.2.4 The effect of PBG and laevulinic acid on the turnover of the intermediate

It had already been established that ALA causes turnover and release of the intermediate and so other molecules were considered as candidates to cause turnover. PBG, the product of the catalysed reaction, was used in order to determine whether the product can complete the turnover of the enzyme.

Enzyme with the associated labelled molecule from the previous experiment (300μg, 10nmoles) was incubated with excess unlabelled PBG (300nmoles). As before, after incubation, the sample was subjected to gel filtration on a PD10 column and 0.5ml fractions collected. Again the fractions were assayed for enzyme activity, protein concentration and carbon-14 counts (Figure 6.7). It was observed that PBG did cause turnover of the intermediate bound, as again the ^{14}C-radiolabel was observed to elute in the later fractions from the column, with only 10% of the ^{14}C-radiolabel remaining associated with the enzyme.

In order to conform that the intermediate was being converted to PBG, the ALAD inhibitor laevulinic acid (LA) was used. LA is a substrate analogue of ALA which lack the amine group of the substrate. Enzyme associated with ^{14}C-radiolabel (300μg, 10nmoles) was incubated with excess LA (600nmoles) and separation was conducted as previously on a PD10 gel filtration column. Similarly to the incubations with ALA and PBG, the ^{14}C-labelled fractions occurred in the latter fractions (Figure 6.7), indicating that the ^{14}C-radiolabel had dissociated from the enzyme, with only 8% of the ^{14}C-radiolabel remaining associated with the enzyme. However, due to the small amounts of molecule in the latter fractions, release of PBG could not be tested directly with Erlich`s reagent which is not sensitive enough for such small amounts of PBG.

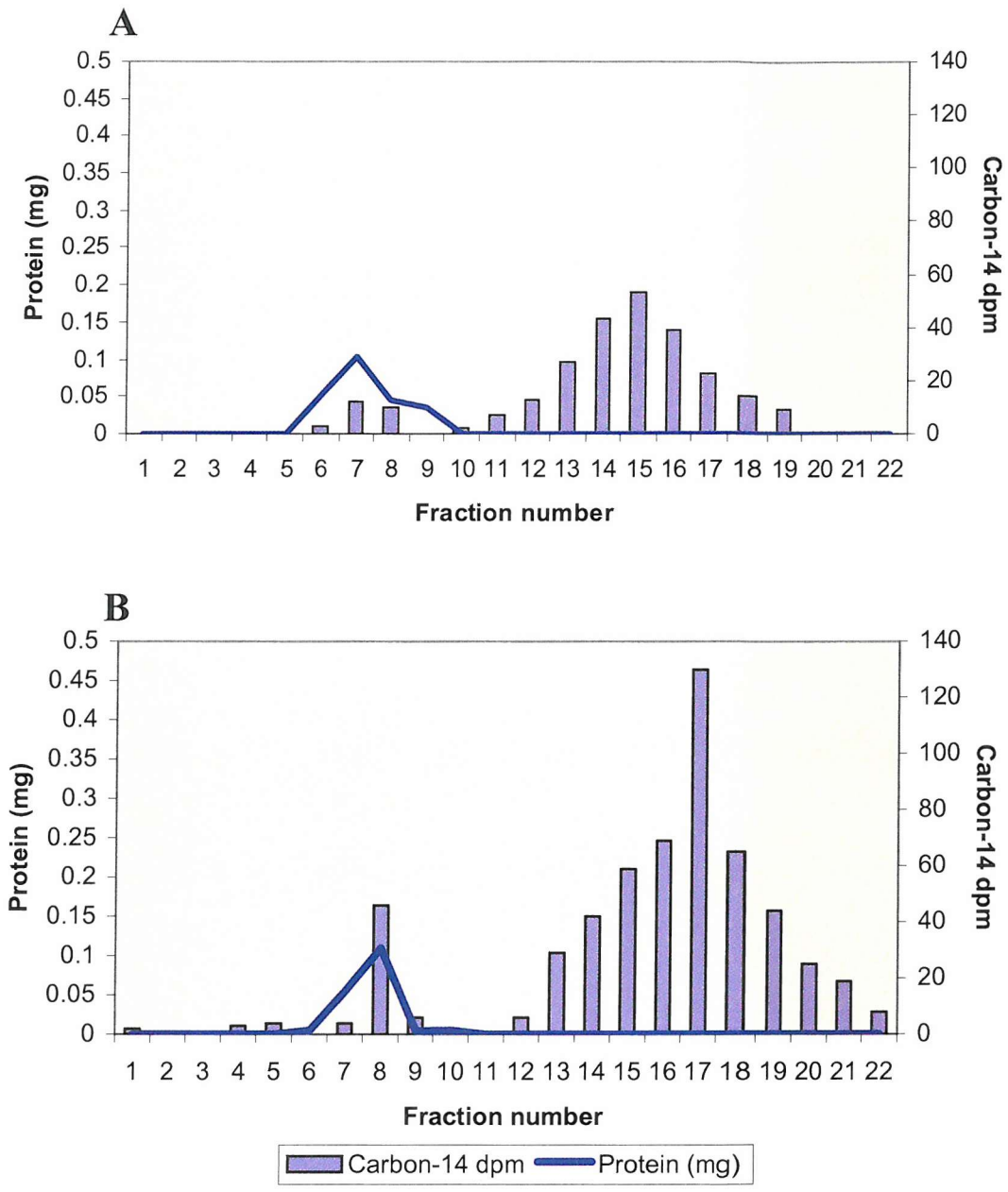


Figure 6.7 Graphs to show the effect of incubation with laevulinic acid and PBG on the association of radiolabelled ALA with human recombinant ALAD.

Graph A shows the results from incubation of the enzyme associated with ^{14}C-radiolabel with PBG. The majority of the ^{14}C-radiolabel is found in the latter fractions (fractions 11-22), indicating that the ^{14}C-labelled intermediate has been displaced from the enzyme.

Graph B shows the results from incubation of the enzyme associated with ^{14}C-radiolabel with LA. The majority of the ^{14}C-radiolabel is found in the latter fractions (fractions 11-22), indicating that the ^{14}C-labelled intermediate has been displaced from the enzyme.

In order to test for PBG within the latter fractions from the experiment conducted using LA, the fractions were pooled and loaded onto a 2ml Dowex-50 column, which had been pre-equilibrated with 10ml of AnalaR water. After loading of the sample, the column was washed with a further 10ml of AnalaR water after which the PBG was eluted using a 0.1M solution of acetic acid. Single drops of the eluent were collect into Eppendorf tubes and each fraction was tested for PBG using capillary tubes: A small volume was allowed to enter the capillary and the same end was then briefly dipped into a small drop of Erlich's reagent until an equal volume had entered the tube. Erlich's positive fractions were subsequently pooled and run on a TLC plate in 4:1:1 mix of acetic acid:isopropanol:water (Figure 6.8).

Figure 6.8 The TLC plate showing the migration of the concentrated sample from the previous experiment against PBG.

The TLC clearly shows that the intermediate has been converted to PBG when LA has been added to the enzyme associated with the ^{14}C-labelled intermediate.

6.3 Discussion

Previous work by Mills-Davies (2000) and Jaffe (1988,1990) has suggested the presence of a ligand that was tightly bound at the active site of ALAD. The exact nature of this ligand

had not been determined, however Jaffe postulated that the pyrrole-like compound observed by NMR studies could be a derivative of PBG, which had been modified by a contaminating protein. Mills-Davies suggested that the molecule found bound at the active site was PBG, due to the nature of the electron density available. Butler conducted experiments to determine the nature of the molecule at the active site of erythrocyte human ALAD, results for reaction with Ehrlich's reagent were negative indicating that the molecule was not PBG.

Native mass spectroscopy indicated a molecular mass close to that expected if the enzyme had four molecules of PBG and 4 Zn ions associated with it. However, in light of recent research by Erskine *et al.* (2003) on the nature of the postulated intermediate, the data can better be attributed to the octameric form of the enzyme associated with four molecules of the proposed intermediate and four Zn^{2+} ions. This data agrees with the observations of half-site reactivity for the human erythrocyte that has previously been reported (Mills-Davies, 2000).

The experiments in this chapter confirm that the molecule that has been observed at the active site of both erythrocyte human ALAD and yeast ALAD is a true intermediate of the reaction mechanism by which ALAD catalyses the reaction of two molecules of ALA to PBG. It has been confirmed that the molecule can be released from the enzyme by addition of ALA, PBG and LA in excesses.

In addition, through the experiments conducted with LA, it has been shown that the turnover of the intermediate to PBG is possible and is likely to be triggered by the presence either ALA or PBG *in vivo*. Since ALA, PBG and LA all cause release of the intermediate, it is clear that the amine group is not essential to cause the dissociation and conversion to PBG. However, it seems likely that the transformation of the intermediate into a released product is a part of the specific enzyme mechanism. Thus the bound intermediate is maintained throughout a fairly rigorous and lengthy purification process and also

throughout the crystallisation process and only is released by an active site specific ligand such as PBG, ALA or LA.

Observations of half-site reactivity are primarily based upon the crystal structure of the Zn^{2+} dependant human erythrocyte ALAD, where the intermediate appears to be found bound at the active site of one subunit within the dimer. In contradiction, the results gained here indicate that two molecules of ALA are found associated with every monomer. This could well be the result of half-site reactivity, whereby the formation of the intermediate at four sites in the octamer is the rate-limiting step of the mechanism. This formation of the intermediate could allow two molecules of ALA to bind at the other four sites, however, the reaction would not be able to continue until the intermediate in the other four sites had been subjected to turnover. Alternatively the binding of substrate to the available active sites could cause the turnover of the intermediate to PBG.

In my opinion it would be of great interest to crystallise the human erythrocyte ALAD with additional ALA so as to potentially obtain crystals with molecules associated at every active site, half with the intermediate bound, and half with ALA bound. Crystallisation of the human erythrocyte ALAD, after first using sodium borohydride in conjunction with an incubation step with ALA so as to trap ALA to the active site, may be a second course of action to obtain this information.

With respect to the turnover experiments, further work would be helpful to quantify the exact concentrations required for turnover of the intermediate and the length of time required for this process to occur. Of particular interest would be experiments to determine whether the bacterial and plant forms of ALAD follow the same process. The Mg^{2+} dependant enzymes would indeed be a very interesting area to study in respect of this phenomenon.

Chapter Seven: Crystallisation of a recombinant human ALAD

7.1 Introduction

The first structure of a mammalian ALAD was solved by Mills-Davies in 2000 (Accession No. 4.2.1.24) (Mills-Davies, 2000). The human erythrocyte ALAD structure revealed a ligand, bound at the active site of one subunit within each dimer, that was reminiscent of the product PBG. When PBG was modelled into the electron density present at the active site, the match was good although not perfect. However, the resolution of the structure was only to 2.8Å and there were certain areas of the structure where modelling was difficult due to absence of any clear electron density.

Since there were areas of the human erythrocyte structure that were unclear. particularly, in monomer b which did not show a ligand molecule bound, attempts were made by Butler (Butler, 2003) to crystallise a recombinant form of the human ALAD in order to collect better diffraction data and hopefully to obtain a more complete structural view of the enzyme.

Based upon the crystallisation conditions employed for the crystallisation of the native human erythrocyte ALAD, Butler(Butler 2003) designed crystal screens using 0.1M Mes buffer at pH 6.2, pH6.4 and pH6.8. The concentration of ammonium sulphate was varied from 1M to 1.6M and the dioxane concentration from 0-10% (see Table 7.1 for secondary screen conditions). Resulting crystals were then screened for the best candidates for freezing and data collection. One large crystal (~1mm in length) was subsequently frozen with the addition of 40% glycerol as cryoprotectant and taken to the ESRF at Grenoble, where diffraction data to 2.8Å was collected.

A	B	C	D	E	F	
0.1M Mes pH 6.2 0% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.2 0% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.4 0% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.4 0% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.8 0% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.8 0% dioxane 1.6M Ammonium sulphate	1
0.1M Mes pH 6.2 1% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.2 1% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.4 1% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.4 1% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.8 1% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.8 1% dioxane 1.6M Ammonium sulphate	2
0.1M Mes pH 6.2 5% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.2 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.4 5% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.4 5% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.8 5% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.8 5% dioxane 1.6M Ammonium sulphate	3
0.1M Mes pH 6.2 10% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.2 10% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.4 10% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.4 10% dioxane 1.6M Ammonium sulphate	0.1M Mes pH 6.8 10% dioxane 1.0M Ammonium sulphate	0.1M Mes pH 6.8 10% dioxane 1.6M Ammonium sulphate	4

Table 7.1 Secondary screen conditions used for the crystallisation of recombinant human ALAD.

7.2 Data processing

The resulting data images were processed using MOSFLM, following the guidelines discussed in Chapter Two.

The putative space group was predicted through the use of the integrated autoindex program of MOSFLM. The program predicted a space group of P422 with cell dimensions of 127.08 x 127.08 x 91.23. This higher symmetry space group allows the collection of fewer images, typically around 45° in this case, in order to obtain a full dataset and therefore the most amount of unique data. An added advantage to this space group is that all the required data can normally be collected from a single, crystal as was the case in these experiments.

MOSFLM was then used to refine the detector parameters, using least square refinement and the profile fitting parameters were determined and post-refinement of the crystal orientation, beam parameters and cell parameters were performed.

Once data had been processed, SCALA was utilised to examine the quality of the data, such as distribution of the reflections in relation to the resolution, completeness of the data, mosaicity and intensities. Finally, the intensities were converted to structure-factor amplitudes. The statistics for the recombinant human ALAD can be viewed in Table 7.2.

Statistics for entire dataset	
Resolution range (Å)	29.35-2.80
R_{merge} (%)	13.4
Completeness (%)	99.7
Multiplicity	9.2
Mean I/σ (I)	3.4
Statistics for outer shell	
Resolution range (Å)	2.95-2.80
R_{merge} (%)	10.7
Completeness (%)	99.7
Multiplicity	9.1
Mean I/σ (I)	1.7

Table 7.2 The statistics for the 2.8Å dataset collected and processed from a recombinant human ALAD crystal. The data indicated the presence of two subunits per asymmetric unit.

7.3 Molecular replacement of recombinant human ALAD

Molecular replacement was carried out using the program MOLREP. Since the structure of the human erythrocyte ALAD had already been solved, this was obviously used as the search model, thereby reducing associated problems with molecular replacement. A radius of integration of 30Å gave the best results (Tables 7.3 and 7.4) and a model with the new coordinates was outputted by MOLREP. The model was opened in MOLPACK and the symmetry related subunits were checked for the correct orientation and packing in comparison to the previously solved native human erythrocyte ALAD structure.

Peak number	α	β	γ	Rf/σ
1	28.45	0.00	104.64	6.97
2	88.86	0.00	0.00	6.05
3	60.67	0.00	0.00	3.23
4	85.33	87.50	112.14	2.22
5	59.70	11.96	204.94	1.99
6	33.06	13.18	210.00	1.99
7	5.73	45.79	94.75	1.98
8	78.29	51.05	162.95	1.97
9	40.23	88.15	112.40	1.97
10	26.71	90.00	250.15	1.93

Table 7.3 A table listing the top ten rotation peaks for the molecular replacement of human recombinant ALAD.

Two potential rotation peak search hits (highlighted in yellow) can be observed as having Rf/ σ above the standard values.

Peak number	X	Y	Z	Density/σ
1	0.985	0.496	0.952	82.42
2	0.485	0.996	0.953	18.96
3	0.485	0.996	0.452	17.74
4	0.984	0.496	0.009	13.60
5	0.984	0.496	0.270	12.01
6	0.987	0.496	0.062	12.01
7	0.986	0.496	0.163	11.77
8	0.985	0.996	0.954	11.45

Table 7.4 A table listing the calculated translation peaks for the molecular replacement of human recombinant ALAD

The proposed translation peak search hit (highlighted in yellow) can be seen to have a very high density/σ value in comparison to the other proposed peaks.

7.4 Refinement of the molecular replacement solution

The molecular replacement solution was displayed at all times using TURBO FRODO and refinement was undertaken using a combination of SHELX and RESTRAIN.
Refinement using SHELX involved rigid-body refinement followed by torsion-angle simulated annealing and individual B-factor refinement.

After each refinement step, as the phases improved, rebuilding of the output model was undertaken in TURBO-FRODO and the residues were rebuilt into observed electron density, ensuring that angles and bond lengths were of acceptable parameters. Such model building was undertaken using the F_o-F_c and the $2F_o$-F_c maps.

Since the search model used had certain residues removed and stretches of the sequence had been replaced with polyalanine sequences, due to lack of distinctive electron density, the first task was to rebuild these residues using the available electron density.

After two rounds of refinement, the phases improved sufficiently to provide enough electron density so as to build the catalytic Zn^{2+} into the density at each active site. Additionally, as was observed with the native erythrocyte human ALAD structure, there was electron density indicating a bound ligand at the active site of one subunit within each dimer. Unfortunately, the electron density for the ligand molecule did not improve through the refinement process enough to model in an exact structure of the molecule. However, as before, it resembled a porphobilinogen like structure.

7.5 The structure of recombinant human ALAD

The structure of the recombinant form of human erythrocyte ALAD is, as expected, very similar to that of native human erythrocyte ALAD. Improvements in the recombinant human erythrocyte ALAD structure have been made with the inclusions of four residues in the monomer containing the bound ligand molecule, and twenty-eight residues in the second monomer. It had not been possible to build these into the human erythrocyte ALAD

structure. Additionally, it has been possible to build the catalytic zinc ion into the active sites in both of the monomers within the dimer, not only the active site containing the bound ligand molecule. The gross appearance of the native and the recombinant human ALAD forms are, however, very similar.

7.5.1 The recombinant human ALAD monomer

The monomer of the recombinant human ALAD exhibits a structure very similar to that of other ALADs whose structures have been solved. The monomer adopts the classical TIM barrel fold formed by eight β-sheets interspersed by eight segments of α-helix (Figure 7.1a). The overall structure is formed from the following secondary elements numbered from the *N*- to the *C*-terminus: β1, α2, β4, α3, β5, α4, β6, α5, β7, α6, β8, α7, β9, α9, β10 and α10. The residues comprising these elements are listed in Table 7.5. The active site of the monomer is formed from the loop regions that connect the *C*-terminal end of the β-sheets and α-helices. The other main component of the monomer is the *N*-terminal arm region which extends away from the barrel region and comprises the first twenty-eight amino acids of the sequence.

Additionally, there is a further structural element that is present in the four monomers with the ligand molecule bound at the active site. This is the loop region that has been previously documented for the native erythrocyte human ALAD and is observed to act as a "lid" to the active site. This comprises amino acid residues 197 to 220.

Symbol	Human A residues	Human B residues	Secondary Structure
	1-7	1-7	Coil
	8-10	8-10	310-helix
	11-13	11-13	turn
α1	14-20	14-20	α-helix
	21-30	21-30	Coil
β1	31-38	31-38	β-strand
	39-43	3-43	turn
β2	44-46	44-46	β-strand
	47-53	47-53	β-hairpin turn
β3	54-57	54-57	β-strand
	58-61	58-61	coil
α2	62-70	62-70	α-helix
	71-74	71-74	turn
β4	75-80	75-80	β-strand
	81-99	81-99	coil
α3	100-111	100-111	α-helix
	112-115	112-115	turn
β5	116-121	116-121	β-strand
	122-142	122-142	coil
α4	143-159	143-159	α-helix
	160-163	160-163	turn
β6	164-167	164-167	β-strand
	168-173	168-173	turn
α5	174-184	174-184	α-helix
	185-191	185-191	turn
β7	192-199	192-199	β-strand
	200-205	200-205	coil
Active site loop	206-212	206-212	α-helix
	213-223	213-223	coil
	224-230	224-230	coil
α6	231-243	231-243	α-helix
	244-248	244-248	turn
β8	249-253	249-253	β-strand
	254-257	254-257	310-helix
α7	258-267	258-267	α-helix
	268-272	268-272	turn
β9	273-278	273-278	β-strand
α8	279-290	279-290	α-helix
	291-295	291-295	turn
α9	296-310	296-310	α-helix
	311-313	311-313	turn
β10	314-317	314-317	β-strand
	318-319	318-319	turn
α10	320-325	320-325	α-helix
	326-328	326-328	turn

Table 7.5 The secondary structure elements of recombinant human erythrocyte ALAD.

7.5.2 The recombinant human ALAD dimer

The strongest interactions within the octamer are those between pairs of monomers that result in the formation of dimers. These interactions are principally between the *N*-terminal arm regions of one monomer (residues 1-20) and the α6 helix (residues 231-243) of the neighbouring monomer within the dimer (Figure 7.1b). The way in which the interactions between monomers occur gives the dimer the overall shape of a twisted number "69".

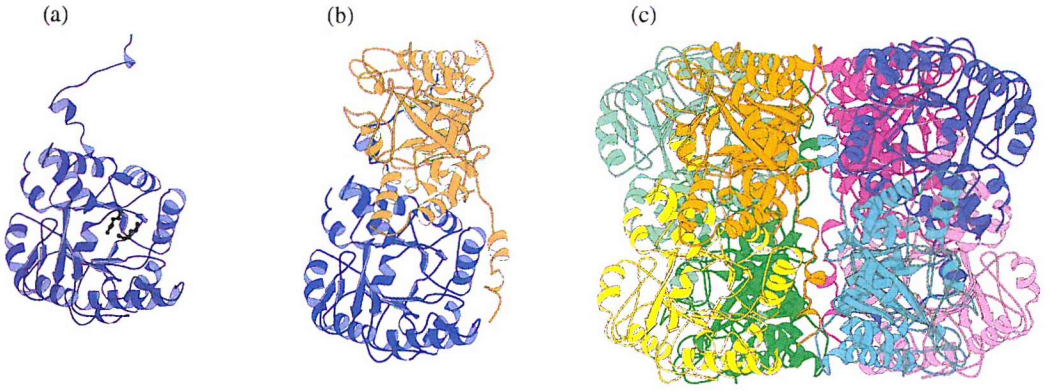


Figure 7.1 The structure of recombinant human ALAD.

(a) The structure of the monomer is based around the TIM barrel fold with an extended *N*-terminal arm region. The two active site lysines are shown in black.

(b) The structure of the dimer gives the appearance of the number "69" with the main interactions between the monomers mediated by their *N*-terminal arm regions.

(c) The octamer is best described as a tetramer of dimers and has a solvent channel of 15-20Å running through its centre.

7.5.3 The recombinant human ALAD octamer

The best description of ALAD is as a tetramer of dimers. The axis of each dimer is orientated towards the centre of the 4-fold axis of the octamer, through which runs a large solvent filled channel 15-20Å diameter. The majority of the interactions between the dimers are again mediated by the outside of the *N*-terminal arm regions as is shown in Figure 7.1c.

7.5.4 The recombinant human ALAD active site

As with other ALAD structures, the active site cleft of the recombinant human ALAD is comprised of amino acid residues located at the *C*-terminal ends of adjacent β-strands and is dominated by two invariant lysine residues, Lys199 and Lys252. The lysine side chains emerge from a hydrophobic pocket predominantly filled with tyrosine side chains. However, the active site also contains a number of invariant polar residues which have been indicated to have possible roles in substrate binding and catalysis. Based on similar structures, Lys252 would form the Schiff base to the P-side substrate and, indeed, has a more hydrophobic environment than its neighbouring Lys199. Ser279 and Tyr318 are implicated in substrate binding and orientation within the active site, based upon their proximity to the active site lysines. These two residues are again conserved between all ALAD structures that have been solved as yet. Additionally, Phe208 may also define packing of the substrate within the active site (Figure 7.2) since the phenyl ring can interact with the methylenes of both substrate molecules.

The catalytic zinc is coordinated between cysteine residues 122,124 and 132. Interestingly, the zinc was found to be present in both subunits of the dimer, not just in the monomer with the ligand density at the active site as noted previously (Mills-Davies, 2000). There was no evidence of a second "structural" zinc in the structure.

There was evidence for a ligand molecule bound at the active site of one monomer within each dimer. The ligand molecule, proposed as a catalytic intermediate, could be modelled into the active site cleft (Figure 7.3) but the results were not absolutely conclusive, since sufficient electron density was not available to confirm the exact molecule bound.

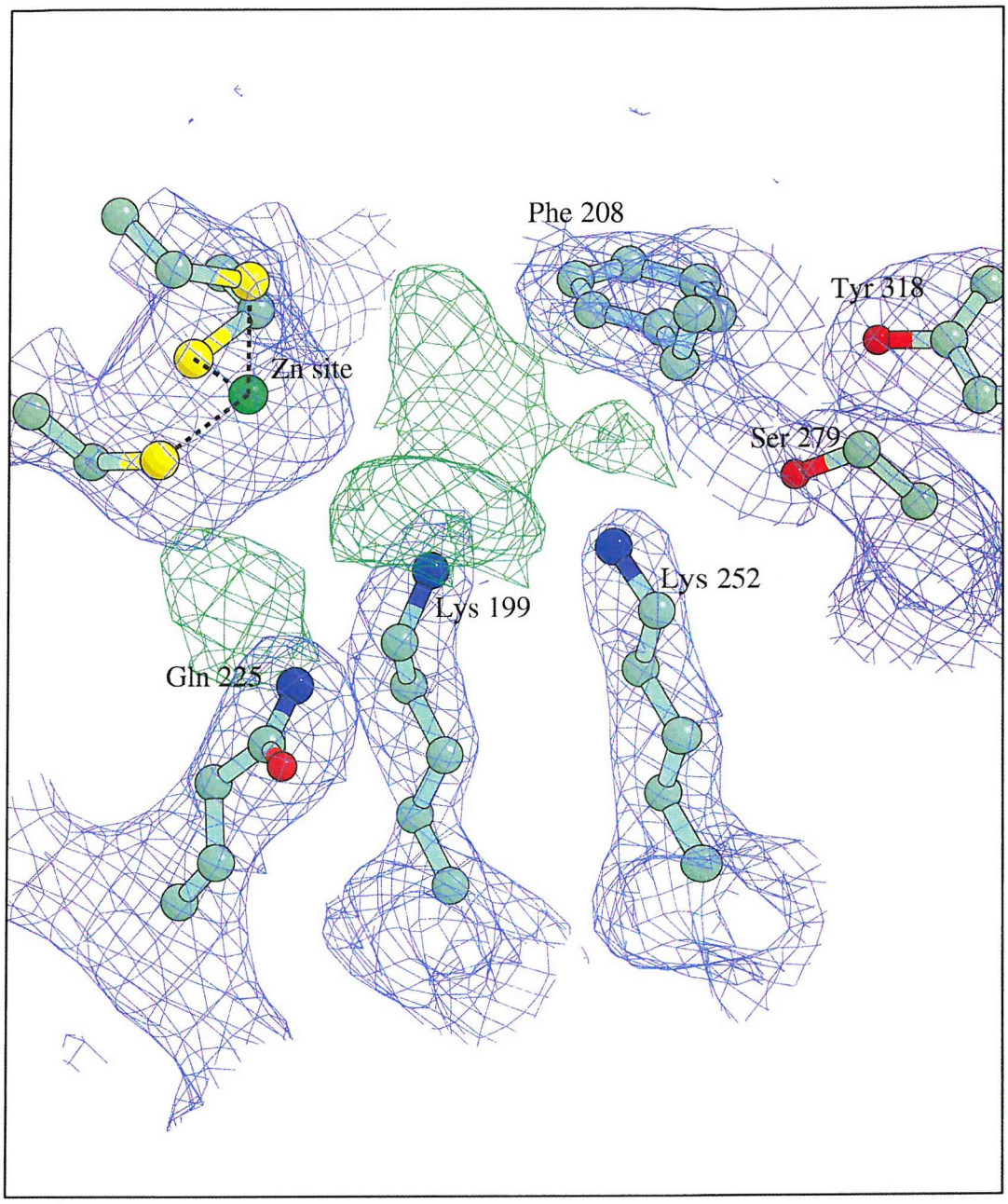


Figure 7.2 A view of the active site of recombinant human ALAD

The active site lysines can be seen, with the P-site lysine (Lys252) apparently interacting with the bound ligand molecule (residues are shown modelled into the F_o-F_c electron density maps (shown in blue)). Unfortunately, as can be seen from the figure, the $2F_o$-F_c electron difference density map (shown in green) is not conclusive enough to determine the type of molecule bound. The catalytic zinc can be observed to be tetrahedrally coordinated between the three invariant cysteine residues. Phe208, Ser279 and Tyr318 are all located at the boundaries of the active site and may also have a role in substrate packing.

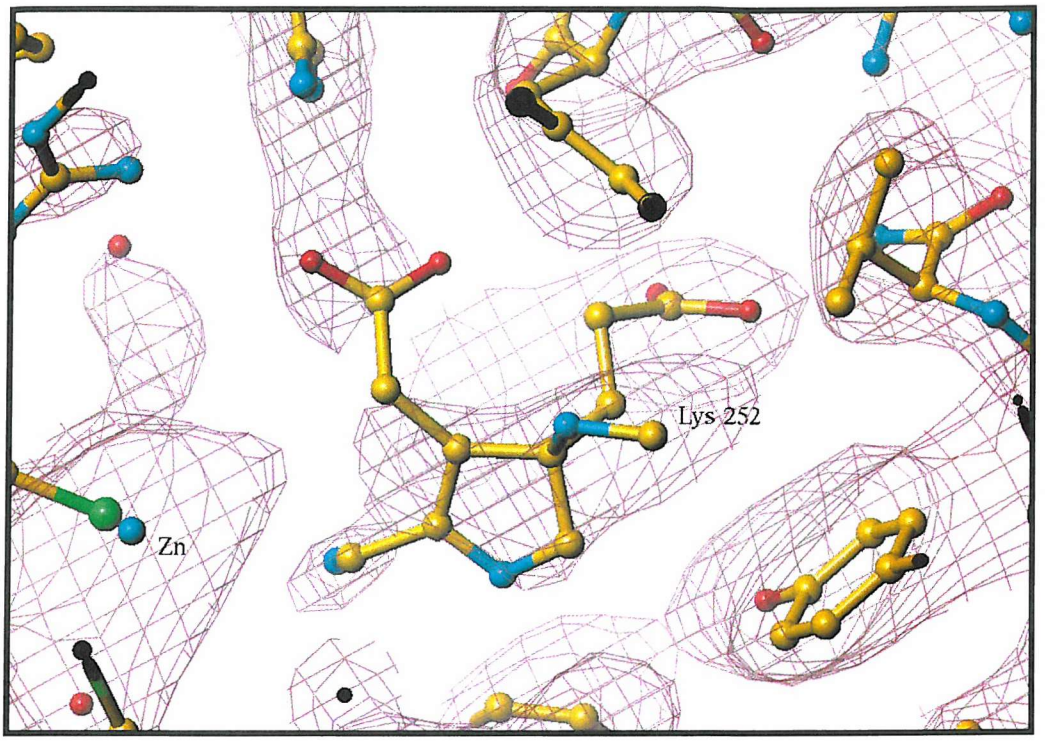


Figure 7.3 The proposed intermediate bound at the active site of human recombinant ALAD.
The proposed intermediate can be seen bound to Lys252 with was no evidence of binding to the other invariant Lys199 which is located in the foreground to the left of Lys252 (omitted for clarity).

7.6 Discussion

The purpose of this study was to improve the original human ALAD model first solved by Mills-Davies (Mills-Davies, 2000) in which there were stretches of the protein that were unresolved due to poor electron density. This has been achieved and those sections of the protein that were poorly resolved have been defined, giving a clear indication of the importance and confirmation of the active sites of both subunits within a dimer.

Unfortunately, the resolution of the structure bound at the active site was not improved and, despite our best efforts, the precise structure of the molecule bound at the active site of the monomer could not be elucidated conclusively. Nevertheless, the proposed intermediate recently proposed by Erskine *et al.* (2003) for the yeast enzyme could be modelled into the

density present at half of the active sites of the human enzyme with the intermediate bound to Lys252 but with no covalent link to Lys199, the other invariant lysine residue at the active site of the human erythrocyte ALAD.

The structure confirms the similarities observed between species of ALAD, with the majority of the contacts of the overall structure mediated between the *N*-terminal arms of neighbouring monomers forming dimers. The overall structure is that of an octamer with a 15-20Å solvent channel running through the centre. There was no evidence for a structural zinc binding site as has been previously been reported (Jaffe, 2000) from kinetic analysis of the enzyme. There was electron density for Zn^{2+} at the catalytic sites in all eight subunits of the octamer, in contrast to the structure determined by Mills-Davies, where a large portion of the structure of the second monomer was disordered. The fact that there is density at the active site of one monomer of each dimer provides new evidence for the possibility of half-site reactivity for the Zn^{2+} dependant forms of the enzyme, as has recently been suggested by the structure produced by Erskine *et al.*(2003).

Future work would have to include further modification of the crystallisation conditions, so as to improve the crystals obtained and therefore the resolution of the data achievable. Better quality data, to a higher resolution, should conclusively prove the nature of the molecule bound and the interaction of the molecules with the surrounding residues within the active site.

Chapter Eight: Final Conclusions

The work described above encompasses extensive investigations on the ALAD enzymes from *P. sativum* and from human and *E. coli* sources.

A new gene construct, encoding a recombinant *P. sativum* ALAD, has been synthesised with suitable codon bias at the 5'-end to facilitate expression in *E. coli*. The DNA sequence lacks the chloroplast targeting sequence (Boese, 1991) that could have led to difficulties in protein folding. The nature of the chloroplast targetting sequence was deduced by comparing the known *N*-terminal sequence of the native pea enzyme with the predicted sequence deduced from the DNA sequence of the *P. sativum* gene and the *N*-termini of *E. coli* and human forms of ALAD. The newly synthesised *P. sativum* clone has been used to produce milligram amounts of purified protein by the design of a novel growth and purification protocol and these measures have reduced the overall time required for the purification process compared with previously published protocols.

The research has revealed the conditions for optimal activity of the *P. sativum* ALAD and factors affecting the catalysis of PBG formation. The studies have also highlighted large differences in pH optima and temperature optima (see figure 8.1) between the ALAD from *E. coli*.

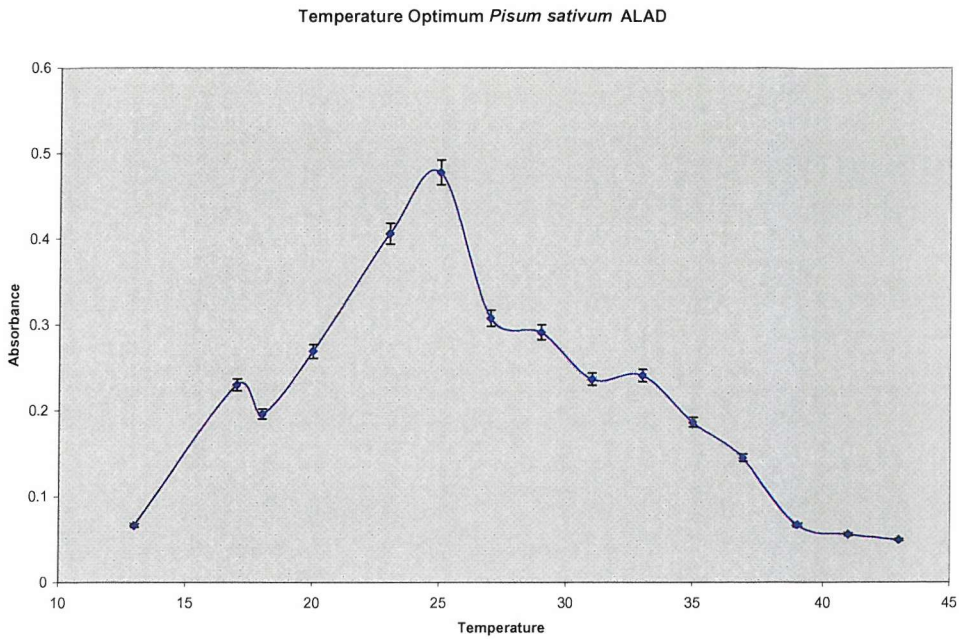


Figure 8.1 The effects of temperature on recombinant pea ALAD over a 30°C temperature range, with a temperature optimum at ~25°C. In comparison, the *E. coli* ALAD shows a temperature optimum of ~48°C.

The properties of the substrate, 5-aminolaevulinic acid, a very reactive compound, have been investigated by NMR with respect to its chemical dimerisation to form a pyrazine under various buffer conditions that have traditionally been used for the assay of ALAD enzymes. Results have shown that the chemical dimerisation at more alkaline buffer pHs and at higher temperatures is significant enough to affect assay results substantially. This has ramifications with regard to the way that ALAD assays need to be conducted in the future.

The stability of the purified *P. sativum* enzyme has been investigated extensively and the dissociation of the octameric enzyme over time has been observed. From these results it was possible to devise conditions that allowed enzyme to be effectively stored over a period of a year, rather than a week, without significant loss of activity. These storage conditions, that involved rapid freezing in liquid nitrogen, could be applied for use with other proteins that are unstable under standard conditions.

The main aim of the study in this thesis was to crystallise the *P. sativum* ALAD and to solve the X-ray structure. Crystals of the enzyme were obtained quite early in the PhD programme (see figure 8.2) but they did not diffract well enough to allow a structure to be determined. A substantial amount of effort went into modifying the crystallisation using a range of different conditions to optimise crystal quality. Crystals that diffracted to below 3Å were obtained that showed high promise. The range of techniques and conditions used for this study would be useful for optimising the crystallisation conditions for other proteins and the research has highlighted areas of interest, such as the ability to grow crystals over large pH ranges and the different crystal forms observed under such conditions.

Figure 8.2 Example of a crystal yielded by secondary crystallisation screening.

Although a data set for *P. sativum* ALAD at 3Å has been obtained, it has not yet been possible to solve the phase problem and a 3-dimensional structure has not yet been obtained. Modification of a recombinant bacterial growth protocol has allowed simple and repeatable preparation of a selenomethionine labelled *P. sativum* ALAD variant, but this did not yield suitable crystals for calculation of phases. The project is therefore well set up for the continued attempts to solve the structure of *P. sativum* ALAD. This is important since a plant ALAD structure is not yet available. The modified conditions used to prepare selenomethionine labelled ALAD have been adapted successfully in the laboratory for the study of other enzymes.

The studies into the inhibition of *P. sativum* ALAD, using novel knowledge-based compounds, has not yielded any significant enzyme inhibitors, however, it has given some insight into which direction further research should proceed with regard to the refinement of inhibitors structure for species-specific inhibition. However, until a 3-dimensional structure is available for higher plant ALAD, it will be difficult to design species-specific inhibitors.

The thesis also incorporates a collaborative study, with Dr Butler, on the structure determination of recombinant human ALAD that has been solved at a resolution of 2.8Å and the structure refined. Previous studies in the laboratory had solved the structure of native human ALAD isolated from erythrocytes. The recombinant structure has given us more information with regard to the active site lid region of the human protein and has confirmed the presence of a tightly bound enzyme intermediate complex (Figure 8.3). Studies into the nature of the compound bound at the active site of the have identified it as a true intermediate of the reaction mechanism of Zn^{2+} dependent enzymes. This strengthens the mechanism proposed by Erskine *et al.* (2003) whereby the C-C bond is formed prior to the formation of the C-N bond in the reaction mechanism (see figure 6.3). It also gives evidence that the turnover of this intermediate is the rate limiting step of the reaction mechanism and that the product can only be released by a subsequent turnover. This has been demonstrated conclusively by turnover experiments using radioactively labelled intermediate and substrate.

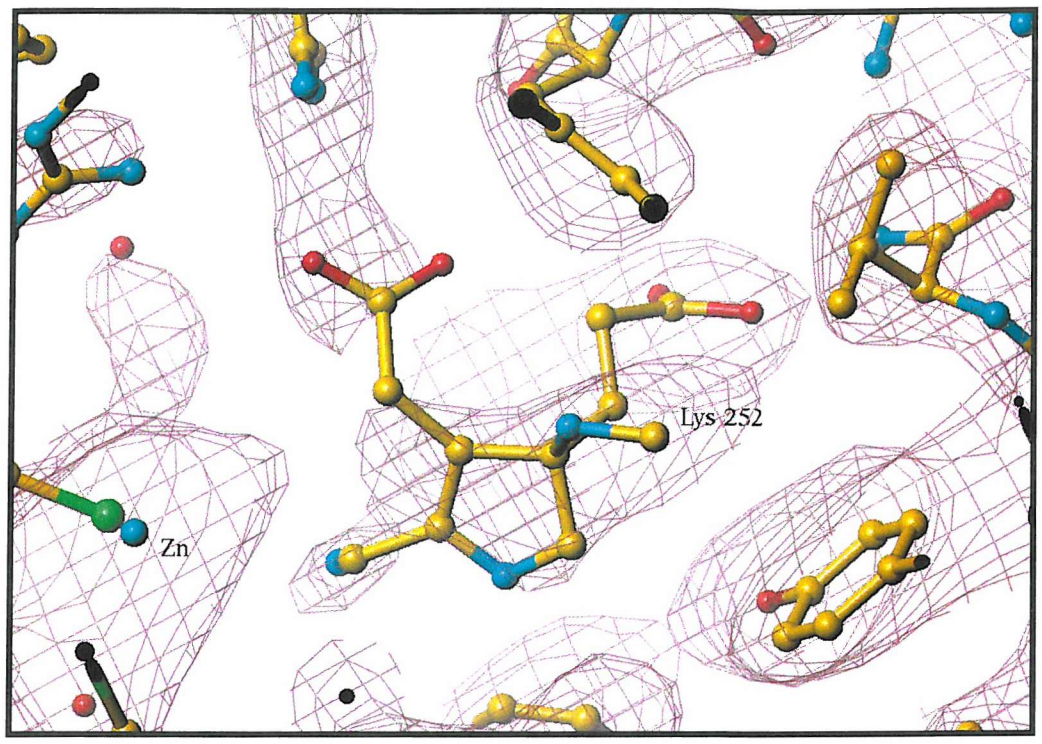


Figure 8.3 The proposed intermediate bound at the active site of human recombinant ALAD.
The proposed intermediate can be seen bound to Lys252 with was no evidence of binding to the other invariant Lys199 which is located in the foreground to the left of Lys252 (omitted for clarity).

In my opinion, there is still much to research in respect to the enzyme ALAD. It is not yet clear whether the reaction mechanism for Zn^{2+} dependant enzymes is exactly the same as that for the Mg^{2+} dependant enzymes, and indeed the role of a metal ion in the reaction mechanism of any ALAD.

The changes in quaternary structure of the enzyme in response to specific factors, such as protein concentration and pH, requires further study in order to investigate the possible change from octameric structure to hexameric. Such changes have been suggested in the case of specific mutations to the human enzyme – for instance it has been suggested that the human Phe 12 Leu mutant exists in a hexameric form rather than as an octamer. In the plant enzyme it is possible that a change from octameric to hexameric structure may play a role in the physiological regulation of the enzyme *in vivo*.

References

Altschul, S., Madden, T., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucl. Acids. Res. **25**(17): 3389-3402.

Appleton, D., Duguid, A. B., Lee, S.-K., Ha, Y.-J., Ha, H.-J. and Leeper, F. J. (1998). "Synthesis of Analogues of 5-Aminolaevulinic Acid and Inhibition of 5-Aminolaevulinic Acid Dehydratase." J. Chem. Soc. **Perkin Trans. 1**: 89-101.

Baldwin, T. (1984). "E. coli TB1-host for pUC plasmids." Focus **6**: 7.

Banerjee, R. and S. W. Ragsdale (2003). "The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes." Annu Rev Biochem **72**: 209-47.

Battersby, A. R. (1978). "The discovery of nature's biosynthetic pathways." Experientia **34**(1): 1-13.

Battersby, A. R. (1986). "Biosynthesis of the pigments of life." Ann N Y Acad Sci **471**: 138-54.

Battersby, A. R., Fookes, C. J., Matcham, G.W. and McDonald, E. (1980). "Biosynthesis of the pigments of life: formation of the macrocycle." Nature **285**(5759): 17-21.

Beale, S. I. and Weinstein, J. D. (1991). "Biochemistry and regulation of photosynthetic pigment formation in plants and bacteria." Biosynthesis of Tetrapyrroles **19**: 155-236.

Bhaguat, L.P. (2001). "Preparation of 3-(2-aminoethylthio)methyl-4-oxo-4-(3-pyridyl)butanoic acid derivatives as neuroprotective agents." Patent WO 2000-2000JP8090, US 99-450245.

Bishop, J. R., Cohen, P.J., Boyer, S.H., Noyes, A.N., Frelin, L.P. (1986). "Isolation of a rat liver delta-aminolaevulinic acid dehdrase (ALAD) cDNA clone: evidence for unequal ALAD gene dosage among inbred mouse strains." Proc Natl Acad Sci U S A **83**: 5568-5572.

Boese, Q.F., Spano, A.J., Li, J.M., Timko, M.P. (1991). "Aminolevulinic acid dehydratase in pea (Pisum sativum L.). Identification of an unusual metal-binding domain in the plant enzyme." J. Biol. Chem. **266**(26): 17060-17066.

Bollivar, D. W. (2003). "Intermediate Steps in Chlorophyll Biosynthesis: Methylation and Cyclization." In: The Porphyrin Handbook, Eds. Kadish, K.M., Smith, K.M., Guilard, R. **13**: 49-68.

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

Breinig, S., Kervinen, J., Stith, L., Wasson, A.S., Fairman, R., Wlodawer, A., Zdanov, A., Jaffe, E.K. (2003). "Control of tetrapyrrole biosynthesis by alternate quaternary forms of porphobilinogen synthase." Nature Structural Biology **10**: 757-763.

Brunger, A. T. (1990). "Extension of molecular replacement - a new search strategy based on Patterson correlation refinement." Acta Crystallographica Section A **A46**: 46-57.

Butler, A.R. and George, S. (1992). "The nonenzymatic cyclic dimerisation of 5-aminolevulinic acid." Tetrahedron **48**: 7879-7886.

Butler, D. (2003). "Studies on recombinant human 5-aminolaevulinic acid dehydratase and recombinant human porphobilinogen deaminase." PhD Thesis, Division of Biochemistry and Molecular Biology, Southampton University.

Chadwick, D. J. and Ackrill, K. (1994). "The Biosynthesis of the Tetrapyrrole Pigments." Wiley **Ciba Foundation Symposium 180**.

Chauhan, S. and O'Brian, M. R. (1993). "Bradyrhizobium japonicum delta-aminolevulinic acid dehydratase is essential for symbiosis with soybean and contains a novel metal-binding domain." J Bacteriol **175**(22): 7222-7.

Chauhan, S., Titus, D.E. and O'Brian, M.R. (1997). "Metals control activity and expression of the heme biosynthesis enzyme delta-aminolevulinic acid dehydratase in Bradyrhizobium japonicum." J Bacteriol **179**(17): 5516-20.

Cheung, K.M., Spencer, P., Timko, M.P., Shoolingin-Jordan, P.M. (1997). "Characterization of a recombinant pea 5-aminolevulinic acid dehydratase and comparative inhibition studies with the Escherichia coli dehydratase." Biochemistry **36**(5): 1148-56.

Collaborative Computational Project, Number 4. (1994). "The CCP4 Suite: Programs for Protein Crystallography." Acta Cryst. **D50**: 760-763.

Cornah, J. E., Terry, M.J., Smith, A.G. (2003). "Green or red: what stops traffic in the tetrapyrrole pathway?" Trends in Plant Science **8**(5): 224-230.

Cox, T. C., Bawden, M.J., Martin, A., May, B.K. (1991). "Human erythroid 5-aminolevulinate synthase: promoter analysis and identification of an iron-response element in the mRNA." EMBO J. **10**: 1891-1902.

Dailey, H. A. (1990). "Biosynthesis of Heme and Chlorophylls." New York: McGraw-Hill.

Derman, A. I., Prinz, W.A., Belin, D., and Beckwith, J. (1993). "Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli." Science **262**: 1744-1747.

Echelard, Y., J. Dymetryszyn, et al. (1988). "Nucleotide sequence of the hemB gene of Escherichia coli K12." Mol Gen Genet **214**(3): 503-8.

Erskine, P.T., Coates, L., Newbold, R., Brindley, A.A., Stauffer, F., Wood, S.P., Warren, M.J., Cooper, J.B., Shoolingin-Jordan, P.M., Neier, R. (2001). "The X-ray structure of yeast 5-aminolaevulinic acid dehydratase complexed with two diacid inhibitors." FEBS Letters **503**(2-3 SU -): 196-200.

Erskine, P. T., Coates, L., Butler, D., Youell, J.H., Brindley, A.A., Wood, S.P., Warren, M.J., Shoolingin-Jordan, P.M., Cooper, J.B. (2003). "X-ray structure of a putative reaction intermediate of 5-aminolaevulinic acid dehydratase." Biochem. J. **373**: 733-738.

Erskine, P.T., Newbold, R., Brindley, A.A., Wood, S.P., Shoolingin-Jordan, P.M., Warren, M.J., Cooper, J.B. (2001). "The x-ray structure of yeast 5-aminolaevulinic acid dehydratase complexed with substrate and three inhibitors." J Mol Biol **312**(1): 133-41.

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

Erskine, P.T., Norton, E., Cooper, J.B., Lambert, R., Coker, A., Lewis, G., Spencer, P., Sarwar, M., Wood, S.P., Warren, M.J., Shoolingin-Jordan, P.M. (1999). "X-ray

structure of 5-aminolevulinic acid dehydratase from Escherichia coli complexed with the inhibitor levulinic acid at 2.0 A resolution." Biochemistry **38**(14): 4266-76.

Erskine, P.T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, I.J., Sarwar, M., Spencer, P., Thomas, P., Warren, M.J., Shoolingin-Jordan, P.M., Wood, S.P., Cooper, J.B. (1997). "X-ray structure of 5-aminolaevulinate dehydratase, a hybrid aldolase." Nat Struct Biol **4**(12): 1025-31.

Evans, J.N., Burton, G., Fagerness, P.E., Mackenzie, N.E., Scott, A.I. (1986). "Biosynthesis of porphyrins and corrins. 2. Isolation, purification, and NMR investigations of the porphobilinogen-deaminase covalent complex." Biochemistry **25**(4): 905-12.

Evans, J.N., Davies, R.C., Boyd, A.S., Ichinose, I., Mackenzie, N.E., Scott, A.I., Baxter, R.L. (1986). "Biosynthesis of porphyrins and corrins. 1. 1H and 13C NMR spectra of (hydroxymethyl)bilane and uroporphyrinogens I and III." Biochemistry **25**(4): 896-904.

Ferreira, G. C. (1995). "Heme biosynthesis: biochemistry, molecular biology, and relationship to disease." J Bioenerg Biomembr **27**(2): 147-50.

Ferreira, G. C. (1999). "5-Aminolevulinate synthase and mammalian heme biosynthesis. In: Ferreira, G.C., Moura, J.J.G., Franco, R. (Eds.), Iron Metabolism. Inorganic Biochemistry and Regulatory Mechanisms." Wiley-VCH, Weinheim: 15-34.

Ferreira, G. C. (1999). "Ferrochelatase." Int J Biochem Cell Biol **31**(10): 995-1000.

Frankenberg, N., Erskine, P.T., Cooper, J.B., Shoolingin-Jordan, P.M., Jahn, D., Heinz, D.W. (1999). "High resolution crystal structure of a Mg2+-dependent porphobilinogen synthase." J Mol Biol **289**(3): 591-602.

Frankenberg, N., Heinz, D.W., Jahn, D. (1999). "Production, purification, and characterization of a Mg2+-responsive porphobilinogen synthase from Pseudomonas aeruginosa." Biochemistry **38**(42): 13968-75.

Fraser, P. J., Curtis, P.J (1987). "Specific pattern of gene expression during induction of mouse erythroleukemia cells." Genes Dev. **1**: 855-861.

French, S., Wilson, K. (1978). "On the treatment of negative intensity observations." Acta. Cryst. **A34**: 517-525.

Frere, F., Schubert, W.D., Stauffer, F., Frankenberg, N., Neier, R., Jahn, D., Heinz, D.W. (2002). "Structure of porphobilinogen synthase from Pseudomonas aeruginosa in complex with 5-fluorolevulinic acid suggests a double Schiff base mechanism." J Mol Biol **320**(2): 237-47.

Frydman, B., Frydman, R.B., Valasinas, A., Levy, E.S., Feinstein, G. (1976). "Biosynthesis of uroporphyrinogens from porphobilinogen: mechanism and the nature of the process." Philos Trans R Soc Lond B Biol Sci **273**(924): 137-60.

Frydman, R. B. and G. Feinstein (1974). "Studies on porphobilinogen deaminase and uroporphyrinogen 3 cosynthase from human erythrocytes." Biochim Biophys Acta **350**(2): 358-73.

Fu, Z.-Q., Du Bois, G.C., Song, S.P., Harrison, R.W. Weber, I.T. (1999). "Improving the diffraction quality of MTCP-1 crystals by post-crystallization soaking." Acta Cryst. D **55**: 5-7.

Gavira, J.A., Toh, D., Lopez-Jaramillo, J., Garcia-Ruiz, J.M., Ng, J.D. (2002). "*Ab initio* crystallographic structure determination of insulin from protein to electron density without crystal handling." Acta Cryst. **D58**: 1147-1154.

Gibbs, P.N., Chaudhry, A.G., Jordan, P.M. (1985). "Purification and properties of 5-aminolaevulinate dehydratase from human erythrocytes." Biochem J **230**(1): 25-34.

Gibbs, P.N. and Jordan, P.M. (1986). "Identification of lysine at the active site of human 5-aminolaevulinate dehydratase." Biochem J **236**(2): 447-51.

Gonzalez-Dominguez, M., M. A. Freire-Picos, Cerdan, M.E. (2001). "Haem regulation of the mitochondrial import of the Kluyveromyces lactis 5-aminolaevulinate synthase: an organelle approach." Yeast **18**(1): 41-8.

Granick, S. (1958). "Porphyrin biosynthesis in erythrocytes. I. Formation of dgr-aminolevulinic acid in erythrocytes." J. Biol. Chem. **232**(2): 1101-1118.

Granick, S. and S. I. Beale (1978). "Hemes, chlorophylls, and related compounds: biosynthesis and metabolic regulation." Adv Enzymol Relat Areas Mol Biol **46**: 33-203.

Granick, S. and D. Mauzerall (1958). "Porphyrin biosynthesis in erythrocytes. II. Enzymes converting dgr-aminolevulinic acid to coproporphyrinogen." J. Biol. Chem. **232**(2): 1119-1140.

Guex, N. a. P., M.C. (1997). "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling." Electrophoresis **18**: 2714-2723.

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

Jaffe, E. K. (2000). "The porphobilinogen synthase family of metalloenzymes." Acta Crystallogr D Biol Crystallogr **56 (Pt 2)**: 115-28.

Jaffe, E.K., Ali, S., Mitchell, L.W., Taylor, K.M., Volin, M., Markham, G.D. (1995). "Characterization of the role of the stimulatory magnesium of Escherichia coli porphobilinogen synthase." Biochemistry **34**(1): 244-51.

Jaffe, E.K., Kervinen, J., Martins, J., Stauffer, F., Neier, R., Wlodawer, A., Zdanov, A. (2002). "Species-specific inhibition of porphobilinogen synthase by 4-oxosebacic acid." J Biol Chem **277**(22): 19792-9.

Jaffe, E.K. and Markham, G.D. (1988). "13C NMR studies of methylene and methine carbons of substrate bound to a 280,000-dalton protein, porphobilinogen synthase." Biochemistry **27**(12): 4475-81.

Jaffe, E.K., Markham, G.D., Rajagopalan, J.S. (1990). "15N and 13C NMR studies of ligands bound to the 280,000-dalton protein porphobilinogen synthase elucidate the structures of enzyme-bound product and a Schiff base intermediate." Biochemistry **29**(36): 8345-50.

Jarret, C., Bobálová, J., Stauffer, F., Henz, M., and Neier, R. (1998). "Inhibition Studies of Porphobilinogen Synthase from Escherichia coli Using Diacids." ECHET 98 **Article 103**.

Jarret, C., Stauffer, F., Henz, M.E., Marty, M., Luond, R.M., Bobálová, J., Schurmann, P., Neier, R. (2000). "Inhibition of Escherichia coli porphobilinogen synthase using analogs of postulated intermediates." Chem Biol 7(3): 185-96.

Jones, T.A. (1978). "A graphics model building and refinement system for macromolecules." J. Appl. Cryst. **11**: 268-272.

Jordan, P.M. (1991). "The biosynthesis of 5-aminolaevulinic acid and its transformation into uroporphyrinigen III." In: Biosynthesis of Tetrapyrroles, Ed. Jordan, P.M. **19**: 1-59.

Jordan, P.M. (1994). "Highlights in haem biosynthesis." Curr Opin Struct Biol **4**(6): 902-11.

Jordan, P.M. and Gibbs, P.N. (1985). "Mechanism of action of 5-aminolaevulinate dehydratase from human erythrocytes." Biochem J **227**(3): 1015-20.

Jordan, P.M. and J.S. Seehra (1980). "Mechanism of action of 5-aminolaevulinic acid dehydratase: Stepwise order of addition of the two molecules of 5-aminolaevulinic acid in the enzymic synthesis of porphobilinogen." J.Chem.Soc.Chem.Comm. 240-242.

Jordan, P.M. and J.S. Seehra (1980). "13C NMR as a probe for the study of enzyme-catalysed reactions : Mechanism of action of 5-aminolevulinic acid dehydratase." FEBS Letters **114**(2 SU -): 283-286.

Kannangara, C.G., Gough, S.P., Bruyant, P., Hoober, J.K., Kahn, A., von Wettstein, D. (1988). "$tRNA^{glu}$ as a cofactor in delta-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis." Trends Biochem Sci **13**: 139-143.

Kashlan, O.B., Scott, C.P., Lear, J.D., Cooperman, B.S. (2002). "A comprehensive model for the allosteric regulation of mammalian ribonucleotide reductase. Functional consequences of ATP- and dATP-induced oligomerization of the large subunit." Biochemistry **41**: 462-474.

Kashlan, O. B., Cooperman, B.S. (2003). "Comprehensive model for allosteric regulation of mammalian ribonucleotide reductase: refinements and consequences." Biochemistry **42**: 1696-1706.

Kervinen, J., Dunbrack, R.L. Jr, Litwin, S., Martins, J., Scarrow, R.C., Volin, M., Yeung, A.T., Yoon, E., Jaffe, E.K.. (2000). "Porphobilinogen synthase from pea: expression from an artificial gene, kinetic characterization, and novel implications for subunit interactions." Biochemistry **39**(30): 9018-29.

Kervinen, J., Jaffe, E.K., Stauffer, F., Neier, R., Wlodawer, A., Zdanov, A. (2001). "Mechanistic basis for suicide inactivation of porphobilinogen synthase by 4,7-dioxosebacic acid, an inhibitor that shows dramatic species selectivity." Biochemistry **40**(28): 8227-36.

Laemmli, U.K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**: 680-685

Laura L. Eggink, R. L., Daniel C. Brune, Judy Brusslan, Akihiro Yamasato, Ayumi Tanaka, and J Kenneth Hoober (2004). "Synthesis of chlorophyll b: Localization of chlorophyllide a oxygenase and discovery of a stable radical in the catalytic subunit." BMC Plant Biol. **4**(1): 5-20.

Leeper, F. J. (1989). "The biosynthesis of porphyrins, chlorophylls, and vitamin B12." Nat Prod Rep **6**(2): 171-203.

Leslie, A. G. (1999). "Integration of macromolecular diffraction data." Acta Cryst. **D55**: 1696-1702.

Li, J.M., Umanoff, H., Proenca, R., Russell, C.S., Cosloy, S.D. (1988). "Cloning of the Escherichia coli K-12 hemB gene." J Bacteriol **170**(2): 1021-5.

Liedgens, W., Grutzmann, R., Schneider, H.A. (1980). "Highly efficient purificaton of the labile plant enzyme 5-aminolevulinate dehydratase (EC 4.2.1.24) by means of monoclonal antibodies." Z Naturforsch [C] **35**(11-12): 958-62.

Mauzerall, D. C. (1998). "Evolution of porphyrins." Clin Dermatol **16**(2): 195-201.

May, B.K., Bhasker, C.R., Bawden, M.J., Cox, T.C. (1990). "Molecular regulation of 5-aminolevulinate synthase. Diseases related to heme biosynthesis." Mol Biol Med 7(5): 405-21.

May, B.K., Borthwick, I.A., Srivastava, G., Pirola, B.A., Elliott, W.H. (1986). "Control of 5-aminolevulinate synthase in animals." Curr Top Cell Regul **28**: 233-62.

McRee, D.E. (1993). "Practical protein crystallography." San Diego, California Academic Press.

Meskauskiene, R., Nater, M., Goslings, D., Kessler, F., op den Camp, R., Apel, K. (2001). " FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*." PNAS **98**(22): 12826-12831.

Mignotte, V., Eleouet, J.F., Riach, N., Romeo, P.-H. (1989). "*Cis*- and *trans*-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene." Natl. Acad. Sci. USA **86**: 6548-6552.

Miller, I.J., Bieker, J.J. (1993). "A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the *Kruppel* family of nuclear proteins." Mol. Cell. Biol. **13**: 2776-2786.

Mills-Davies, N.L. (2000). "Structure of human erythrocyte 5-aminolaevulinic acid dehydratase, the second enzyme in the biosynthesis pathway of haem." PhD Thesis, Division of Biochemistry and Molecular Biology, Southampton University.

Myers, A.M., Crivellone, M.D., Koerner, T.J., Tzagoloff, A. (1987). "Characterization of the yeast HEM2 gene and transcriptional regulation of COX5 and COR1 by heme." J Biol Chem **262**(35): 16822-9.

Nandi, D.L., Baker-Cohen, K.F., Shemin, D. (1968). "Delta-aminolevulinic acid dehydratase of Rhodopseudomonas spheroides." J Biol Chem **243**(6): 1224-30.

Nandi, D.L. and Shemin, D. (1968). "Delta-aminolevulinic acid dehydratase of Rhodopseudomonas spheroides. 3. Mechanism of porphobilinogen synthesis." J Biol Chem **243**(6): 1236-42.

Nandi, D.L. and Shemin, D. (1968). "Delta-aminolevulinic acid dehydratase of Rhodopseudomonas spheroides. II. Association to polymers and dissociation to subunits." J Biol Chem **243**(6): 1231-5.

Nandi, D.L. and Shemin, D. (1968). "dgr-Aminolevulinic Acid Dehydratase of Rhodopseudomonas spheroides. III. MECHANISM OF PORPHOBILINOGEN SYNTHESIS." J. Biol. Chem. **243**(6): 1236-1242.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Cryst. **A50**: 157-163.

Neier, R. (1996). "Chemical Synthesis of Porphobilinogen and Studies of its Biosynthesis." In: Advances in Nitrogen Heterocycles **2**: 35-146.

Norton, E.B. (1999). "Structural characterisation of the active site of Eschericia coli 5-aminolaevulinic acid dehydratase." PhD Thesis, Division of Biochemistry and Molecular Biology, Southampton University.

Novo, M., Huttmann, G. and Diddens, H. (1996). "Chemical instability of 5-aminolevulinic acid used in the flourescence diagnosis of bladder tumours." Journal of Photochemistry and Photobiology **34**: 143-148.

Papenbrock, J., Pfundel, E., Mock, H.P., Grimm, B. (2000). "Decreased and increased expression of the subunit CHL I diminishes Mg chelatase activity and reduces chlorophyll synthesis in transgenic tobacco plants." Plant J **22**(2): 155-164.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., Sweet, R.M. (1994). "Crystal structure of globular domain of histone H5 and its implications for nucleosome binding." Nature **362**: 219 - 223.

Raux, E., Schubert, H.L., Roper, J.M., Wilson, K.S., Warren, M.J. (1999). "Vitamin B12: Insights into Biosynthesis's Mount Improbable." Bioorganic Chemistry **27**(2): 100-118.

Raux, E., Schubert, H.L., Warren, M.J. (2000). "Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum." Cell Mol Life Sci **57**(13-14): 1880-93.

Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Cryst. **D57**: 1373-1382.

Rhodes, G. (2000). "Crystallography made crystal clear." San Diego, California Elsevier Science.

Roessner, C.A., Santander, P.J., Scott, A.I. (2001). "Multiple biosynthetic pathways for vitamin B12: variations on a central theme." Vitam Horm **61**: 267-97.

Rossmann, M.G. and Blow, D.M. (1962). "The detection of subunits within the crystallographic asymmetric unit." Acta Cryst. **15**: 24-31.

Rüdiger, W. (2003). "The Last Steps of Chlorophyll Synthesis." In: The Porphyrin Handbook, Eds. Kadish, K.M., Smith, K.M., Guilard, R. **13**: 71-105.

Sassa, S. (1998). "ALAD porphyria." Semin Liver Dis **18**(1): 95-101.

Schaumburg, A., Schneider-Poetsch, H.A., Eckerskorn, C. (1992). "Characterization of plastid 5-aminolevulinate dehydratase (ALAD; EC 4.2.1.24) from spinach (Spinacia oleracea L.) by sequencing and comparison with non-plant ALAD enzymes." Z Naturforsch [C] **47**(1-2): 77-84.

Schick, B., Jurnack, F. (1994). "Extension of the diffraction resolution of crystals." Acta Cryst. D **50**: 563-568.

Schoenhaut, D.S., Curtis, P.J. (1989). "Structure of a mouse erythroid 5-aminolevulinate synthase gene identification of erythroid-specific DNase I hypersensitive sites." Nucleic Acids Res **17**: 7013-7028.

Scott, A.I. (2001). "Reflections on the discovery of nature's pathways to vitamin B12." Chem Rec **1**(3): 212-27.

Senior, N.M., Brocklehurst, K., Cooper, J.B., Wood, S.P., Erskine, P., Shoolingin-Jordan, P.M., Thomas, P.G., Warren, M.J. (1996). "Comparative studies on the 5-aminolaevulinic acid dehydratases from Pisum sativum, Escherichia coli and Saccharomyces cerevisiae." Biochem J **320 (Pt 2)**: 401-12.

Spencer, P. and Jordan, P.M. (1993). "Purification and characterization of 5-aminolaevulinic acid dehydratase from Escherichia coli and a study of the reactive thiols at the metal-binding domain." Biochem J **290 (Pt 1)**: 279-87.

Spencer, P. and Jordan, P.M. (1994). "5-Aminolaevulinic acid dehydratase: characterization of the alpha and beta metal-binding sites of the Escherichia coli enzyme." Ciba Found Symp **180**: 50-64; discussion 64-9.

Stauffer, F., Zizzari, E., Engeloch-Jarret, C., Faurite, J.P., Bobalova, J., Neier, R. (2001). "Inhibition studies of porphobilinogen synthase from Escherichia coli differentiating between the two recognition sites." Chembiochem **2**(5): 343-54.

Tabor, S., Richardson, C.C. (1985). "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes." Proc Natl Acad Sci U S A. **82**: 1074-1078.

Tishler, P.V. (1999). "The effect of therapeutic drugs and other pharmacologic agents on activity of porphobilinogen deaminase, the enzyme that is deficient in intermittent acute porphyria." Life Sci **65**(2): 207-14.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., Orkin, S.H. (1989). "Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells." Nature **39**: 446-451.

Vagin, A., Teplyakov, A. (1997). "MOLREP: an automated program for molecular replacement." J. Appl. Cryst. **30**: 1022-1025.

Volland, C. and Urban-Grimal, D. (1988). "The presequence of yeast 5-aminolevulinate synthase is not required for targeting to mitochondria." J Biol Chem **263**(17): 8294-9.

Vothknecht, U.C., Kannangara, C.G., von Wettstein, D. (1998). "Barley glutamyl tRNAGlu reductase: Mutations affecting haem inhibition and enzyme activity." Phytochemistry **47**(4 SU -): 513-519.

Warren, M.J. and Scott, A.I. (1990). "Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors." Trends Biochem Sci **15**(12): 486-91.

Wetmur, J.G., Bishop, D.F., Cantelmo, C., Desnick, R.J. (1986). "Human delta-aminolevulinate dehydratase: nucleotide sequence of a full-length cDNA clone." Proc Natl Acad Sci U S A **83**: 7703-7.

Willows, R. D. (2003). "Biosynthesis of chlorophylls from protoporphyrin IX." Nat Prod Rep **20**(3): 327-41.

Willows, R. D., Hansson, M. (2003). "Mechanism, Structure, and Regulation of Magensium Chelatase." In: The Porphyrin Handbook, Eds. Kadish, K.M., Smith, K.M., Guilard, R. **13**: 1-43.

Witty, M., Wallace-Cook, A.D., Albrecht, H., Spano, A.J., Michel, H., Shabanowitz, J., Hunt, D.F., Timko, M.P., Smith, A.G. (1993). "Structure and expression of chloroplast-localized porphobilinogen deaminase from pea (Pisum sativum L.) isolated by redundant polymerase chain reaction." Plant Physiol **103**(1): 139-47.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S, DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z., Graham, M.W. (1989). "Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants." Nucleic Acids Res. **17**: 3469-3478.