Functional analysis of PaxP and PaxQ, two cytochrome P450 monooxygenases required for paxilline biosynthesis in *Penicillium paxilli*

A Thesis presented in partial fulfilment of
the requirements for the degree of
Master of Science in Genetics
at Massey University, Palmerston North,
New Zealand.

Rohan George Thomas Lowe

2002

Abstract

The indole-diterpene paxilline is a potent mammalian tremorgenic mycotoxin and a known inhibitor of maxi-K ion channels. The gene cluster encoding the enzymes for the synthesis of this compound was recently cloned from *Penicillium paxilli* (Young *et al.* 2001). The cluster comprises a set of core genes required for indole-diterpene biosynthesis, including two cytochrome P450 monooxygenases, *paxP* and *paxQ*. Targeted deletion of *paxP* and *paxQ* resulted in mutant strains that accumulate paspaline and 13-desoxypaxilline, respectively, confirming that both genes are involved in paxilline biosynthesis. The aim of the current work is to establish *in vitro* that PaxP and PaxQ catalyse the monooxygenation of paspaline and 13-desoxypaxilline, respectively. To achieve this, cDNA copies of both genes were cloned into pGEX-6P-3, to generate pRL2 and pRL4, and the corresponding glutathione-S-transferase (GST) fusion proteins over-expressed in *E. coli*. However, both GST-fusion proteins accumulated as insoluble inclusion bodies when cultures were incubated at 18°C, 25°C and 37°C. Attempts to express a soluble form of the GST-PaxP by co-expressing this fusion with the chaperones, GroES and GroEL, or by expressing in *E. coli*, Origami B, a strain (*trxB*, *gor*, *lacY*) designed to facilitate expression of active and soluble proteins, were unsuccessful. GST-PaxP was able to be solubilised by the addition of 0.25% *N*-laurylsarcosine, and retained some glutathione binding activity, however, the yield was too low to carry out further experiments. GST and thioredoxin fusion expression constructs were designed in which the putative N-terminal trans-membrane region of PaxP and PaxQ was removed to aid solubility in *E. coli*. These N-terminal modified fusion proteins were still expressed as insoluble protein.

Acknowledgments

Firstly I would like to acknowledge my supervisor Barry Scott. You have provided me with a fantastic lab to work in, and strong support thought my study. I am particularly proud and appreciative that I began my research career in your lab.

I must also thank all the past and present lab members of Scott Base who I had the pleasure of working with. Lisa, you taught me everything about research and the lab environment when I arrived. You always had time for celebration and commiseration and I thank you for your time spent with me. Carolyn, you are the backbone of this lab and I am glad I had your experience as a resource during my study. However, science is not your only skill and I appreciate every beverage consumed and moment enjoyed at 14 Ascot Street. Jonathan, what would I have done without you? Probably been a hell of a lot more bored at the very least. I have had a great time in the lab and a lot of it is due to your friendship. We have shared plenty of good stories, good results, even more bad results and some great practical jokes. I appreciate every one. To Paula and Janet, thanks for good times and great friendship during your time in RB lab. I must also congratulate Andrea; you put up with my crappy jokes and schemes and yet always manage to have a smile on your face. Thank you.

I would like to thank Stan Moore for helping me with the design of N-terminal deletion constructs and multiple sequence alignments. I also thank Geoff Jameson and Gill Norris for advice on recombinant protein expression and Emily Parker for advice on metabolic chemistry.

To my family. Thanks for being supportive and providing me with something I can always look forward to when I manage a visit northward. I am proud of you all and I always work to be a son/brother you can be proud of.

I must also thank all the friends and flatmates who I have spent time with during my study. Special thanks go to James and Phillipa, you guys have always been there for me.

Thanks must also go to the all the people from the Institute of Molecular Biosciences who I have met and worked with along the way. There are a lot of you, but if you get a smile or greeting from me as we pass in the corridor, you know you are on the list!

Table of contents

Abbreviations

APS	ammonium persulphate
bp	base pairs
cDNA	complementary deoxyribonucleic acid
CPR	cytochrome P450 reductase
DEPC	diethylpyrocarbonate
DMAPP	dimethyallyldiphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FPP	farnesyldiphosphate
gDNA	genomic deoxyribonucleic acid
GGPP	geranylgeranyldiphosphate
GST	glutathione-s-transferase
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
IPP	isopentenyldiphosphate
IPTG	isopropylthiogalactoside
kb	kilobase pairs
LB media	Luria Bertaini media
MCS	multiple cloning site
mRNA	messenger ribonucleic acid
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
PAGE	polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
psi	pounds per square inch
RNA	ribonucleic acid

RT-PCR	reverse-transcriptase polymerase chain reaction
SDS	sodium dodecyl sulphate
SRS	substrate recognition site
TEMED	tetramethylethylenediamine
tRNA	transfer ribonucleic acid
UV	ultra violet
X-Gal	5-Bromo-4-chloro-3-indolyl-beta-D-galactoside

List of tables

List of Figures

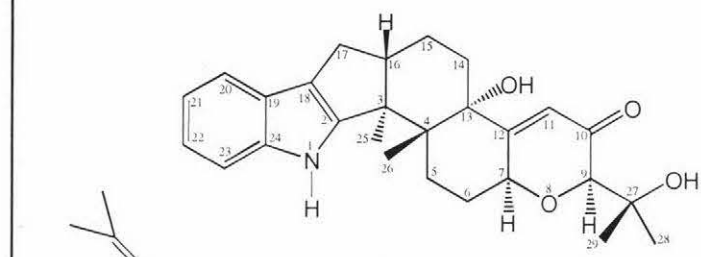
Chapter 1 Introduction

1.1 Indole-diterpenoids and paxilline

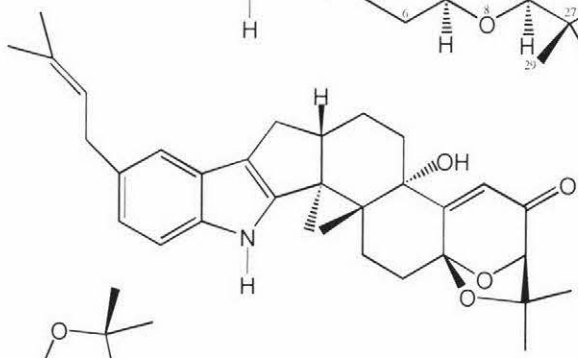
The indole-diterpenoid group of molecules are secondary metabolic products synthesised by fungi. Indole-diterpenes have diverse structural properties and varied pharmacological effects (Steyn and Vleggaar, 1985) including the production of tremors in animals. Indole diterpenes are usually classified into four structural groups, the lolitrems (Gallagher *et al.*, 1984), penitrems (de Jesus *et al.*, 1983), janthitrems (Gallagher *et al.*, 1980) and the paspalitrems (Dorner *et al.*, 1984) (Figure 1.1). This classification could also include the terpendoles (Gatenby *et al.*, 1999; Huang *et al.*, 1995; Tomoda *et al.*, 1995), shearinines (Belofsky *et al.*, 1995) and sulpinines (Laakso *et al.*, 1992). Paxilline was first isolated and identified as a tremorgenic metabolite from *Penicillium paxilli* in 1974 (Cole *et al.*, 1974). Paxilline has a relatively simple indole diterpenoid structure and because of this it is proposed to be an important intermediate in the biosynthesis of other groups of indole-diterpenoids such as lolitrem B (Weedon and Mantle, 1987), an indole-diterpenoid associated with the agricultural livestock disorder 'ryegrass staggers'. This condition is a neurotoxic disorder caused predominantly by the ingestion of lolitrem B produced by the fungal endophyte *Neotyphodium lolii* contained in ryegrass pastures (Gallagher *et al.*, 1984).

Paxilline occurs along with lolitrem B in endophyte infected ryegrasses (Weedon and Mantle, 1987). Their co-synthesis and the fact that every atom in paxilline can be observed at an equivalent position in the lolitrem B molecule suggest that lolitrem B is derived from a biosynthetic precursor of paxilline. Radiolabelling experiments involving the biosynthesis of penitrems and janthitrems also support the hypothesis that paxilline is an intermediate in the biosynthesis of other indole diterpenoid groups (Penn and Mantle, 1994). It is this relation to other important groups of tremorgenic mycotoxins that underlines the importance of understanding the biosynthesis of paxilline. Insights into paxilline biosynthesis will likely be able to be applied to other more complex indole-diterpenoids.

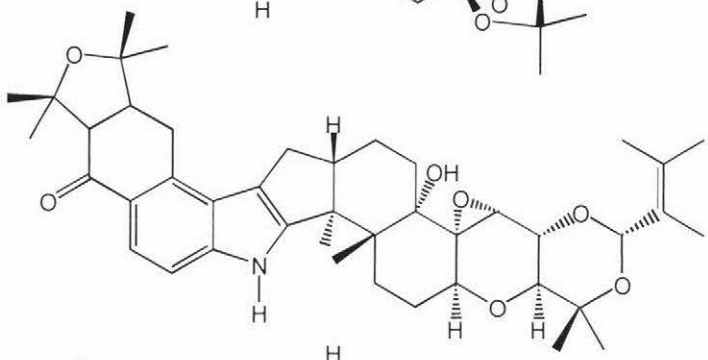
Fig 1.1 **The structures of selected indole-diterpenoids**



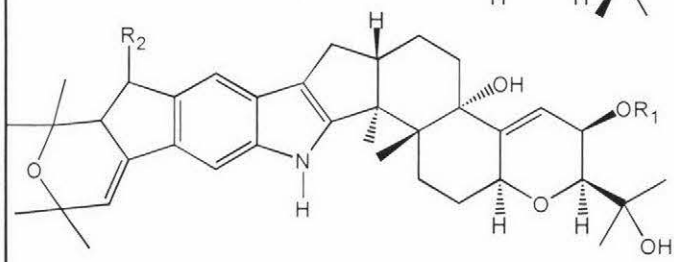
Paxilline



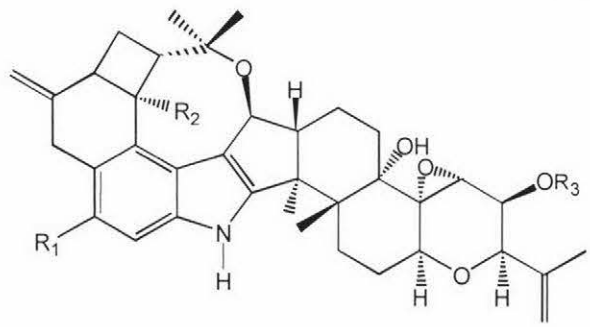
Paspalitrem A



Lolitrem B



Janthitrems



Penitrems

Structural overview of selected indole-diterpenoid molecules.

Chemical structures were derived from those of Steyn and Vleggaar (1985).

1.1.1 Paxilline biosynthesis

The metabolic schemes for the biosynthesis of paxilline are based upon the structural analysis of various indole-diterpenes isolated from fungi and the incorporation of radiolabeled intermediates into paxilline (Mantle and Weedon, 1994; Munday-Finch *et al.*, 1996). These schemes propose that the primary precursors for paxilline biosynthesis are geranylgeranyl diphosphate (GGPP) and an indole group originating from tryptophan. However, *P. paxilli* grown in submerged culture has been shown to incorporate only 5% of radiolabeled tryptophan into paxilline (Laws and Mantle, 1989). In addition, the work of Byrne *et al.* (2000) reveals that in *Nodulosporium sp.* the indole group required for the biosynthesis of nodulosporic acid can be derived from anthranilic acid, while radiolabeled tryptophan failed to be incorporated. Based on these studies Byrne *et al.* (2000) proposed that indole-3-glycerol phosphate (derived from anthranilic acid and ribose) was a shared intermediate for nodulosporic acid and tryptophan synthesis. These results suggest that the indole moiety for paxilline biosynthesis is likely to be indole-3-glycerol phosphate.

The synthesis of GGPP in *Penicillium paxilli* has been shown to be dependant on a GGPP synthase dedicated to secondary metabolism and independent of the GGPP synthase required for primary metabolism (Young *et al.*, 2001). The significance of a dedicated GGPP synthase, is that paxilline production could begin with either the C5 isoprene carbon units of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) or the C15 farnesyl diphosphate (FPP) (Figure 1.2) as both are known substrates for GGPP synthase enzymes (Wiedemann *et al.*, 1993). Therefore, the current hypothesis for paxilline biosynthesis is as follows (Figure 1.2 and 1.3). The synthesis of GGPP begins with the sequential condensation of three acetyl-CoA units to form 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA). This is then converted to mevalonate (a six carbon unit) in an irreversible reaction catalysed by HMG-CoA reductase. IPP and DMAPP, an isomer of IPP, are produced by the decarboxylation and phosphorylation of mevalonate. The condensation of IPP and DMAPP (head to tail) generates geranyl diphosphate (GPP), addition of another IPP to GPP produces farnesyl diphosphate (FPP) and addition of a third IPP molecule creates GGPP. The chemistry of indole addition to GGPP and the associated cyclisation events are still unknown.

However, paspaline and 13-desoxypaxilline are likely intermediates of paxilline biosynthesis on the basis of recent work (Section 1.1.2).

Fig 1.2 **An overview of primary isoprenoid biosynthesis: –Acetyl coenzyme A to dimethylallyl pyrophosphate**

A

Overview of the synthesis of IPP from acetyl coenzyme A

Two molecules of acetyl-Coenzyme A (acetyl-CoA) are combined by Acetoacetyl-CoA thiolase (AACT) to form the 4 carbon molecule acetoacetyl-CoA. Another molecule of acetyl-CoA is added to acetoacetyl-CoA by 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). Coenzyme A (CoA) is removed from HMG-CoA by HMG-CoA reductase (HMGR) to form the six carbon mevalonate. Mevalonate is converted to isopentenyl pyrophosphate (IPP) in a three step process. Mevalonate is phosphorylated by Mevalonate kinase (MK) to form mevalonate 5-phosphate. Mevalonate-5-phosphate is further phosphorylated by Phosphomevalonate kinase (PMK) to form mevalonate-5-pyrophosphate. A single carbon is removed from mevalonate-5-pyrophosphate as CO_2 by Mevalonate-5-pyrophosphate decarboxylase (MDC) producing the five carbon molecule IPP. IPP isomerase (IPPI) converts IPP to dimethylallyl pyrophosphate (DMAPP) and vice versa (Chappell, 1995; Lange *et al.*, 2000; Stryer, 1995).

Enzyme abbreviations:

AACT	Acetoacetyl-Coenzyme A thiolase
HMGS	3-Hydroxy-3-methylglutaryl-Coenzyme A synthase
HMGR	3-Hydroxy-3-methylglutaryl-Coenzyme A reductase
MK	Mevalonate kinase
PMK	5-Phosphomevalonate kinase
MDC	Mevalonate-5-pyrophosphate decarboxylase
IPPI	Isopentenyl pyrophosphate isomerase

Colour Key:
The alpha carbon of Acetyl-CoA is coloured red to help visualise the incorporation of the two carbon skeleton into subsequent catalytic products.
The decarboxylation of mevalonate by MDC is highlighted in blue.
The carbon-carbon double bond acted upon by IPPI is highlighted in green.

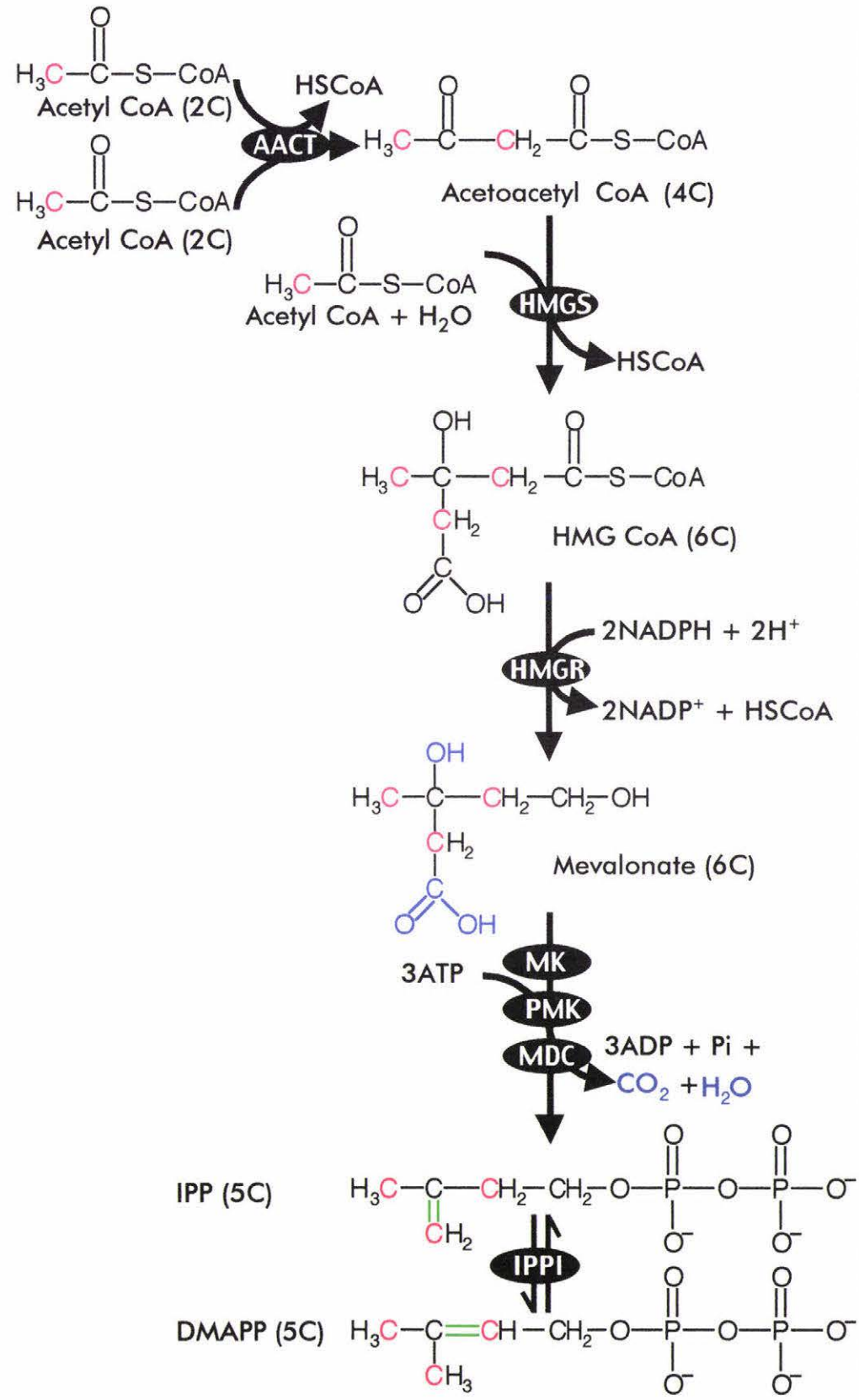

Acetyl CoA (2C)
HSCoA
AACT
Acetyl CoA (2C)
Acetoacetyl CoA (4C)
Acetyl CoA + H2O
HMGS
HSCoA
HMG CoA (6C)
2NADPH + 2H+
HMGR
2NADP+ + HSCoA
Mevalonate (6C)
3ATP
MK
PMK
MDC
3ADP + Pi +
CO2 +H2O
IPP (5C)
IPPI
DMAPP (5C)

Fig 1.2 **An overview of primary isoprenoid biosynthesis: -Dimethylallyl pyrophosphate to geranylgeranyl pyrophosphate**

B

Overview of the synthesis of geranylgeranyl pyrophosphate (GGPP) from dimethylallyl pyrophosphate (DMAPP).

The synthesis of long chain isoprenoid molecules can be achieved by the combination of various shorter intermediate molecules. This figure shows one possible pathway for the synthesis of GGPP.

DMAPP and isopentenyl pyrophosphate (IPP) condense to form a ten carbon geranyl pyrophosphate molecule. The DMAPP molecule is ionised through the loss of pyrophosphate to generate a carbonium ion.. The DMAPP carbonium ion is attacked by the double bond of IPP, both molecules condense and eliminate a proton to form a 10 carbon GPP molecule (Stryer, 1995).

This reaction, involving ionisation, condensation and elimination occurs for each successive addition of IPP to DMAPP, GPP or farnesyl pyrophosphate (FPP). In this manner long chain isoprenoids can be constructed from sequential addition of five carbon intermediates to the substrate molecule. The size of synthesised isoprenoids is limited by the specific prenyltransferase enzyme responsible for each catalysis.

FPP synthase catalyses both condensation of DMAPP and IPP, but also GPP and IPP in order to synthesise the 15 carbon FPP. In another case, Sonanesyl diphosphate synthase sequentially condenses IPP to form a 45 carbon isoprenoid (Ohnuma *et al.*, 1996).

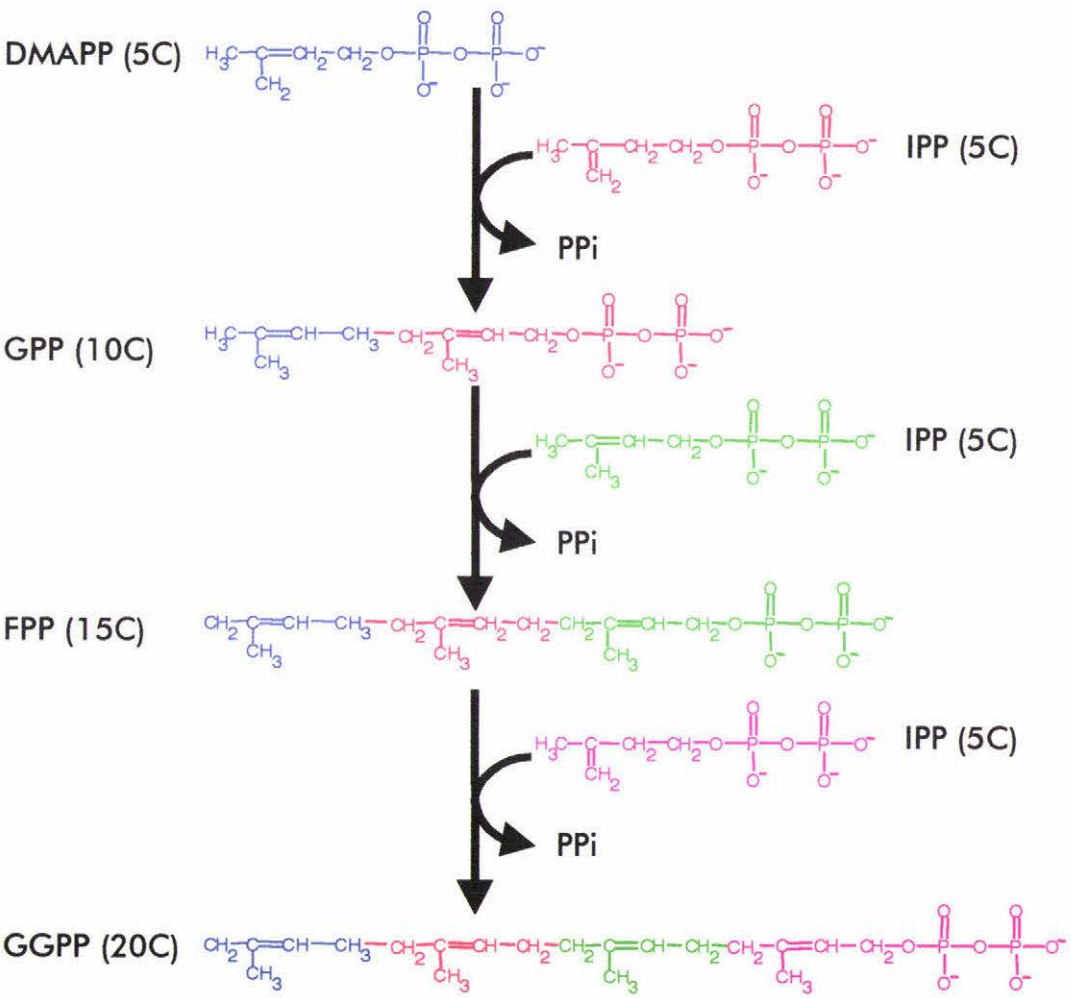
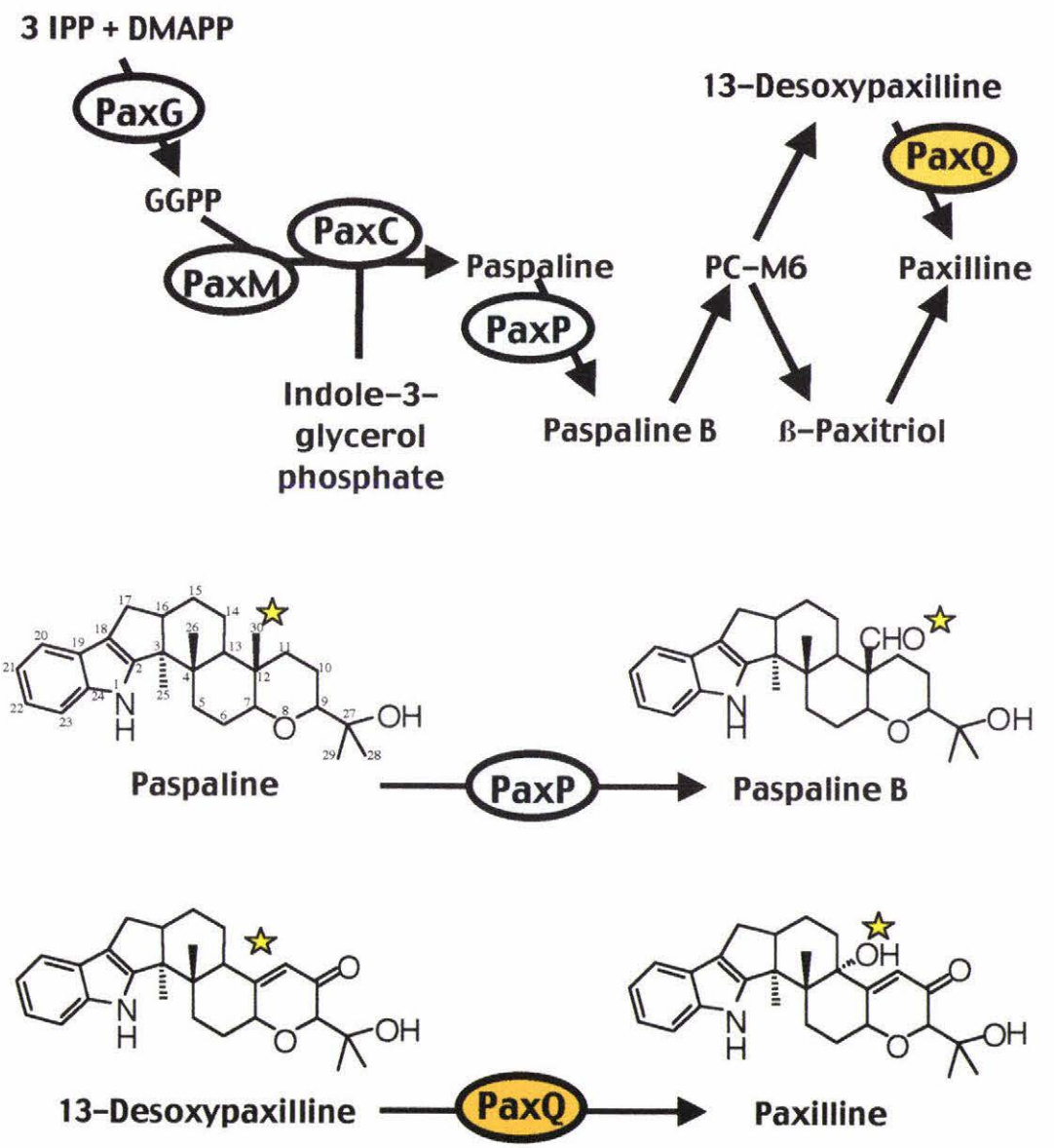

DMAPP (5C)
IPP (5C)
PPi
GPP (10C)
IPP (5C)
PPi
FPP (15C)
IPP (5C)
PPi
GGPP (20C)

Fig 1.3 **Proposed scheme for paxilline biosynthesis**

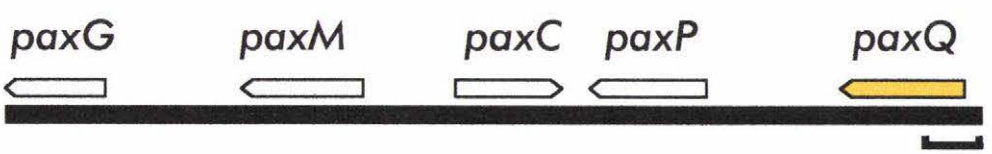


The scheme for paxilline biosynthesis is based upon the isolation of indole diterpenoids from *Penicillium* and other fungi (Munday-Finch, 1996), radiolabeling studies (Mantle, 1994. Byrne, 2000) and gene replacement of *paxP* and *paxQ* (McMillan, Carr, Young and Scott, unpublished results). The catalytic conversion illustrated for PaxP may require a series of catalytic reactions performed by a single enzyme.

1.1.2 The paxilline biosynthetic gene cluster

The enzymes responsible for paxilline biosynthesis are encoded by a gene cluster, which contains at least five members essential for paxilline production (Young *et al.*, 2001) (Figure 1.4). Genes proposed to encode enzymes for the early steps in the pathway include a GGPP synthase (*paxG*), a FAD dependant monooxygenase (*paxM*) and a prenyltransferase (*paxC*). Disruption of each of these genes results in a paxilline negative phenotype (McMillan, Young and Scott, unpublished results; Bryant and Scott, unpublished results). PaxG cDNA has been shown to complement an *Erwinia uredovora* carotenoid biosynthetic gene cluster lacking a functional GGPP synthase (Δ*crtE*) (Sandmann *et al.*, 1993). The co-transformation of both *paxG* and the mutated carotenoid biosynthetic cluster into *E. coli* resulted in the formation of carotenoids (Aravalli and Scott, unpublished results). This demonstrated that PaxG is a functional GGPP synthase. However, no identifiable indole-diterpene intermediates have yet been identified for PaxM or PaxC mutants. On the basis of these results and their predicted functions as an FAD-dependant monooxygenase and a prenyltransferase respectively, they are proposed to have a role in the addition of the indole moiety and cyclisation of GGPP.

Fig 1.4 **Overview of the paxilline biosynthetic gene cluster**



Gene	Proposed enzymatic function
paxG	Geranylgeranyl diphosphate synthase
paxM	FAD-dependant monooxygenase
paxC	Prenyl transferase
paxP	Cytochrome P450 monooxygenase
paxQ	Cytochrome P450 monooxygenase

The relative positions and orientations of five members of the paxilline biosynthetic cluster are shown. These five genes have been shown to be necessary for paxilline biosynthesis.

Two cytochrome P450 monooxygenase genes have been identified within the gene cluster responsible for paxilline biosynthesis (Figure 1.4). The deletion of *paxP* or *paxQ* results in mutant strains of *P. paxilli* that cannot synthesise paxilline (McMillan, Carr, Young and Scott, unpublished results). High pressure liquid chromatography (HPLC) analysis of the deletion strains revealed that the *paxP* mutant accumulated paspaline and the *paxQ* mutant accumulated 13-desoxypaxilline. These two indole diterpenoids have been isolated from *P. paxilli* previously and are listed as intermediates in a proposed metabolic grid for paxilline biosynthesis (Figure 1.3) (Munday-Finch *et al.*, 1996). Two cytochrome P450 enzymes required for the synthesis of the diterpene plant hormone gibberellin have been shown to perform multiple catalytic steps. P450-4 catalysed three steps (Tudzynski *et al.*, 2001) and P450-1 was shown to catalyse at least four sequential steps in gibberellin biosynthesis and possibly created up to 12 different products (Rojas *et al.*, 2001). This means that PaxP, PaxQ and possibly a dehydrogenase may be the only enzymes required to synthesise paxilline from paspaline (Figure 1.3). If PaxP and PaxQ are multifunctional enzymes then it is possible that the paxilline gene cluster contains fewer genes than was initially proposed by Young *et al.* (2001). This study focuses on further biochemical characterisation of PaxP and PaxQ.

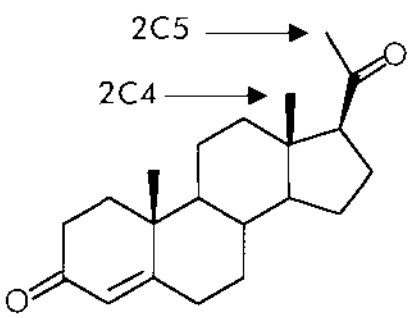
1.2 Cytochrome P450 enzymes

Cytochrome P450 enzymes are a large super-family of haem-thiolate proteins involved in the metabolism of a wide variety of compounds (Degtyarenko, 1995). They were first described in rat liver microsomes and are characterised by a large absorption maximum at 450 nm in the presence of carbon monoxide (Omura and Sato, 1964).

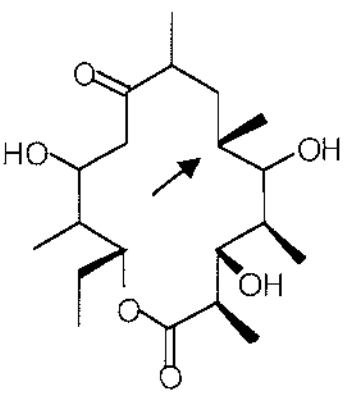
1.2.1 The function of cytochrome P450 enzymes

A diverse range of reactions are catalysed by cytochrome P450 enzymes, including monooxygenation, hydroxylation, dealkylation, epoxidation and reduction. Some examples of substrates modified by P450 enzymes include, progesterone (P450 2C5), 6-deoxerythronolide (P450eryF), camphor (P450cam), terpineol (P450terp) and fatty acids (P450BM-3) (Figure 1.5).

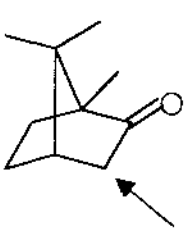
Fig 1.5 **The substrates of various hydroxylating cytochrome P450 enzymes**



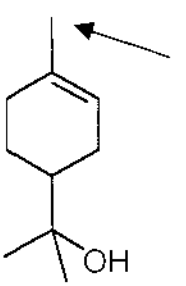
A Progesterone (P450 2C5/2C4)



B 6-deoxerythronolide B (P450eryF)

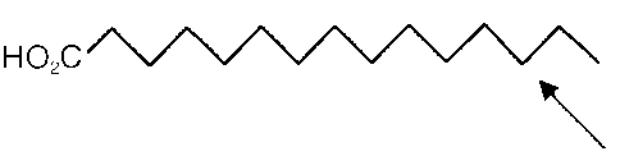


C Camphor (P450cam)



D Terpineol (P450terp)

E Fatty acids (P450 BM-3)



The site of modification is highlighted with an arrow.

The general reaction for hydroxylation is as follows:

$$\mathbf{R}H + O_2 + NAD(P)H + H^+ \longrightarrow \mathbf{R}OH + H_2O + NAD(P)^+$$

RH is the substrate and **R**OH is the product.

1.2.2 Cytochrome P450 reductase

Electrons donated from NAD(P)H are used to activate oxygen and allow its incorporation into the substrate molecule. The electrons are usually donated to the enzymes via a separate enzyme system. The electron donating system has either two parts (adrenodoxin and adrenodoxin reductase) in the case of mitochondrial and prokaryotic P450s or one member (cytochrome P450 reductase, CPR) for P450s in the endoplasmic reticulum (ER) (van den Brink *et al.*, 1998). In both cases the electron transfer system binds the NAD(P)H and passes the two electrons stepwise to the P450 haem. The CPR protein is roughly 80 kDa in mass (Masters and Okita, 1980) and contains one flavin dinucleotide (FAD) and one flavin mononucleotide (FMN) prosthetic group per enzyme. CPR is an integral membrane protein with a 6 kDa membrane spanning region and a 74 kDa cytosolic region (van den Brink *et al.*, 1998). The membrane spanning region anchors the protein to the ER or nuclear envelope (Kasper, 1971) ensuring it is oriented to allow interaction with a cytochrome P450 enzyme. The cytosolic domain contains two functional regions, one for binding FMN and the other binding both FAD and NADPH (Porter and Kasper, 1986). Electrons are transferred from NADPH to FAD, FMN and finally to the P450 enzyme. Only one *cpr* gene is found in most organisms which indicates a high degree of functional conservation in the enzyme, this is illustrated by the fact that mammalian P450 enzymes can be effectively reduced by *Saccharomyces cerevisiae* CPR *in vivo* (Yabusaki, 1995).

1.2.3 The cytochrome P450 reaction cycle.

The microsomal cytochrome P450 reaction cycle (Figure 1.6) involves:

(i) Binding of the substrate (Step 1) and reduction of the two flavin prosthetic groups of cytochrome P450 reductase by NADPH.

(ii) Transfer of one of the two available electrons to the cytochrome P450 (Step 2).

(iii) Binding of O_2 to give a ferrous cytochrome P450-dioxygen complex (Step 3).

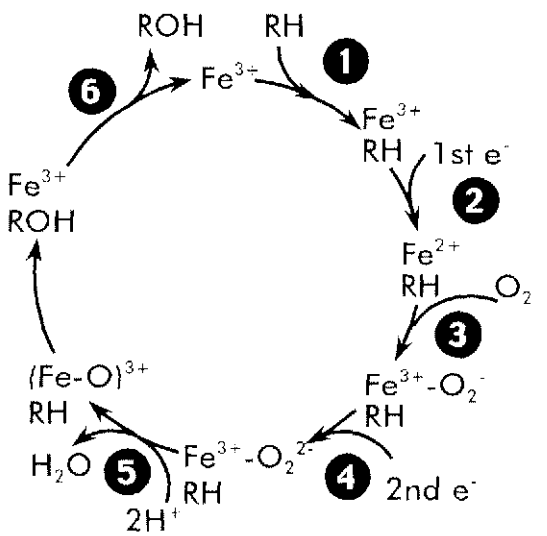
(iv) Transfer of a second electron from cytochrome P450 reductase, or an electron from cytochrome b_5 to the complex (Step 4).

(v) Cleavage of the oxygen-oxygen bond, and incorporation of one of the oxygen atoms to form a molecule of H_2O (Step 5).

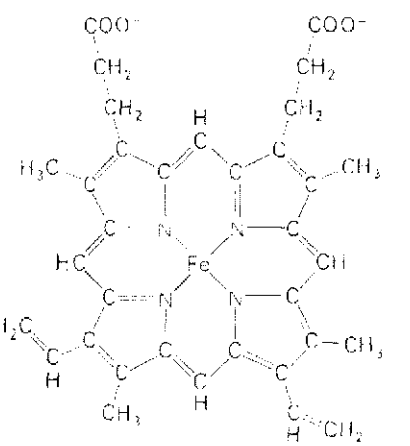
(vi) Transfer of the second oxygen atom to the substrate.

(vii) Dissociation of the product (Step 6).

Fig 1.6 **The cytochrome P450 reaction cycle and haem cofactor.**



The cytochrome P450 reaction cycle



The cytochrome P450 haem cofactor

The P450 reaction cycle.
Substrate is shown as **R**, electrons as **e**$^-$.

The cytochrome P450 haem cofactor.

The reaction cycle begins with the binding of substrate. This binding event alters the electron spin state of the haem prosthetic group (Figure 1.6) and aids the uptake of the first electron. The binding of substrate facilitates electron uptake because the ferric (Fe^{3+}) haem group in the P450 is more readily reduced to the ferrous (Fe^{2+}) state when the iron atom is coordinated to one rather than two axial ligands. Binding of substrate effects a change from a hexacoordinate to a pentacoordinate state and thus favours reduction of the haem group (Ortiz de Montellano, 1987).

The haem co-factor contains an iron atom which is reduced from the ferric to the ferrous state by the addition of the first electron. The oxidative reactions begin with molecular

oxygen binding to the ferrous P450 with coordination to iron *trans* to the cysteine thiolate. Transfer of the second electron then occurs (Ortiz de Montellano, 1987). The next step is not yet completely defined. It involves the splitting of the bound oxygen molecule along with the uptake of two protons to form an "activated oxygen" plus the release of H_2O. The addition of oxygen to the substrate is believed to involve removal of hydrogen from the substrate and the recombination of the resulting hydroxyl and carbon radicals to form the oxygenated product (Porter and Coon, 1991).

1.3 Cytochrome P450 structural elements

Structural data from cytochrome P450 enzymes has now identified substrate binding sites, water channels, membrane binding regions, sites for redox partner interaction and substrate intermediate formation. This level of understanding has revealed that while P450s have a high degree of sequence diversity their structural folds are highly conserved.

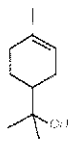
1.3.1 The cytochrome P450 superfamily

There are four different classes of P450s, the divisions determined by the type of redox partner used (Degtyarenko, 1995; Peterson and Graham, 1998). Bacterial and mitochondrial P450s are grouped into Class I. They utilise an FAD-containing NAD(P)H ferredoxin reductase and require an iron-sulphur protein to shift electrons between the reductase and the P450. P450cam is an example of a Class I enzyme (Figure 1.7). Class II P450 enzymes are present in the ER of eukaryotic cells, and require a FAD/FMN-containing NADPH P450 reductase to deliver electrons from NADPH. PaxP and PaxQ are predicted to be Class II P450s based on sequence similarity. Class III enzymes modify endoperoxide or hydroperoxide substrates. The presence of oxygen in these substrates negates the need for NAD(P)H or a reductase enzyme partner. Nitric oxide reductase (P450nor, isolated from *Fusarium oxysporum*) is the only known member of the Class IV P450s (Figure 1.7). It reduces nitric oxide (NO) and obtains electrons directly from NADH.

Fig 1.7 **Overview of available cytochrome P450 enzyme structures**

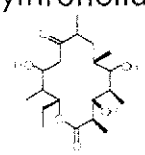
A Name: P450cam Substrate: Camphor
Organism: *Pseudomonas putida*
PDB code: 1DZ8,
(Poulos *et al.*, 1987)

B Name: P450terp Substrate: Terpineol
Organism: *Pseudomonas sp.*
PDB code: 1CPT
(Hasemann *et al.*, 1994)

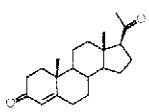


C Name: P450BM-3 haem domain. Substrate: Fatty acids
Organism: *Bacillus megaterium*
PDB code: 1BU7
(Ravichandran *et al.*, 1993; Sevrioukova *et al.*, 1999)

D Name: P450eryf Substrate: 6-Deoxerythronolide B
Organism: *Saccharopolyspora erythraea*
PDB code: 1OXA
(Cupp-Vickery and Poulos, 1995)



E Name: P450 2C5dh Substrate: Progesterone
Organism: *Oryctolagus cuniculus*
PDB code: 1DT6
(Williams *et al.*, 2000)

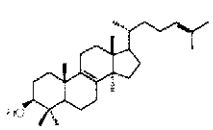


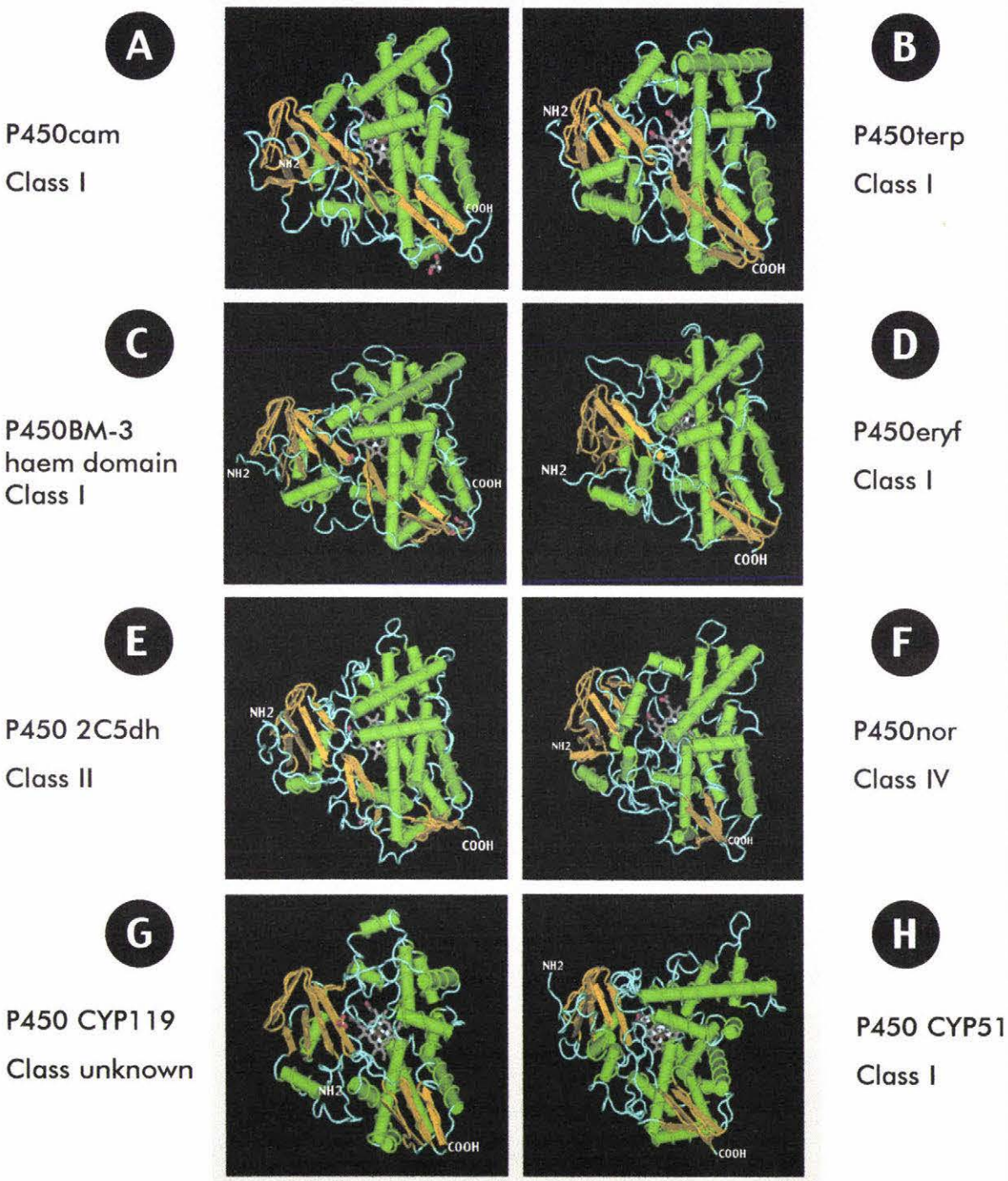
F Name: P450nor Substrate: Nitric oxide
Organism: *Fusarium oxysporum*
PDB code: 1ROM
(Park *et al.*, 1997)

N-O•

G Name: P450 CYP119 Substrate: Unknown
Organism: *Sulfolobus solfactaricus*
PDB code: 1F4U
(Yano *et al.*, 2000),

H Name: P450 CYP51 Substrate: Lanosterol
Organism: *Mycobacterium tuberculosis*
PDB code: 1EA1
(Podust *et al.*, 2001)





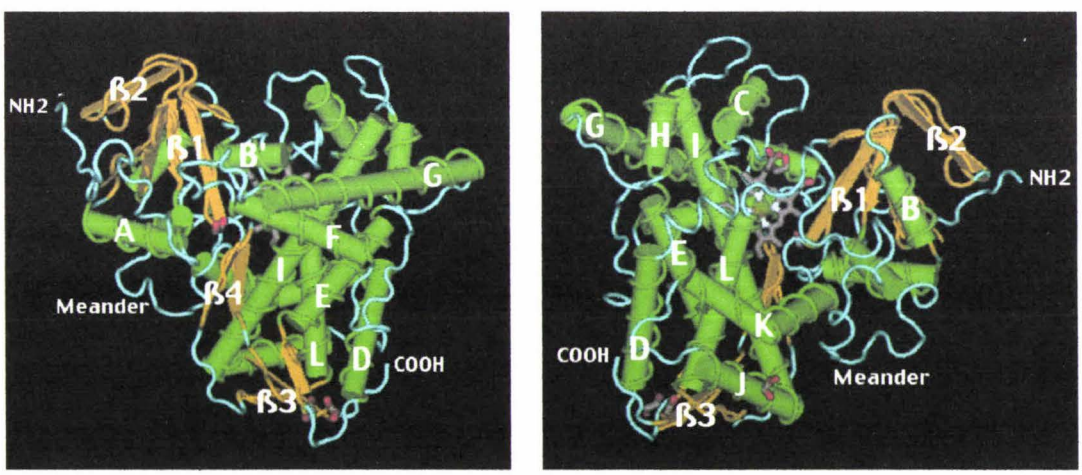
All structures are presented as a distal view. Beta sheet structures are drawn in orange, helical structures in green. General secondary structure designations are described in Figure 1.8. Views were generated using CN3D software (NCBI).

1.3.2 Structural features of the cytochrome P450 enzyme

The general structural fold of the cytochrome P450 super-family includes two regions, one that is predominantly alpha helical (α domain) and another smaller region composed of mostly beta-sheet structures (ß domain) (Poulos, 1995). These regions are not strictly domains, as they are both part of one globular enzyme, but they do exhibit different structural compositions. A general consensus of the known P450 structures shows that most contain 13 helices and four ß sheets (Figure 1.7).

The haem domain of cytochrome P450BM-3 from *Bacillus megaterium* is considered a good model for eukaryotic microsomal P450s and has high sequence identity with PaxP and PaxQ (amongst the available P450 structures) (Figure 1.8). The structure of the haem domain was determined in 1993 (Ravichandran *et al.*, 1993) and P450BM-3 is now one of the most well characterised P450s. P450BM-3 contains both a haem (P450) domain and a corresponding reductase in one holoenzyme.

Fig 1.8 **Overview of the P450BM-3 haem domain structure**



Distal view **Proximal view**

P450BM-3 haem domain drawn from the PDB coordinates of structure 1BU7A. The P450 haem can be seen in the centre of each view as a grey molecule. Alpha helices are coloured green, beta sheet structures are coloured orange. Secondary structural elements are labelled where possible. Image generated with CN3D software (NCBI).

The conserved core of the protein contains a four helix bundle (involving helices D, E, I, L) plus helices J and K (Figure 1.8). Helix K contains a Glu-x-x-Arg motif found in all P450s, which is thought to help stabilise the core of the protein. The L helix is involved in haem binding. The binding of fatty acid and eicosanoid substrates by P450BM-3 is performed by a hydrophobic access channel. The channel is created by beta sheets ß1, ß2, part of ß4 and includes α helices F and B' (Ravichandran *et al.*, 1993).

A summary of the cytochrome P450 enzymes for which 3D structures have been determined is shown in Figure 1.7. Six of the structures are from prokaryotic organisms including one thermophillic archea (CYP119). The two eukaryotic examples include a fungal P450 (P450nor) and a mammalian P450 (P450 2C5). These eight structures underline the structural conservation in the cytochrome P450 superfamily. Sequence conservation among these examples is low, yet all exhibit the same basic folds and the same basic structural elements.

1.3.2.1 The haem binding region

The α-domain contains the haem cofactor, which is essential for activity. The most conserved region of the enzyme is that which binds the haem cofactor. The haem forms a linkage to the protein via a thiolate bond between the haem iron and a sulphur present in a cysteine residue. The thiolate linkage forms the fifth bond to the iron centre, the other four bonds are to the porphyrin ring of the haem (Figure 1.6) and a sixth bond is formed with a water or oxygen molecule during catalysis. The presence of the cysteine residue is absolutely conserved in all P450 enzymes. A consensus sequence of F(G/S)xGx(H/R)x**C**xGxx(I/L/F)A is characteristic of the haem binding site (the conserved cysteine is listed in bold) (Graham-Lorence and Peterson, 1996). Another conserved structural element of P450s is the "meander", named as it initially appeared to be a random loop. Further study of this region revealed it to be a spatially conserved structure with sequence variation depending on the class of the P450.

1.3.2.2 Substrate binding regions

The variable regions of the peptide sequence are involved with binding of substrates and redox partner enzymes. Several areas have been identified as substrate interacting

regions. Analysis of the P450BM-3 structure (Ravichandran *et al.*, 1993), indicated that sheets ß1, ß2, ß4 and helices B' and F would form the substrate binding pocket. The pocket extends from the surface to close to the active site haem (Figure 1.8).

Using multiple sequence alignments and hydropathy index calculations Gotoh (1992) identified six putative substrate recognition sites (SRS's), based on the assumption that sequence variation was due to adaptation to different substrates. Regions within the B' helix, F helix , I helix and beta sheet ß1 were suggested as SRS's

1.3.2.3 Eukaryotic cytochrome P450 enzymes

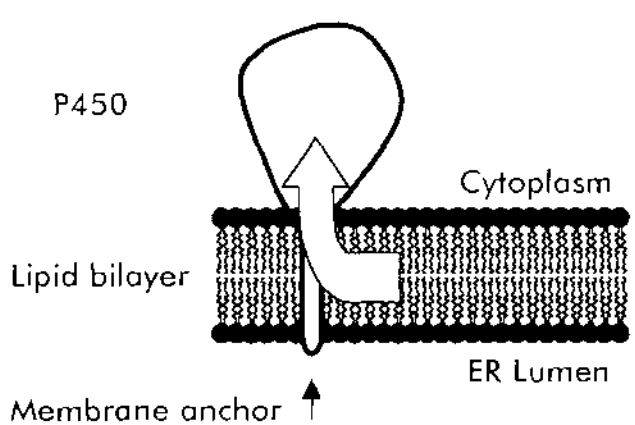
The structure of rabbit cytochrome P450 2C5 (Williams *et al.*, 2000) has more relevance than BM-3 to the potential structures of PaxP and PaxQ as it is a eukaryotic microsomal P450. It is the only mammalian cytochrome P450 structure for which there is a known 3D structure. In order to increase the solubility of the protein and promote crystallisation, this protein was expressed and crystallised without the N-terminal membrane anchor sequence. The modified protein was referred to as 2C5dh. Computer simulations using the 2C5dh structure showed a clear position for the progesterone substrate. It was bound by residues which all fell into the six SRS's predicted by Gotoh (Gotoh, 1992). The structural importance of residues in the active site is demonstrated by alanine-113 located in SRS-1. In other 2C family P450s, a valine is found at that position. Mutation of the valine in 2C4 to an alanine, perturbs the activity of 2C4 to that of 2C5, so that hydroxylation occurs at a slightly altered position on the substrate (Figure 1.5, part A).

1.3.2.4 Membrane topology of cytochrome P450

The cytochrome P450 2C5 structure is the first determined from a membrane binding P450. The N-terminal membrane anchor was removed in the crystallised enzyme (2C5dh) but structural features remain that indicate its membrane topology. All microsomal P450s are integral membrane proteins, attaching to the membrane mainly via a single transmembrane domain at the N-terminus of the enzyme. Mitochondrial P450s lack this domain and yet still bind membranes, an interaction that can be disrupted by high salt conditions. This suggests that the membrane interaction is determined by more than just the N-terminus. The exterior surface of the enzyme is thought to bind the membrane in a reversible fashion, a theory supported by the fact that

N-terminal modified 2C5dh still binds membranes in a salt reversible manner, similar to mitochondrial P450s (Williams *et al.*, 2000). Epitope tagging experiments with Class II P450s indicate that the N-terminal domain, F-G, A-B and B-C loops are involved in membrane attachment. The F-G loop in 2C5 is more suited to membrane attachment than the same region in BM-3 as it contains 21 residues compared to the 12 found in BM-3, and possesses more hydrophobic residues as well, 12 compared to 5 (Williams *et al.*, 2000). The N-terminal region of 2C5 is attached to the rest of the enzyme by a short (nine residue) linker region. The structure of this area was not able to be resolved, which indicates it has a high degree of flexibility. The linker contains basic residues that halt the translocation of the enzyme into the ER lumen (Szczesna-Skorupa *et al.*, 1988). This results in a transmembrane region with the remainder of the enzyme exposed to the cytoplasm (Figure 1.9). These membrane interaction regions are thought to orient the P450 so the substrate binding channel is buried in the membrane. Another putative access point has been determined which would allow passage from the cytoplasm. It was proposed that this could function as an exit path for substrates which have increased hydrophilicity after catalysis (Williams *et al.*, 2000).

Fig 1.9 **Orientation of membrane bound P450s**



The substrate binding and access channels are buried in the lipid bilayer.

The grey arrow depicts the path of substrate access.

1.4 Sequence alignment of fungal cytochrome P450 enzymes.

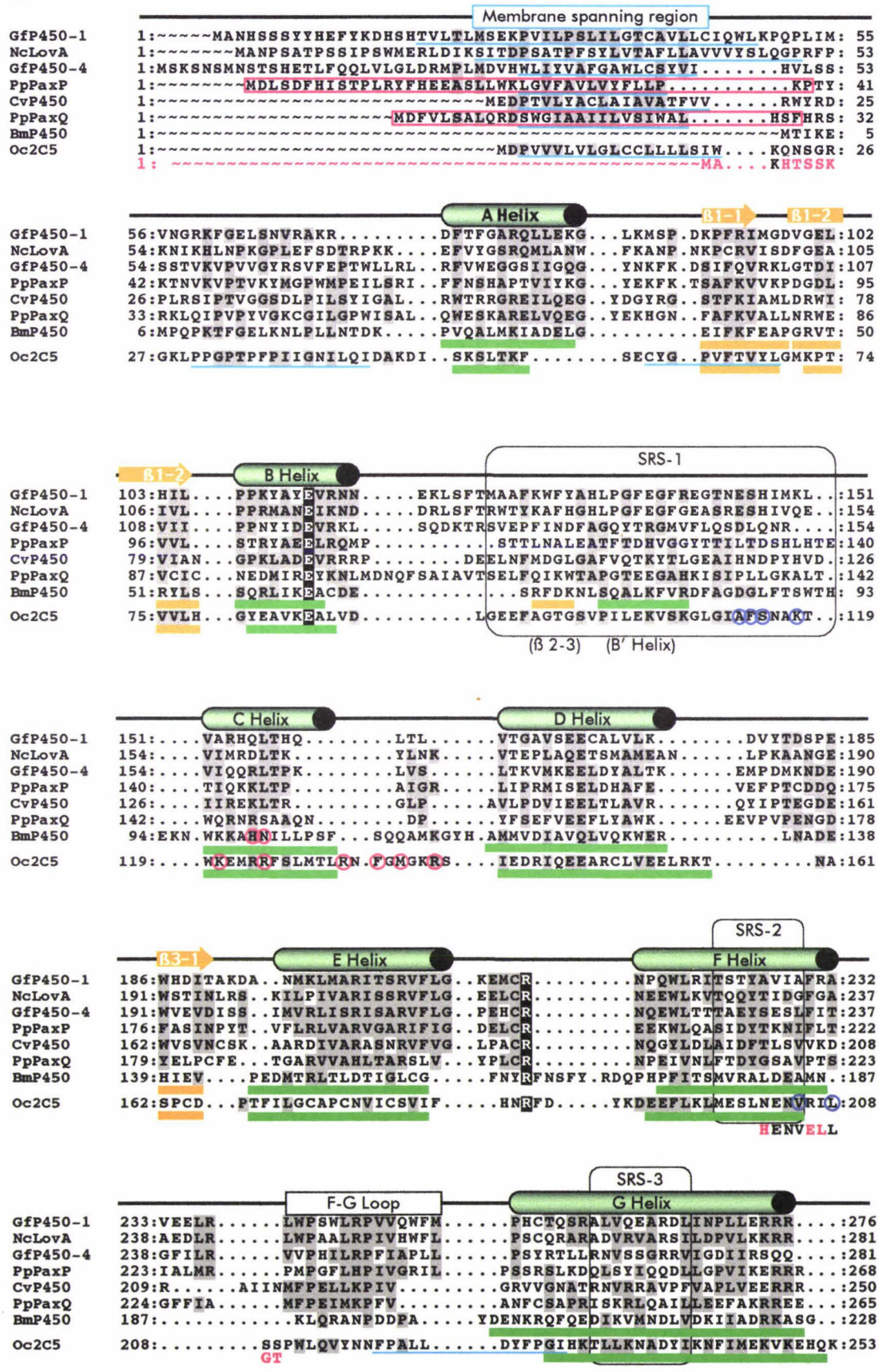
To further analyse the properties of the PaxP and PaxQ enzymes, a multiple sequence alignment was generated for PaxP and PaxQ including various fungal cytochrome P450 enzymes and two enzymes for which the structure was known (Figure 1.10). The alignment was performed so that regions proposed to be important for the enzymatic function of PaxP and PaxQ could be highlighted and better understood.

Fungal cytochrome P450 amino acid sequences were selected on the basis of their sequence identity to PaxP and PaxQ. The fungal enzyme sequences included were: *Gibberella fujikuroi P450-1* (GfP450-1 (Tudzynski and Hölter, 1998) accession no. CAA75565); *Neurospora crassa lovA* (NcLovA; accession no. CAB91316); *G. fujikuroi P450-4* (GfP450-4 (Tudzynski *et al.*, 2001) accession no. CAA76703); *Coriolus versicolor CYP512A1* (CvP450 (Ichinose *et al.*, 2002) accession no. BAB59027). Enzymes included for structural reference were P450BM-3 (PDB id 1BU7) and CYP2C5 (PDB id 1DT6). Although P450BM-3 is a prokaryotic enzyme, it is considered to be a good model for eukaryotic systems. CYP2C5 is a mammalian P450 and is the only membrane bound P450 to be structurally determined. From the cytochrome P450 enzymes for which a structure has been determined P450BM-3 and CYP2C5 share the highest degree of sequence identity with PaxP and PaxQ. The helix and beta sheet secondary structures of BM-3 and CYP2C5 are highlighted with green and orange bars respectively (Figure 1.10). The tertiary structure of P450BM-3 is shown in Figure 1.8, while a dissected view of the secondary structure of CYP2C5 is shown in Figure 1.10 part B.

The CYP2C5 enzyme, which was crystallised by Williams *et. al.* (2000), was modified in two regions (Cosme and Johnson, 2000). The N-terminal transmembrane region was deleted and the C terminus of the F helix modified. The two modified regions are shown in red type under the Oc2C5 sequence. These modifications were carried out to facilitate the heterologous expression of soluble enzyme in *E. coli*.

Fig 1.10 Sequence alignment of fungal cytochrome P450 enzymes

A



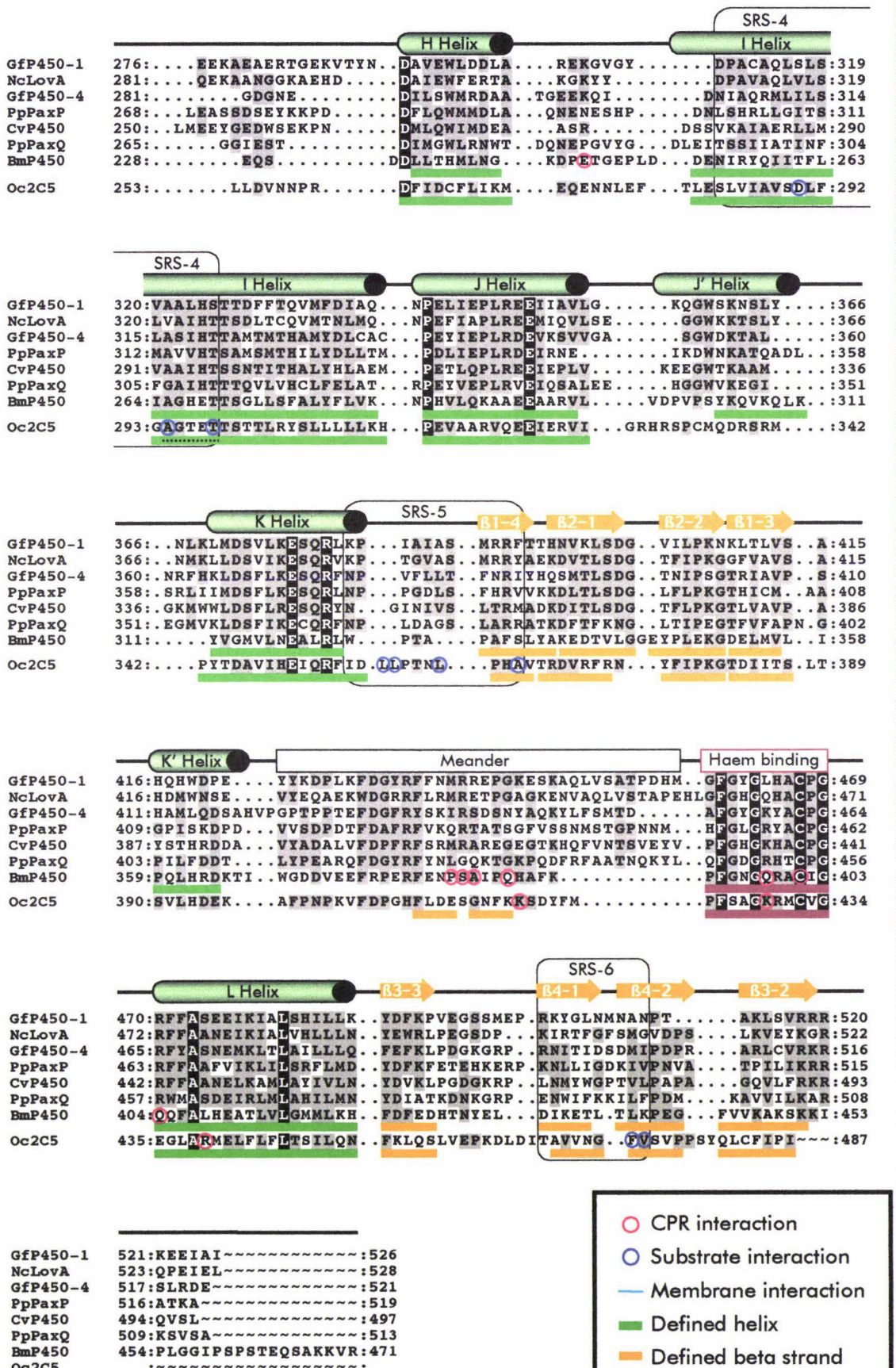
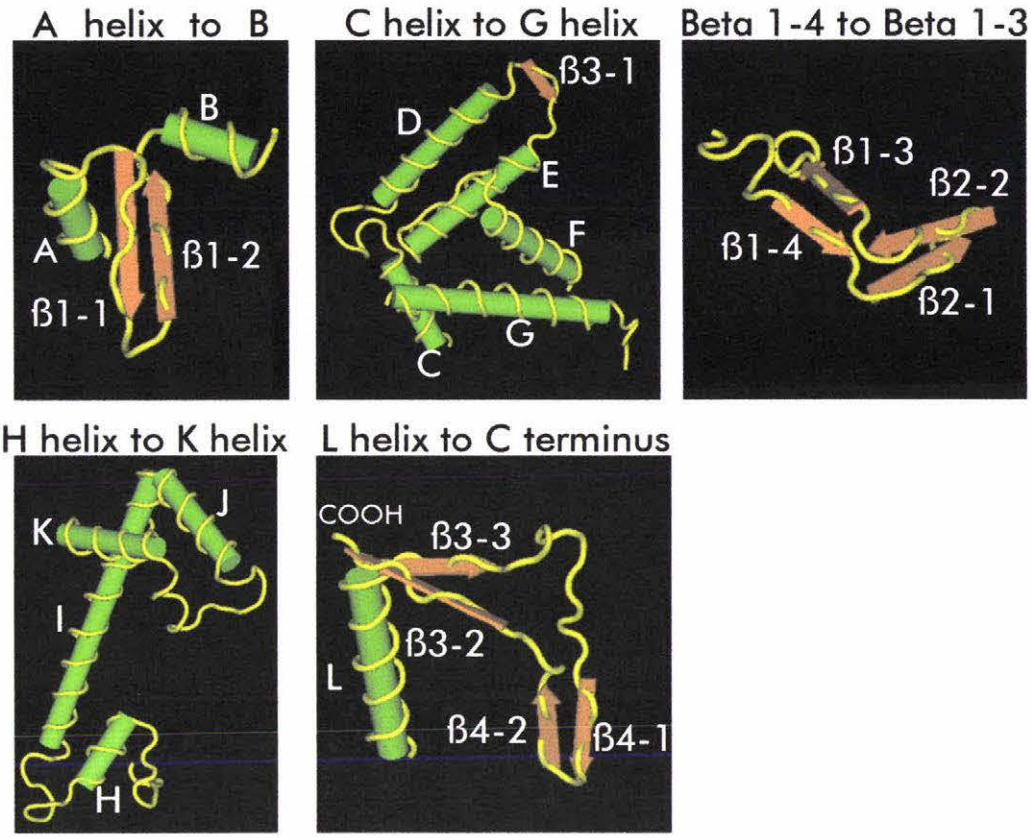

SRS-4
H Helix
I Helix
GfP450-1 276:....EEKAEAERTGEKVTYN..DAVEWLDDLA....REKGVGY.....DPACAQLSLS:319
NcLovA 281:....QEKAANGGKAEHD.....DAIEWFERTA....KGKYY.....DPAVAQLVLS:319
GfP450-4 281:.....GDGNE.....DILSWMRDAA..TGEEKQI.....DNIAQRMLILS:314
PpPaxP 268:...LEASSDSEYKKPD.....DFLQWMMDLA...QNENESHP.....DNLSHRLLGITS:311
CvP450 250:..LMEEYGEDWSEKPN.....DMLQWIMDEA....ASR.....DSSVKAIAERLLM:290
PpPaxQ 265:.....GGIEST.....DIMGWLRNWT..DQNEPGVYG...DLEITSSIIATINF:304
BmP450 228:.....EQS.....DDLLTHMLNG....KDPETGEPLD...DENIRYQIITFL:263
Oc2C5 253:.....LLDVNNPR.....DFIDCFLIKM....EQENNLEF...TLESLVIAVSDLF:292
SRS-4
I Helix
J Helix
J' Helix
GfP450-1 320:VAALHSTTDFFTQVMFDIAQ...NPELIEPLREEIIAVLG.....KQGWSKNSLY....:366
NcLovA 320:LVAIHTTSDLTCQVMTNLMQ...NPEFIAPLREEMIQVLSE.....GGWKKTSLY....:366
GfP450-4 315:LASIHTTAMTMTHAMYDLCAC...PEYIEPLRDEVKSVVGA.....SGWDKTAL.....:360
PpPaxP 312:MAVVHTSAMSMTHILYDLLTM...PDLIEPLRDEIRNE.....IKDWNKATQADL..:358
CvP450 291:VAAIHTSSNTITHALYHLAEM...PETLQPLRREIEPLV.....KEEGWTKAAM.....:336
PpPaxQ 305:FGAIHTTTQVLVHCLFELAT...RPEYVEPLRVEIQSALEE.....HGGWVKEGI.....:351
BmP450 264:IAGHETTSGLLSFALYFLVK...NPHVLQKAAEEAARVL.....VDPVPSYKQVKQLK..:311
Oc2C5 293:GAGTETTSTTLRYSLLLLLKH...PEVAARVQEEIERVI...GRHRSPCMQDRSRM....:342
K Helix
SRS-5
ß1-4
ß2-1
ß2-2
ß1-3
GfP450-1 366:..NLKLMDSVLKESQRLKP...IAIAS..MRRFTTHNVKLSDG..VILPKNKLTLVS..A:415
NcLovA 366:..NMKLLDSVIKESQRVKP...TGVAS..MRRYAEKDVTLSDG..TFIPKGGFVAVS..A:415
GfP450-4 360:.NRFHKLDSFLKESQRFNP...VFLLT..FNRIYHQSMTLSDG..TNIPSGTRIAVP..S:410
PpPaxP 358:.SRLIIMDSFLKESQRLNP...PGDLS..FHRVVKKDLTLSDG..LFLPKGTHICM..AA:408
CvP450 336:.GKMWWLDSFLRESQRYN...GINIVS..LTRMADKDITLSDG..TFLPKGTLVAVP..A:386
PpPaxQ 351:.EGMVKLDSFIKECQRFNP...LDAGS..LARRATKDFTFKNG..LTIPEGTFVFAPN.G:402
BmP450 311:.....YVGMVLNEALRLW....PTA....PAFSLYAKEDTVLGGEYPLEKGDELMVL..I:358
Oc2C5 342:....PYTDAVIHEIQRFID.LLPTNL....PHAVTRDVRFRN...YFIPKGTDIITS.LT:389
K' Helix
Meander
Haem binding
GfP450-1 416:HQHWDPE....YYKDPLKFDGYRFFNMRREPGKESKAQLVSATPDHM..GFGYGLHACPG:469
NcLovA 416:HDMWNSE....VYEQAEKWDGRRFLRMRETPGAGKENVAQLVSTAPEHLGFGHGQHACPG:471
GfP450-4 411:HAMLQDSAHVPGPTPPTEFDGFRYSKIRSDSNYAQKYLFSMTD.....AFGYGKYACPG:464
PpPaxP 409:GPISKDPD...VVSDPDTFDAFRFVKQRTATSGFVSSNMSTGPNNM...HFGLGRYACPG:462
CvP450 387:YSTHRDDA...VYADALVFDPFRFSRMRAREGEGTKHQFVNTSVEYV..PFGHGKHACPG:441
PpPaxQ 403:PILFDDT....LYPEARQFDGYRFYNLGQKTGKPQDFRFAATNQKYL..QFGDGRHTCPG:456
BmP450 359:PQLHRDKTI..WGDDVEEFRPERFENPSAIPQHAFK.....PFGNGQRACIG:403
Oc2C5 390:SVLHDEK....AFPNPKVFDPGHFLDESGNFKKSDYFM.....PFSAGKRMCVG:434
SRS-6
L Helix
ß3-3
ß4-1
ß4-2
ß3-2
GfP450-1 470:RFFASEEIKIALSHILLK..YDFKPVEGSSMEP.RKYGLNMNANPT.....AKLSVRRR:520
NcLovA 472:RFFAANEIKIALVHLLLN..YEWRLPEGSDP...KIRTFGFSMGVDPS....LKVEYKGR:522
GfP450-4 465:RFYASNEMKLTLAILLLQ..FEFKLPDGKGRP..RNITIDSDMIPDPR....ARLCVRKR:516
PpPaxP 463:RFFAAFVIKLILSRFLMD..YDFKFETEHKERP.KNLLIGDKIVPNVA....TPILIKRR:515
CvP450 442:RFFAANELKAMLAYIVLN..YDVKLPGDGKRP..LNMYWGPTVLPAPA....GQVLFRKR:493
PpPaxQ 457:RWMASDEIRLMLAHILMN..YDIATKDNKGRP..ENWIFKKILFPDM....KAVVILKAR:508
BmP450 404:QQFALHEATLVLGMMLKH..FDFEDHTNYEL...DIKETL.TLKPEG...FVVKAKSKKI:453
Oc2C5 435:EGLARMELFLFLTSILQN..FKLQSLVEPKDLDITAVVNG..GVSVPPSYQLCFIPI~~~:487
GfP450-1 521:KEEIAI~~~~~~~~~~~~:526
NcLovA 523:QPEIEL~~~~~~~~~~~~:528
GfP450-4 517:SLRDE~~~~~~~~~~~~:521
PpPaxP 516:ATKA~~~~~~~~~~~~~~:519
CvP450 494:QVSL~~~~~~~~~~~~~~:497
PpPaxQ 509:KSVSA~~~~~~~~~~~~~:513
BmP450 454:PLGGIPSPSTEQSAKKVR:471
Oc2C5 :~~~~~~~~~~~~~~~~~~:
CPR interaction
Substrate interaction
Membrane interaction
Defined helix
Defined beta strand

Fig 1.10 **Summary of Oc2C5 secondary structure**

B



A Sequence alignment of fungal cytochrome P450 enzymes

Sequences were aligned and shaded as described in Section 2.14. Coloured bars indicate the positions of structural features determined for the haem protein of P450 BM-3 (Ravichandran *et al.*, 1993; Sevrioukova *et al.*, 1999) and CYP2C5 (Williams *et al.*, 2000). Cylinders and arrows show predicted regions of secondary structure for the aligned sequences. Substrate recognition sites (SRS) described by Gotoh (1992) are boxed and labelled. Possible transmembrane regions are underlined in light **blue**. Sequences include: *Gibberella fujikuroi P450-1* (GfP450-1 (Tudzynski and Hölter, 1998) accession no. CAA75565). *Neurospora crassa lovA* (NcLovA; accession no. CAB91316). *G. fujikuroi P450-4* (GfP450-4 (Tudzynski *et al.*, 2001) accession no. CAA76703). *P. paxilli paxP* (PpPaxP (Young *et al.*, 2001) accession no. AAK11528). *Coriolus versicolor CYP512A1* (CvP450 (Ichinose *et al.*, 2002) accession no. BAB59027). *P. paxilli paxQ* (PpPaxQ (Young *et al.*, 2001) accession no. AAK11527). *Bacillus megaterium* P450BM-3 (BmP450 (Sevrioukova *et al.*, 1999) PDB 1BU7, accession no. 6729906). *Oryctolagus cuniculus CYP2C5* (Oc2C5 (Pendurthi *et al.*, 1990) PDB 1DT6, accession no. O4RBP4).

Multiple sequence alignments were generated using the University of Wisconsin GCG PILEUP program. PILEUP generated pairwise alignments of each sequence which were combined into a final alignment using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970).

B Summary of Oc2C5 secondary structure

Selected regions of the CYP2C5 3D structure (PDB Id 1DT6) are shown with helices drawn in green and beta strands drawn in orange. Structural elements for each section of the alignment are labeled in white. Structures drawn with CN3D (NCBI)

1.4.1 Sequence alignment construction

The polypeptide sequences of the selected enzymes were aligned using the PILEUP program from the GCG bioinformatics package (Section 2.14). A gap open penalty of 4 and gap length penalty of 1 was used to constrain the alignment. The alignment was shaded on the basis of amino acid functional similarity (Section 2.14). Comparison of aligned regions with the known secondary structure of P450BM-3 and CYP2C5 allowed the position of helices, beta strands and other conserved regions in the other proteins to be predicted. Information for the positioning of secondary structure was primarily taken from that published by Graham-Lorence and Peterson (1996). Additional information was derived from the PaxP and PaxQ tertiary structure predictions made by the FUGUE computer program (Shi *et al.*, 2001) (http://www-cryst.bioc.cam.ac.uk/~fugue/). Predicted structural features for the fungal enzymes are highlighted with cylinders (helices), arrows (beta strands) and boxes (other features) above the aligned sequences. Gotoh (1990) predicted the polypeptide regions that form substrate recognition sites (SRS) in cytochrome P450 enzymes. The six SRS were mapped onto the alignment using secondary structure and sequence similarity to determine their positions. A possible transmembrane region which is characteristic of eukaryotic cytochrome P450s was also assigned to each sequence (underlined in light blue). The basis for this tentative prediction was the presence of a continuous sequence of uncharged amino acids located at the N-terminus of the aligned polypeptides. The N-terminal regions deleted from PaxP and PaxQ for the Δ1-39PaxP and Δ1-29PaxQ expression constructs (Section 3.4) are boxed in red and include the proposed membrane spanning region.

Alignment of fungal P450 enzymes revealed the most highly conserved regions were those responsible for haem binding and cytochrome P450 reductase (CPR) interaction. Regions at the N-terminal of the polypeptide generally tended to be less conserved. The SRS regions also showed reduced homology.

1.4.2 Cytochrome P450 secondary structure

The haem binding region contains the absolutely conserved cysteine residue and shows the highest degree of sequence conservation in the enzyme. The essential cysteine is required for the formation of a covalent linkage to the haem iron. The alignment of

fungal enzymes showed good agreement with the consensus sequence F(G/S)xGx(H/R) xCxGxx(I/L/F)A as published by Graham-Lorence and Peterson (1996). Variation from the consensus occurred at the sixth position with a tyrosine in place of the predicted histidine or arginine in the PpPaxP and GfP450-4 sequences. Also, a methionine and a tyrosine were present in PpPaxQ and GfP450-4 respectively, at the 13th position of the consensus. Both substitutions can be considered conserved in nature.

The K helix contains the absolutely conserved ExxR motif. These two residues form a salt bridge and are positioned facing into the meander region of the enzyme, which interacts with electron donating enzymes.

The I helix is the longest helix in the enzyme. It sandwiches the haem group between itself and the haem-binding region. The I helix contains the consensus sequence (A/G)Gx(E/D)T (underscored on the alignment). PaxP and PaxQ follow this consensus closely, with the main exception being a histidine at the fourth position of the motif. This residue position faces the haem group in both CYP2C5 and P450BM-3 but is closest to the hydrophobic vinyl side chains of the haem, rather than the negatively charged propionate side chains. The centre of the I helix in P450BM-3 is disrupted from residues Ile 264 to Thr 269. In this region, the helix is slightly bent and stretched out. CYP2C5 also exhibits a similar structure in the same region. The helix distortion creates a groove which binds a water molecule in P450BM-3, CYP2C5 and P450Cam (Ravichandran *et al.*, 1993; Williams *et al.*, 2000). This region is highly conserved in all the sequences aligned, which suggests it performs an important enzymatic function.

The L helix is the most conserved helix in the alignment. The high degree of conservation is expected considering its close proximity to the haem group and redox-partner binding site. The arginine seen at the beginning of the L helix in fungal enzymes is likely to be involved in redox partner binding.

The C helix contains two conserved residues, listed as Trp 119 and Arg 123 in the Oc2C5 sequence. The arginine is charged-paired to one of the propionate side chains of the haem group (Figure 1.6). This arginine is present in PaxQ and substituted for a similarly charged Lysine in PaxP. The conserved tryptophan is thought to be involved

in electron transfer and yet is usually absent from prokaryotic P450 enzymes (Lewis and Hlavica, 2000). Interestingly, this residue is present in PaxQ, yet absent from PaxP.

The F-G loop of CYP2C5 has been implicated in membrane interaction (Williams *et al.*, 2000). This loop has a high degree of conservation among the aligned fungal enzymes. The fungal sequences in this region contain amino acids which are predominantly hydrophobic or neutral in character.

1.4.3 Redox partner interactions

Regions involved in the binding of the CPR enzyme also exhibit sequence conservation. Sevrioukova *et al.* (1999) have solved the 3D structure of the haem domain of P450BM-3 in complex with the corresponding P450BM-3 FMN binding domain (PDB id 1BVY). This complex revealed regions important for the interaction between the electron donating and accepting enzymes. Specifically, helices C, L and the meander region of the haem domain were shown to provide the necessary protein-protein interactions. Contacts formed include, two hydrogen bonds, one salt bridge and several water-mediated hydrogen bonds. The residues involved are His 101, Asn 102, Glu 245, Pro 383, Ser 384, Ala 385, Gln 388, Gln 398, Cys 401 and Gln 404. These residues are circled in red in Figure 1.10 A. Of these residues, the most conserved among the aligned sequences is Cys401 (required for haem binding) and His 101. Histidine 101 forms a salt bridge with Glu 494 of the flavin domain and is conserved among cytochrome P450s as a basic amino acid.

Mutagenesis of the CYP 2B4 enzyme identified residues that are important for its interaction with CPR. The CYP 2B4 residues were mapped onto the closely related CYP 2C5 structure by Williams *et al.* (2000), the positions which are revevant to CPR binding are highlighted with red circles in Figure 1.10 A. The residues identified are Lys 120, Arg 124, Arg 131, Phe 133, Met 135, Arg 138, Lys 418, Lys 429 and Arg 439.

1.4.4 The substrate binding site

Amino acids important for the substrate binding site of CYP2C5 were also characterised by Williams *et al.* (2000). A computer simulation of progesterone binding was used to identify the amino acids that held the substrate in place. The residues identified

included Ile 102, Ala 113, Phe 114, Val 205, Leu 208, Asp 290, Ala 294, Thr 298, Leu 358, Leu 359, Leu 363 and Phe 474. These residues have been highlighted in Figure 1.10 A with pink circles. All of these residues are positioned in or near the six SRS regions predicted by Gotoh (1992). Concordant with the hydrophobic nature of cytochrome P450 substrates, SRS residues generally tend to be hydrophobic. However, these SRS regions are also some of the most variable sections of the enzyme. The substrate binding pocket of P450BM-3 is defined as sheet ß1 plus residues 14 to 25, B' and F helices and also beta sheet 4.

1.5 Aim of this study

The aim of this study was to determine the substrates used and the products synthesised by PaxP and PaxQ. To achieve this goal, both enzymes were expressed as recombinant fusion proteins in *E. coli*. The initial hypothesis was based upon gene replacement studies and a proposed metabolic pathway for paxilline biosynthesis. The hypothesis was that PaxP catalysed the conversion of paspaline to paspaline B and PaxQ catalysed the conversion of 13-desoxypaxilline to paxilline. The purification of these enzymes and subsequent *in vitro* enzyme assays would allow the precise identification of substrates and synthesised products.

Chapter 2 Materials and methods

2.1 Media

All media were made up in sterile H_2O and autoclaved at 121°C for 15 minutes.

2.1.1 Luria-Bertani medium

Luria-Bertani (LB) broth contained; 0.85 mM NaCl, 1% (w/v) Bactotryptone, 0.5% (w/v) yeast extract (Difco) (pH 7.0–7.5). LB agar contained an additional 1.5% (w/v) agar (Gibco).

2.1.2 SOC medium

SOC medium contained; 2% (w/v) Bactotryptone, 0.5% (w/v) yeast extract (Difco), 20 mM glucose, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2.5 mM KCl.

2.1.3 Media additions

2.1.3.1 Ampicillin

Stock ampicillin solution contained; 0.29 M ampicillin. A working concentration of 0.29 mM was used. The stock solution was stored at -20°C.

2.1.3.2 Chloramphenicol

Chloramphenicol stock solution contained; 30 mM chloramphenicol dissolved in methanol. A working concentration of 60 µM was used. Stock solution was stored at -20°C.

2.1.3.3 IPTG

Isopropyl-beta-D-thiogalactopyranoside (IPTG) stock solution contained 0.84 M IPTG. The solution was sterilised by passing through a 0.45 µm filter (Gelman) and stored at -20°C.

2.1.3.4 Kanamycin

Kanamycin stock solution contained 20 mM kanamycin. A working concentration of 60 µM was used. The solution was sterilised by passing through a 0.45 µm filter (Gelman) and stored at -20°C.

2.1.3.5 Spectinomycin

Spectinomycin stock solution contained 0.3 M spectinomycin. A working concentration of 0.3 mM was used. The solution was sterilised by passing through a 0.45 µm filter (Gelman) and stored at -20°C.

2.1.3.6 Tetracycline

Tetracycline stock solution contained 20 mM tetracycline in methanol. A working concentration of 20 µM was used. The stock solution was stored at -20°C.

2.1.3.7 X-Gal

5-Bromo-4-chloro-3-indolyl-beta-D-galactoside (X-Gal) solution contained; 50 mM X-Gal dissolved in dimethyl-formamide. The stock solution was stored at -20°C.

2.1.4 *E. coli* cell lines

2.1.4.1 BL21

Genotype; F⁻, *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm*. Supplier; Invitrogen

2.1.4.2 Origami B

Genotype; F⁻, *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm lacY1 gor522::Tn10*(Tc^R) *trxB*::kan (DE3). Supplier; Invitrogen.

2.1.4.3 Top10

Genotype; F⁻, *mcr*AΔ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lacX*74 *deoR recA*1 *araD*139 Δ(*ara-leu*)7697 *galU galK rpsL endA*1 *nupG*

Supplier; Invitrogen.

2.1.4.4 XL1-blue

Genotype; *recA*1, *endA*1, *gyrA*46, *thi*, *hsdR*17, *supE*44, *posA*1, λ⁻, *lac*⁻, [F' *proAB*⁺, lac*I*^{q}, *lacZ*ΔM15, Tn10(Tet^R)]. (Bullock *et al.*, 1987)

2.1.5 Plasmids

Plasmids used are listed in Table 2.1. Selected vector maps are shown in Appendix 6.1.

Table 2.1 Plasmids

Name	Insert	Base vector	Size	Selection	Description	Reference
pUC118			3.2 kb	Ampicillin	Cloning vector	(Vieira and Messing, 1987)
pGEM-T-easy			3.0 kb	Ampicillin	T-tailed Cloning vector	Promega
pGEX-6P-3			4.9 kb	Ampicillin	GST fusion expression vector	Amersham Biosciences
pThioHis-A			4.4 kb	Ampicillin	Thioredoxin fusion expression vector	Invitrogen
pSJS1240	tRNAs, Arg (AGA AGG) Ile (AUA).		5.9 kb	Spectinomycin	Rare tRNA expression vector	(Del Tito *et al.*, 1995)
pGroESL	GroES/L chaperonins		6.5 kb	Chloramphenicol	GroES/L expression vector	(Goloubinoff *et al.*, 1989)
pRL1	*paxP* cDNA	pGEM-T-easy	4.6 kb	Ampicillin	*paxP* cDNA clone	This study
pRL2	*paxP* cDNA	pGEX-6P-3	6.4 kb	Ampicillin	GST-PaxP expression vector	This study
pRL3	*paxQ* cDNA	pGEM-T-easy	4.6 kb	Ampicillin	*paxQ* cDNA clone	This study
pRL4	*paxQ* cDNA	pGEX-6P-3	6.4 kb	Ampicillin	GST-PaxQ expression vector	This study
pRL5	Δ1-39*paxP* cDNA	pUC118	4.7 kb	Ampicillin	*npaxP* clone	This study
pRL6	Δ1-29*paxQ* cDNA	pUC118	4.7 kb	Ampicillin	*npaxQ* clone	This study
pRL7	Δ1-39paxP cDNA	pGEX-6P-3	6.3 kb	Ampicillin	GST-nPaxP expression vector	This study
pRL8	Δ1-29paxQ cDNA	pGEX-6P-3	6.3 kb	Ampicillin	GST-nPaxQ expression vector	This study
pRL9	Δ1-39paxP cDNA	pThioHis-A	5.9 kb	Ampicillin	Thioredoxin-nPaxP expression vector	This study
pRL10	Δ1-29paxQ cDNA	pThioHis-A	5.9 kb	Ampicillin	Thioredoxin-nPaxQ expression vector	This study

2.2 Common buffers

All buffers were made in sterile H_2O unless otherwise advised.

2.2.1.1 PBS buffer

PBS buffer contained; 140 mM NaCl, 10 mM Na_2HPO_4, 2.7 mM KCl, 1.8 mM KH_2PO_4, (pH 7.3).

2.2.1.2 STE buffer

STE buffer contained; 150 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM ethylene-diaminetetraacetic acid disodium salt (Na_2EDTA).

2.2.1.3 TE buffer

TE buffer (10:1) contained; 10 mM Tris-HCl (pH 8.0), 1 mM Na_2EDTA.

2.3 Polymerase chain reaction

2.3.1 PCR reagents

All PCR reagents were made in sterile H_2O unless otherwise advised.

2.3.1.1 DNA polymerase

Taq DNA polymerase (Roche), Expand high fidelity PCR system (Roche) or *Pfu* turbo DNA polymerase (Stratagene) were used in PCR reactions.

2.3.1.2 dNTP solution

dNTP solution contained 1.25 mM dATP.Li salt, 1.25 mM dCTP.Li salt, 1.25 mM dGTP.Li salt, 1.25 mM dTTP.Li salt.

2.3.1.3 10 x PCR buffer

Taq DNA polymerase 10 x buffer (Roche) contains 100 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 500 mM KCl.

Cloned *Pfu* DNA polymerase reaction buffer (10 x) (Stratagene) contains 200 mM Tris-HCl (pH 8.8), 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 1% Triton X-100, 1 mg/mL BSA.

2.3.1.4 Primers

Primer solutions were diluted to a final concentration of 10 µM in sterile H_2O. Primers used are listed in Table 2.2. The annealing positions of primers used for the PCR amplification of *paxP* and *paxQ* are shown in Appendix 6.3 and 6.4 respectively.

Table 2.2 Primers

Name	Sequence 5'–3' (1)	Site
CYP-3		*paxP*
CYP-8		*paxP*
M13(lacZ) Forward		pUC vector
M13(lacZ) Reverse		pUC vector
Oligo dT	GAG AGA ATT CGG ATC CTC TAG AGT	Poly-A tail
P2RL1		*paxQ*
P1RL3		*paxP*
P1RL5		*paxP*
P2RL6		*paxQ*
P1RL7		*paxP*
P1RL9		*paxP*
Pax20		*paxQ*
Pax29		*paxP*
Pax30		*paxP*

Pax34	GTG CAA AGA AGA AGT TGC AG	*paxQ*
Pax60	[illegible]	*paxQ*
PaxP2P1	[illegible]	*paxQ*
PaxP2P4	[illegible]	*paxQ*
PaxP2P5	[illegible]	*paxQ*
pGEX 3'	[illegible]	pGEX-6P-3
pGEX 5'	[illegible]	pGEX-6P-3
Random hexamer	NNNN NNN	Random
RL10	GGT ACC [illegible]	*paxQ*
RL11	GGT [illegible]	*paxP*
RL12	[illegible]	*paxP*
RL13	[illegible]	*paxP*
RL16	[illegible]	*paxQ*
SP6	[illegible]	pUC vector
T7 promoter	[illegible]	pUC vector
Trx forward	[illegible]	pThioHis-A
Trx reverse	[illegible]	pThioHis-A

(1) Bold type = mismatch to wild-type sequence

2.3.2 Standard PCR reaction components

PCR reactions (25 µL) were incubated in 0.2 mL PCR tubes using a Corbett PC-960, PC-960G or FTS-960 thermocycler. Reactions contained: 2.5 µL 10 x PCR buffer (Section 2.3.1.3), 1.0 µL dNTP solution (Section 2.3.1.2), 0.5 µL of each primer (Section 2.3.1.4), 0.1 µL DNA polymerase (Section 2.3.1.1), 15.4 µL sterile H_2O and 5 µL DNA template.

Expand High fidelity PCR system DNA polymerase (Roche) was used at 0.2 µL per reaction and Stratagene *Pfu* Turbo DNA polymerase (Stratagene) was used at 0.5 µL per reaction. In these cases the volume of H_2O was adjusted to give a 25 µL final volume.

Thermal cycler settings for all PCR based figures are listed below. The initial denaturing period before amplification was 94°C for 2 minutes.

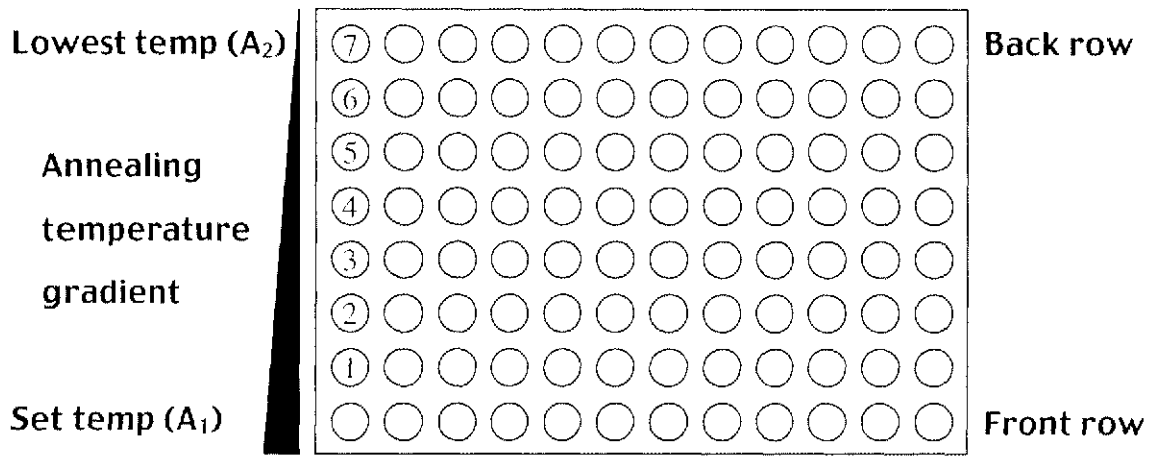
Table 2.3 Thermal cycler settings

Figure	Denaturing	Annealing	Extension	Cycles
3.2	94°C, 30 s	50°C, 30 s	72°C, 2 min	35
3.4 *paxP*	94°C, 30 s	63°C-54°C, 30 s	72°C, 2 min	30
3.4 *paxQ*	94°C, 30 s	70°C-59°C, 30 s	72°C, 2 min 35 s	30
3.5	94°C, 20 s	45°C, 30 s	72°C, 2 min 30 s	35
3.6	94°C, 20 s	45°C, 30 s	72°C, 2 min 30 s	35
3.7	94°C, 30 s	55°C, 30 s	72°C, 2 min 30 s	35
3.8	94°C, 30 s	55°C, 30 s	72°C, 2 min 30 s	25-34
3.9	94°C, 30 s	55°C, 30 s	72°C, 2 min 30 s	33

2.3.3 Gradient PCR thermal cycler settings

The Corbett PC-960G thermal cycler was used to perform gradient PCR. Gradient PCR generates a temperature gradient across the heating block during the annealing step (Figure 2.1). The thermal cycler reports both the maximum and minimum annealing temperatures, which are manually noted when the program is running. The lowest annealing temperature will vary according to the ambient temperature at the time, but is usually 7-10°C lower then the set annealing temperature. The exact annealing temperature for a PCR reaction can be calculated using the formula listed in Figure 2.1.

Fig 2.1 The Gradient PCR heating block



$$A_1 - \left[N\left(\frac{A_1 - A_2}{7}\right)\right] = \text{Annealing temperature for that tube}$$

A_1 = Set annealing temperature

A_2 = Lowest annealing temperature

N = Number of divisions from the front row (1-7)

2.3.4 PCR direct from colony

PCR from a colony was performed as standard PCR with the following modifications. The total reaction volume was reduced to 20 µL including; 2.0 µL 10 x PCR buffer (Section 2.3.1.3), 0.8 µL dNTP solution (Section 2.3.1.2), 0.4 µL of each primer (Section 2.3.1.4), 0.08 µL DNA polymerase (Section 2.3.1.1) and 16.32 µL sterile H_2O. Template was added as *E. coli* cells picked from a colony on a LB agar (Section 2.1.1) plate. Cells to be amplified as a template were touched lightly with a 20 µL pipette tip and the tip briefly immersed in the PCR reaction. Thermal cycler settings were modified with a three minute initial denaturation step at 94°C, followed by standard denaturation, annealing and extension cycles.

2.3.5 Megaprimer PCR

Megaprimer PCR (Aiyar and Leis, 1993) was performed with the same reaction and thermal cycling conditions as standard PCR. The template consisted of a 1:1 molar ratio of the two target regions to be combined in the final reaction. The molar ratio of each segment was calculated on the basis of mass added versus the size of the segment. An overview of the megaprimer theory can be seen in Figure 3.3.

2.3.6 Reverse Transcriptase PCR

Reverse transcriptase (RT) PCR involves synthesis of cDNA from an RNA template and the subsequent amplification of the desired region from that cDNA template. All solutions and equipment used during cDNA synthesis were free of RNase.

2.3.6.1 DEPC treated H_2O

DEPC treated H_2O contained 0.6 mM diethylpyrocarbonate (DEPC). The solution was incubated at 37°C overnight and autoclaved twice at 121°C, 15 psi for 15 minutes.

2.3.6.2 Total cDNA synthesis

The RNA template used for cDNA synthesis was wild-type *Pennicillium paxilli* total RNA isolated from a 60 hour culture (Telfer, 2000). RNA (1 µg) was added to 0.6 µL random hexamer primers (3 µg/µL) (Gibco BRL), and made up to a total volume of 11.7 µL with DEPC-treated H_2O (Section 2.3.6.1). The RNA and primers were incubated at 90°C for 5 minutes and placed on ice. A cocktail consisting of 4 µL of 5 x Expand RT buffer (Roche), 2 µL of 100 mM dithiothreitol (DTT), 0.8 µL of dNTP solution (Section 2.3.1.2) and 20 U of RNase inhibitor (Roche) was mixed with the RNA and primer solution and Expand Reverse transcriptase (50 U) (Roche) added. The reaction was incubated at room temperature for 10 minutes and raised to 42°C in a Corbett thermal cycler for 45 minutes. The resulting cDNA stock was stored at -20°C. Control reactions containing either DEPC-treated H_2O (Section 2.3.6.1) instead of RNA or DEPC-treated H_2O (Section 2.3.6.1) instead of reverse transcriptase enzyme were also performed.

2.3.6.3 Gene specific or mRNA specific cDNA synthesis

Variations on the standard cDNA synthesis protocol were performed in order to generate *paxP*, *paxQ* or mRNA specific cDNA. To synthesise *paxP* or *paxQ* specific cDNA, 2.5 µL of primers P1RL5 or P2RL6 respectively (10 pmol/µL) were added instead of random hexamer primers. To synthesise mRNA specific cDNA, 3 µL of oligo dT primer (10 pmol/µL) was used in place of random hexamer primers.

2.3.6.4 cDNA amplification

cDNA was amplified, using standard PCR conditions, from 10^{-1}, 10^{-2}, 10^{-3} dilutions of the stock solution, to determine the optimal concentration of template for *paxP* and *paxQ* amplification. Control cDNA reactions which were synthesised without reverse transcriptase or without RNA were also amplified undiluted to check for DNA contamination. All cDNA amplifications were performed with separate control reactions of genomic DNA (gDNA).

2.4 Restriction enzyme digestion of DNA

Restriction enzyme digestion was performed in a 25-75 µL total reaction volume. Reactions contained; DNA to be digested, 2-5 units of restriction enzyme, appropriate commercial enzyme buffer and sterile H_2O to the desired total volume. Reactions were incubated at 37°C for 1 hour or until digestion was complete as determined by agarose gel electrophoresis (Section 2.9). Reactions were halted by incubation at 65°C for 10 minutes or by the addition of one quarter volume of sodium dodecyl sulphate (SDS) electrophoresis loading dye (Section 2.9.1.2).

2.5 DNA ligation

A DNA ligation reaction (19 μL) contained 2 μL of 10 x ligation buffer (New England Biolabs), insert DNA, 20 ng of vector and sterile H_2O. DNA ligase (40 Units) (New England Biolabs) in 1 mL was added to the reaction, and incubated at 4°C overnight. Ligation of blunt-ended DNA was performed as above but with the addition of 5% (w/v) polyethelene glycol 6000. The amount of insert DNA added was determined by the desired molar ratio of insert to vector DNA. The amount of insert required for a 1:1 insert:vector ratio was determined by the following equation.

$$\frac{\text{ng of vector x kb of insert}}{\text{kb of vector}} = \text{ng of insert}$$

2.6 Bacterial transformation

E. coli competent cells (Section 2.6.2) were thawed on ice before use. The Gene Pulsar and Pulse controller apparatus (Biorad) were set to deliver 25 μF at 2.5 kV and 200 Ω. DNA (1-2 μL) was added to 40 μL of competent cells, mixed and incubated on ice for one minute. The cell suspension was added to a cold 0.2 cm electroporation cuvette and electroporated once at the above settings. Immediately after electroporation 500 μL of SOC medium (Section 2.1.2) was rapidly added to the cells and incubated at 37°C for 30-60 minutes. After incubation the culture was plated on LB agar (Section 2.1.1) containing antibiotic selection (Section 2.1.3) and incubated at 37°C overnight. Blue white selection was used if supported by the transformed plasmid (Section 2.6.1)

Co-transformation of two different plasmids was performed as for single transformations but with the following modifications. A 1μL aliquot of DNA solution was used for each of the two plasmids. The two 1 μL aliquots of plasmid DNA were added to the same 40 μL of competent cells, mixed and incubated on ice for one minute. The plasmid DNA/competent cell solution was then electroporated and plated as for a single plasmid transformation except for the antibiotic selection utilised. Two different antibiotics which individually selected for the two different plasmids were used to ensure correct plasmid maintenance. Co-transformed plasmids always contained compatible origins of replication to ensure the stable maintennance of both vectors.

2.6.1 Blue white selection

40 µL of X-Gal stock solution (Section 2.1.3.7) and 40 µL of IPTG stock solution (Section 2.1.3.3) were added to LB agar (Section 2.1.1) plates.

2.6.2 Competent cells

Two cultures of *E. coli* were grown overnight in 5 mL of LB medium (Section 2.1.1) at 37°C with antibiotic selection (Section 2.1.3). The two overnight cultures were each added to a 2 litre flask containing 1 litre of LB broth (Section 2.1.1) and antibiotic selection (Section 2.1.3). The culture was incubated at 37°C (300 rpm) for 3 hours until an A_{600} culture density of 0.5-0.6 absorbance units was reached. The culture was chilled on ice for 20 minutes and centrifuged at 4000 x **g** for 10 minutes to harvest the cells. The supernatant was discarded and the cells resuspended in 1 litre of H_2O (4°C). The cells were washed three more times, and were resuspened in 500 mL H_2O (4°C), 20 mL 10% (v/v) glycerol (4°C) and 4 mL 10% (v/v) glycerol (4°C), respectively. The cells were then divided into 40 µL aliquots and stored at -80°C.

2.7 DNA purification

2.7.1 Gel purification

Gel purification of DNA was performed as follows (Thuring *et al.*, 1975). DNA was loaded on a 0.7-1.0% Seaplaque (FMC) low melting point agarose gel in TAE buffer (Section 2.9.1.3). Electrophoresis was performed at a constant 70-80 V. At the completion of electrophoresis the gel was stained in ethidium bromide (Section 2.9.1.1) for 10 minutes and destained in H_2O for 5 minutes. DNA was visualised with longwave UV light and the appropriate band was excised from the gel. The gel slice was weighed and the DNA recovered by one of two methods. Either, using a QIAquick Gel Extraction kit (Qiagen) or by the following method. The gel slice was melted at 65°C and an equal volume of phenol added and vortexed. The mixed solution was placed at -20°C for at least 2 hours. After freezing, the solution was centrifuged for 10 minutes at 13000 rpm in a benchtop centrifuge. The aqueous phase was recovered to a new tube and an equal volume of chloroform added, vortexed and centrifuged for 3 minutes at 13000 rpm. The aqueous phase was recovered to a new tube. One tenth volume of 3 M

sodium acetate and 0.6 volume of isopropanol were added to the aqueous phase, mixed gently and cooled to -20°C for at least 2 hours. The solution was centrifuged for 15 minutes at 4°C to pellet the DNA. The supernatant was removed and an equal volume of 70% (v/v) ethanol added and centrifuged for 2 minutes at 4°C. The supernatant was removed and the DNA pellet dried at 37°C for 15-30 minutes. Once dry, the pellet was resuspended in sterile H_2O or TE buffer (Section 2.2.1.3).

2.7.2 PCR product purification

PCR products were purified using the Concert Rapid PCR Purification System (Invitrogen) to remove PCR reaction components. Plasmid digestions were also treated with this kit to remove small (10-20 bp) DNA fragments and restriction enzyme from the digest.

2.7.3 Phenol/chloroform purification of DNA

An equal volume of phenol and chloroform were added to the DNA to be purified, mixed by vortexing and centrifuged for 3 minutes at 13,000 rpm in a benchtop centrifuge. The aqueous phase was removed to a fresh tube, an equal volume of chloroform added, mixed by vortexing and centrifuged for 1 minute at 13,000 rpm in a benchtop centrifuge. The aqueous phase was removed to a fresh tube, 3 M sodium acetate (0.1 volume) and isopropanol (0.6 volume) were added and mixed gently. The solution was cooled to -20°C for at least 2 hours and centrifuged at 13,000 rpm for 15 minutes (4°C) to pellet the DNA. The supernatant was removed, an equal volume of 70% (v/v) ethanol was added to the DNA pellet and centrifuged for 3 minutes at 13,000 rpm. The supernatant was discarded and the pellet dried at 37°C for 15-30 minutes. The DNA was resuspended in sterile H_2O or TE buffer (Section 2.2.1.3).

2.7.4 Plasmid DNA isolation

E. coli transformed with the plasmid of interest was grown in 5 mL of LB medium (Section 2.1.1) overnight. A 1-5 mL aliquot of culture was processed with the Quantum Plasmid Miniprep kit (Bio-Rad). Isolated plasmid DNA was eluted in sterile H_2O or TE buffer (Section 2.2.1.3).

2.8 DNA quantitation

Double-stranded DNA was quantified with either Hoechst dye (Section 2.8.1) or ethidium bromide (Section 2.8.2) staining and fluorescence.

2.8.1 Flurometric quantitation with hoechst dye

2.8.1.1 Hoechst dye stock

Hoechst dye stock contained; 1.88 mM Hoechst 33258 dye (Amersham Biosciences).

2.8.1.2 10 x TNE buffer

10 x TNE buffer contained; 0.1 M Tris-HCl (pH 7.4), 1 M NaCl, 10 mM Na_2EDTA.

2.8.1.3 Working solution A

Working solution A contained; 1 x TNE buffer (Section 2.8.1.2), 10 µL Hoechst dye stock (Section 2.8.1.1)

2.8.1.4 Flurometric quantitation

DNA was analysed using either a Hoefer TKO-100 or a Hoefer DyNA Quant 200 fluorometer (Amersham Biotech). The fluorometer was calibrated to read zero and 100 ng/µL of calf thymus DNA standard (Amersham Biotech) in 2 mL of Working solution A (Section 2.8.1.3). A 2 µL sample of unknown concentration was then tested.

2.8.2 Quantitation by ethidium bromide staining

Quantitation of DNA by ethidium bromide staining was performed after agarose electrophoresis (Section 2.9) of samples of unknown concentration along with standards of known concentration. Standards used were Low DNA Mass Ladder (Gibco) and digested pUC118 corresponding to 2.5 ng/µL, 5 ng/µL and 10 ng/µL concentrations. The agarose gel was stained in ethidium bromide (Section 2.9.1.1) for 10 minutes and destained for five minutes in water. Ethidium bromide fluorescence was visualised by exposure to shortwave UV light and photographed. An approximation of the unknown sample concentration was made on the basis of its fluorescence compared to that of the standards.

2.9 Agarose gel electrophoresis

2.9.1 Agarose gel electrophoresis solutions

2.9.1.1 Ethidium bromide staining solution

Ethidium bromide staining solution contained 5 µM ethidium bromide.

2.9.1.2 SDS loading dye

SDS loading dye contained: 0.58 M sucrose, 5 mM Na_2EDTA, 35 mM SDS, 3 mM bromophenol blue.

2.9.1.3 1 x TAE electrophoresis buffer

1 x TAE electrophoresis buffer contained: 20 mM Tris, 10 mM glacial acetic acid, 1 mM Na_2EDTA (pH 8.5).

2.9.1.4 1 x TBE electrophoresis buffer

1 x TBE electrophoresis buffer contained: 89 mM Tris, 89 mM boric acid, 2 mM Na_2EDTA (pH 8.2).

2.9.2 Agarose gel electrophoresis method

Agarose gels were made with either agarose (Gibco) dissolved in TBE buffer (Section 2.9.1.4) for routine electrophoresis, or SeaPlaque agarose (FMC) dissolved in 1 x TAE buffer (Section 2.9.1.3) for gel purification electrophoresis. Agarose gel concentrations ranged from 0.7% (w/v) to 2% (w/v) depending on the length of the DNA to be resolved. DNA (5 µL) was mixed with 5 µL of SDS loading dye (Section 2.9.1.2) and loaded into the gel. Electrophoresis was performed at a constant 70-100 V until the dye front had progressed at least 2/3 of the gel length.

Resolved DNA fragments were visualised by soaking the gel in ethidium bromide staining solution (Section 2.9.1.1) for 10 minutes, destaining in H_2O for five minutes and exposed to shortwave UV light. Images were recorded by an Alpha Innotech video capture system. Sizes of DNA fragments were estimated by comparing their migration to that of DNA size markers electrophoresed in parallel.

2.10 Automated sequencing of DNA

DNA templates were sequenced using the MuSeq Massey University DNA Analysis Service. The Sanger dideoxy chain termination method for sequencing reactions (Sanger *et al.*, 1977) was carried out using ABI Prism Big Dye™ chemistry (Applied Biosystems) and analysed on an ABI377 DNA sequencer (Applied Biosystems). Sequence data was viewed using EditView™ 1.01 (Applied Biosystems) and Sequencher™ 4.1 (Gene Codes Corporation) software.

2.11 Recombinant protein expression and analysis

2.11.1 Recombinant protein expression

Recombinant proteins were expressed in *E. coli* by the following method. LB broth (Section 2.1.1) (5 mL) with antibiotic selection (Section 2.1.3) was inoculated with either BL21 or Origami B *E. coli* (Section 2.1.4) transformed with the expression vector of interest. The culture was incubated at 37°C overnight. A 150 µL aliquot of the overnight culture was sub-cultured into 5 mL of LB broth with antibiotic selection (Section 2.1.3) and incubated at 37°C until an A_{600} culture density of 0.5-0.7 absorbance units was reached. The culture was then cooled on ice to ensure the temperature of the culture was not above the desired induction temperature. Recombinant protein expression was induced by the addition of IPTG (Section 2.1.3.3) to a final concentration of 0.1–1 mM depending on the strain and vector being used. The optimal concentration of IPTG was determined by titrating IPTG concentrations until the minimum amount necessary for detectable expression was found. Cultures were incubated at 15°C, 18°C, 22°C, 25°C or 37°C. Recombinant protein expression was induced for either 3 hours at 37°C, overnight at 22°C/25°C, 1 to 2 days at 15°/18°C or alternatively until an A_{600} culture density of 1.4-1.7 absorbance units was reached.

2.11.2 Sample preparation for total protein analysis

Cells induced for recombinant protein expression were harvested by centrifugation of 1 mL of culture at 13,000 rpm in a benchtop centrifuge for 30 seconds, washed in 1 mL PBS buffer (Section 2.2.1.1) and resuspended in 1 mL PBS buffer. For SDS/PAGE

(Section 2.13) analysis of total cellular protein, 15 µL of resuspended cells was added to 15 µL of 2 x sample treatment buffer (Section 2.13.1.5), boiled for 3 minutes and centrifuged at 13,000 rpm in a benchtop centrifuge for 2 minutes. A 10 µL aliquot was then analysed by SDS/PAGE (Section 2.13).

2.11.3 Sample preparation for protein solubility analysis

Cells induced for recombinant protein expression were harvested by centrifugation of 500 µL of culture at 13,000 rpm in a benchtop centrifuge for 30 seconds, washed in 1 mL PBS (Section 2.2.1.1) and resuspended in 200 µL PBS. The cells were cooled on ice and lysed by sonication with a VirSonic sonicator (Virtis) fitted with a microtip probe. The cell suspension was held on ice to prevent heat build-up. Sonication was performed on setting 3 (1-10 scale) for 10 second periods until partial clearing of the cell suspension was evident. Lysed cells were separated into soluble and insoluble fractions by centrifugation at 13,000 rpm for 15 minutes at 4°C. The supernatant was reserved as the soluble fraction and the pellet resuspended in 200 µL PBS as the insoluble fraction. Samples including total cellular, soluble and insoluble fractions were prepared for SDS/PAGE (Section 2.13).

2.11.4 Solubilisation of proteins with *n*-laurylsarkosine

Harvested cells were washed and resuspended as described for protein solubility analysis (Section 2.11.3) but with the following alterations. Before sonication *N*-laurylsarkosine was added to the protein sample to a final concentration of 0.25% (w/v) (8.5 mM). STE buffer (Section 2.2.1.2) was used in place of PBS buffer (Section 2.2.1.1) in all *N*-laurylsarkosine containing solutions.

2.12 Affinity chromatography

2.12.1 Affinity chromatography solutions

2.12.1.1 Glutathione elution buffer

Glutathione elution buffer contained; 10 mM reduced glutathione, 50 mM Tris-HCl (pH 8.0)

2.12.1.2 Haemin solution

Haemin solution contained; 10 μM haemin chloride in PBS buffer (Section 2.2.1.1) (pH 7.0). The solution was made by diluting a 8.8 mM solution of haemin chloride in 0.1 M NaOH, with PBS (pH 7.0), to the stock concentration of 10 μM.

2.12.1.3 PreScission protease cleavage buffer

PreScission protease cleavage buffer contained; 50 mM Tris-HCl (pH 7.0), 150 mM NaCl, 1 mM Na_2EDTA, 1 mM DTT.

2.12.2 Binding of GST fusion proteins to affinity beads

All procedures for affinity purification of GST fusion proteins were performed on ice using solutions chilled to 4°C. Soluble protein fractions (200 μL) were incubated with 10-30 μL of a 50% (w/v) Glutathione-Sepharose 4B affinity bead slurry (Amersham Biosciences) at 4°C on a slow rotary shaker. Incubations varied from 30 minutes to overnight, with the longer period resulting in greater yields of bound GST fusion protein. If the soluble protein fraction had been treated with *N*-laurylsarkosine, Triton X-100 was added to a final concentration of 4% (v/v) before the addition of affinity beads. After the binding period, the beads were separated into supernatant (unbound) and pellet (bound) fractions by centrifugation at 500 x **g** for five minutes (4°C). The pelleted affinity beads were washed three times by the addition of either 1 mL of PBS (Section 2.2.1.1) or STE buffer (Section 2.2.1.2) (for all solutions containing *N*-laurylsarkosine), and centrifuged again. The washed beads were resuspended in 200 μL PBS or STE buffer and unbound and bound samples prepared for SDS/PAGE (Section 2.13). Bound fractions containing affinity beads were treated the same as a soluble protein fraction.

Affinity chromatography in the presence of haemin was performed as above but with 1-50 μL of a 10 μM haemin solution (Section 2.12.1.2) added in addition to the affinity bead slurry.

2.12.3 Elution of GST fusion proteins from affinity beads

Affinity bead slurry with bound GST fusion protein was applied to a column and 1 mL of glutathione elution buffer (Section 2.12.1.1) added per mL of bed volume. The beads

were then incubated at room temperature for 10 minutes to elute the fusion protein. The eluate was collected and the elution step repeated twice, the three eluates were pooled or analysed separately.

2.12.4 PreScission protease cleavage of GST fusion proteins

Affinity bead slurry with bound GST fusion protein was washed with 10 bed volumes of PreScission cleavage buffer (Section 2.12.1.3) prior to digestion. The affinity beads were resuspended in 1 mL of cleavage buffer per mL of bed volume and 2 units of PreScission protease (Amersham Biotech) added per 100 µg of bound protein. The digestion was incubated at 4°C overnight with gentle agitation. Once the digest was complete the bead suspension was pelleted by centrifugation at 500 x **g** for 5 minutes (4°C), the protein of interest was recovered in the supernatant fraction.

2.13 SDS Polyacrylamide gel electrophoresis (SDS/PAGE)

2.13.1 SDS/PAGE solutions

2.13.1.1 Acrylamide bisacrylamide mix

Acrylamide bisacrylamide mix contained; 4 M acrylamide (BDH), 65 mM N,N'-methylenebisacrylamide (Serva).

2.13.1.2 Coomassie brilliant blue staining solution

Coomassie brilliant blue staining solution contained; 0.3 mM Coomassie brilliant blue R-250, 40% (v/v) methanol, 7% (v/v) glacial acetic acid.

2.13.1.3 Coomassie brilliant blue destaining solution

Coomassie brilliant blue destaining solution contained; 40% (v/v) methanol, 7% (v/v) glacial acetic acid.

2.13.1.4 10 x protein electrophoresis buffer

10 x Protein electrophoresis buffer contained; 0.25 M Tris (pH 8.3), 1.92 M glycine, 35 mM SDS.

2.13.1.5 2 x protein sample treatment buffer

Protein sample treatment buffer (2 x) contained; 0.125 M Tris-HCl (pH 6.8), 0.14 M SDS, 30% (v/v) glycerol, 2% (v/v) 2-mercaptoethanol, 0.29 mM bromophenol blue.

2.13.1.6 6 x protein sample treatment buffer

Protein sample treatment buffer (6 x) contained; 0.35 M Tris-HCl (pH 6.8), 0.35 M SDS, 30% (v/v) glycerol, 0.6 M DTT, 0.175 mM bromophenol blue.

2.13.2 Polyacrylamide gel components

Each polyacrylamide gel consists of a resolving section and a stacking section. The resolving section consisted of (10% (w/v) acrylamide gel, 1x); 4 mL of H_20, 3.3 mL of acrylamide-bisacrylamide mix (Section 2.13.1.1) , 2.5 mL of 1.5 M Tris-HCl (pH 8.8), 0.1 mL of 0.35 M SDS, 0.1 mL of 0.44 M ammonium persulphate (APS), 25 µL of tetramethylethylenediamine (TEMED). Stacking section consisted of (4% (w/v) acrylamide gel, 1 x); 3.05 mL of H_20, 0.65 mL of acrylamide bisacrylamide mix (Section 2.13.1.1), 1.25 mL of 0.5 M Tris-HCl (pH 6.8), 0.05 mL of 0.35 M SDS, 0.05 mL of 0.44 M APS, 25 µL TEMED.

2.13.3 Gel casting

The resolving section of the gel was cast first, followed by the casting of the stacking section of the gel on top of the resolving section. H_2O, acrylamide-bisacrylamide and Tris-HCl for the resolving gel were mixed and degassed under vacuum for 10 minutes. The 10% (w/v) SDS, 10% (w/v) APS and TEMED for the resolving gel are added and the solution poured into gel casting mould. The resolving gel should leave enough space so that the distance from the base of the comb to the top of the resolving gel is equal to the depth of the sample once it is loaded. Butanol was quickly layered over the poured resolving gel before it was allowed to set. Once the gel was set, the butanol was washed off with H_2O and left to drain. The stacking section was then prepared in the

same manner to the resolving section and poured over the set resolving gel until the mould was completely full. A comb was inserted before the stacking gel set.

2.13.4 Sample preparation for electrophoresis

The protein sample (15 µL) was added to 15 µL of 2 x sample treatment buffer (Section 2.13.1.5) and boiled for three minutes. The solution was centrifuged for three minutes at 13000 rpm in a benchtop centrifuge and a 10 µL aliquot analysed by SDS/PAGE (Section 2.13).

2.13.5 SDS Polyacrylamide gel electrophoresis

A polyacrylamide gel was loaded into a Hoefer mighty small SE250 PAGE apparatus (Amersham Biotech). Protein electrophoresis buffer (1 x) (Section 2.13.1.4) was poured into the reservoirs, samples were loaded and the electrophoresis performed at a constant 32 mA. The electrophoresis was halted once the dye front had reached the end of the gel.

2.13.6 Coomassie brilliant blue staining

Completed gels were soaked in coomassie brilliant blue staining solution (Section 2.13.1.2) for at least 30 minutes, longer staining periods were used to increase the sensitivity of the stain. Gels were stained, washed briefly in H_2O and added to coomassie brilliant blue destaining solution (Section 2.13.1.3). The destaining solution was replaced as it became saturated with dye. Destaining was halted once the required gel clarity was achieved. Gels were rehydrated to their original size by soaking in H_2O and photographed using an Alpha Innotech video capture system.

2.14 Bioinformatics

2.14.1 Hydropathy analysis

Hydropathy values for polypeptide sequences were calculated with the methods of Kyte-Doolittle (Kyte and Doolittle, 1982) and Goldman (Engelman *et al.*, 1986) using the PEPPLOT program (Gribskov *et al.*, 1986) from the University of Wisconsin Genetics Computer Group (GCG) Bioinformatics package version 9.1.

2.14.2 Multiple sequence alignment

Multiple sequence alignments were generated using the University of Wisconsin GCG PILEUP program. This program calculates pairwise alignments of each sequence and then combines the paired sequences into a final alignment using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970). Alignments were exported from the GCG package as MSF files and shaded using the MACBOXSHADE 2.15 program.

Chapter 3 Results

3.1 Construction of GST-PaxP and GST-PaxQ expression vectors

3.1.1 Initial PCR amplification of *paxP* and *paxQ* cDNA

The strategy initially adopted to overexpress *paxP* and *paxQ* in *E. coli* was to synthesise cDNA copies (Section 2.3.6) of these genes and clone them into the Glutathione-S-transferase vector. The general strategy to achieve this is shown in Figure 3.1. A cDNA template was synthesised using either gene specific primers, P1RL5 for *paxP* and P2RL6 for *paxQ* or random hexamer primers (Table 2.2) annealing to total RNA (Section 2.3.6). While full length and partial genomic copies of *paxP* and *paxQ* were readily amplified (Figure 3.2; lanes 4, 8 and 11), only a partial length copy of *paxQ* could be amplified from cDNA template (lane 10). Additional attempts to amplify full-length cDNAs were unsuccessful. PCR amplifications in this study were performed with Roche High Fidelity DNA polymerase unless otherwise noted. This DNA polymerase is supplied as a mixture of both proofreading and non-proofreading enzymes, providing increased fidelity compared to standard *Taq* DNA polymerase

3.1.2 PCR amplification of *paxP* and *paxQ* from megaprimer cDNA template

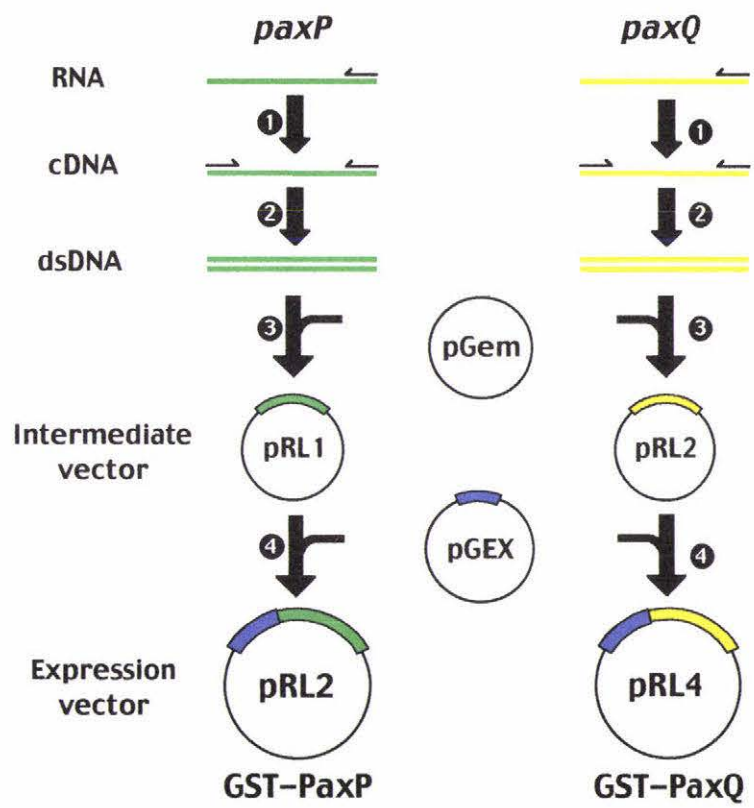
3.1.2.1 The megaprimer method of PCR amplification

Given the lack of success in amplifying full length cDNAs using standard procedures a megaprimer PCR method (Section 2.3.5) was adopted. An overview of the megaprimer method of PCR amplification is shown in Figure 3.3. The first step of megaprimer PCR is the individual synthesis of the two overlapping halves of the target region (products I and II).

Primers are selected to amplify two regions covering the desired target with between 15 bp to 100 bp of overlapping sequence between the two megaprimers (Figure 3.3, step 1). This overlapping sequence allows the megaprimers to anneal to each other during the second round PCR and to be extended to form full-length template

(Figure 3.3, step 3). The second round PCR reaction contains the two megaprimers in a 1:1 molar ratio and standard primers which anneal at the 5' and 3' ends of the full length template. Once the full length target has been synthesised from the megaprimers, amplification of that target will occur.

Fig 3.1 **Strategy to clone a full length cDNA copy of *paxP* and *paxQ*.**



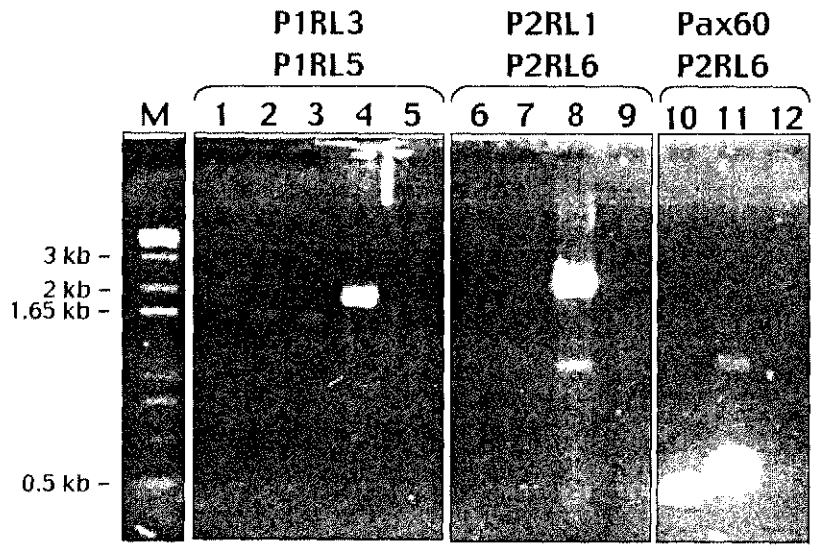
Step 1: cDNA synthesis from total RNA, using reverse transcriptase primed with either P1RL5 (*paxP*), P2RL6 (*paxQ*) or random hexamer primers (total cDNA).

Step 2: PCR amplification from cDNA, using primers P1RL3/P1RL5 (*paxP*) or P2RL1/P2RL6 (*paxQ*) which contain unique restriction endonuclease sites.

Step 3: Clone PCR product into T-tailed vector for ease of cloning and manipulation.

Step 4: Subclone insert into expression vector using *Eco*RI and *Not*I (*paxP*) or *Bam*HI and *Xho*I (*paxQ*) restriction sites.

Fig 3.2 **PCR amplification of *paxP* and *paxQ* from random hexamer and gene specific cDNA**



cDNA templates do not allow full length PCR amplification of *paxP* and paxQ.

PCR of full length *paxP* (primers P1RL3+P1RL5), full length *paxQ* (P2RL1+P2RL6) and partial length *paxQ* (*pax60*+P2RL6) from cDNA and genomic DNA (gDNA) templates. 1% agarose gel.
Lane M, 1kb+ ladder. Templates used per reaction are as follows. Lane 1, 1:2 *paxP* gene specific cDNA. Lane 2, 1:10 *paxP* gene specific cDNA. Lane 3, cDNA synthesised with no RNA added. Lane 4, gDNA. Lane 5, no template. Lane 6, 1:10 random hexamer cDNA. Lane 7, 1:100 random hexamer cDNA. Lane 8, gDNA. Lane 9, no template. Lane 10, 1:10 random hexamer cDNA, Lane 11, gDNA. Lane 12, no template.

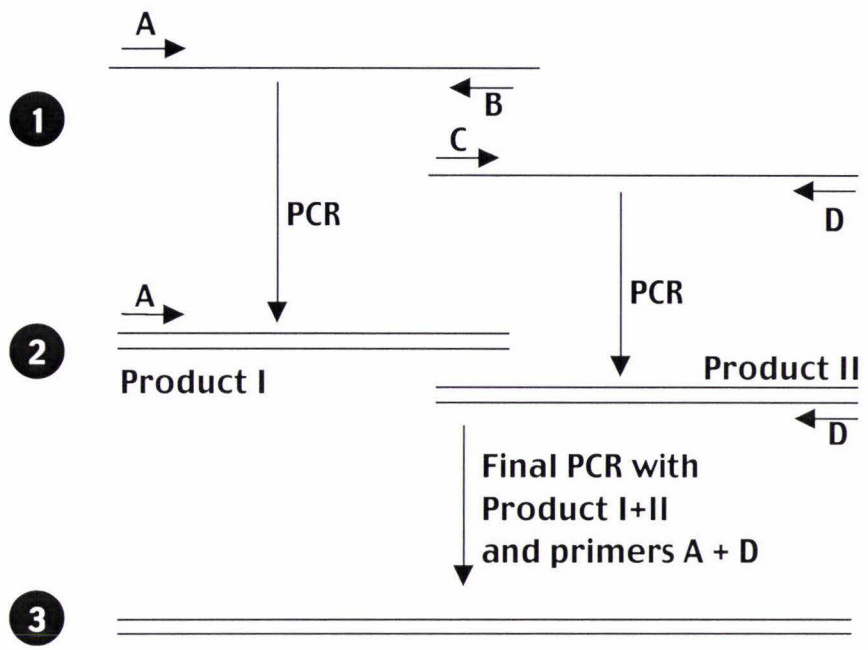
Expected sizes of PCR products are as follows. Lane 1, 1.6 kb. Lane 2, 1.6 kb. Lane 3, no expected product. Lane 4, 1.9 kb. Lane 5, no expected product. Lane 6, 1.6 kb. Lane 7, 1.6 kb. Lane 8, 2.2 kb. Lane 9, no expected product. Lane 10, 0.4 kb. Lane 11, 0.5 kb. Lane 12, no expected product.

See Table 2.3 for incubation conditions.

3.1.2.2 Design of paxP and paxQ megaprimers

The primers chosen to amplify the megaprimers were selected from those already present in stocks. Primers P1RL3 and CYP-3 (Table 2.2) were used to amplify the 5' *paxP* megaprimer while primers CYP-8 and P1RL5 (Table 2.2) were used to amplify the 3' *paxP* megaprimer. This approach was predicted to amplify a 446 bp 5' megaprimer and a 1170 bp 3' megaprimer which share a 17 bp overlapping region.

Fig 3.3 **Overview of the megaprimer method of PCR amplification.**



Megaprimer PCR allows amplification of rare PCR templates via a two step method.

Step 1. Amplify two, partial, overlapping regions of the target region. These smaller regions are more abundant in the cDNA template than a full length region.

Step 2. Combine the two partial regions (megaprimers) as a template for second round PCR.

Step 3. Full length PCR product of target region.

This overlapping region was the minimum size suggested (Aiyar and Leis, 1993) but still fulfilled the requirements for a preliminary test. The primers chosen to amplify the 5' region of *paxQ* were P2RL1 and PaxP2P5, while the primers for amplification of the 3' region was Pax34 and P2RL6 (Table 2.2). These were predicted to amplify a 609 bp 5' region and a 1100 bp 3' region which share an 88 bp overlap. This design is more robust because of a larger overlapping region between the two megaprimers. The primers chosen for the second round megaprimer PCR were P1RL3 and P1RL5 for *paxP* and P2RL1 and P2RL6 for *paxQ* (Table 2.2).

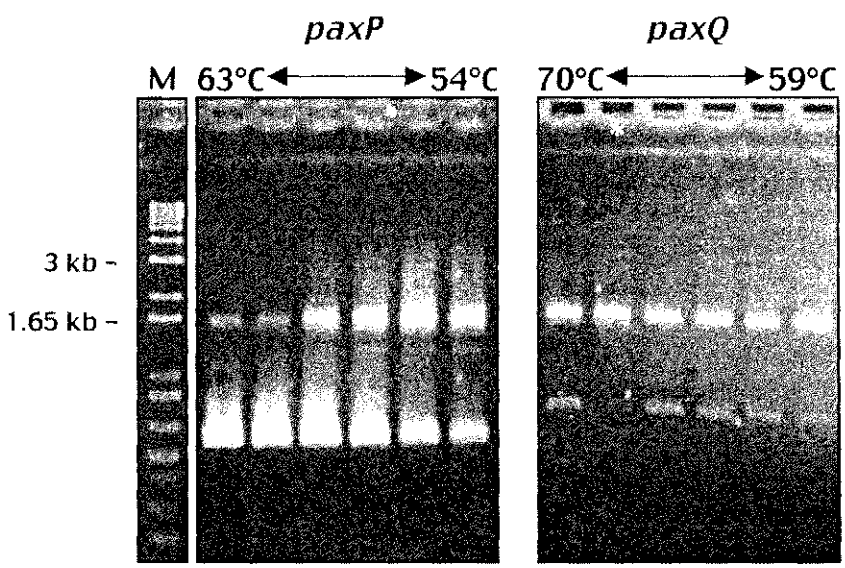
All megaprimer PCR reactions were performed with Roche Expand High Fidelity DNA polymerase. This product contains a mixture of both *Taq* and *Pfu* DNA polymerases. The combination of proofreading (*Pfu*) and non-proofreading (*Taq*) enzymes means that

PCR products are synthesised with and without an additional adenine at the 3' end. Therefore, the use of this DNA polymerase mix ensures PCR products are synthesised with high fidelity and can still be cloned efficiently into T-tailed vectors. The addition of an extra adenine to the 3' region of a PCR product does impact on the second round synthesis of the final megaprimer product. Any products which contain an extra A at the 3' end will not be corrected by *Taq* DNA polymerase, resulting in a base change at that position. However, the *Pfu* DNA polymerase will correct any A-tailed products as it begins to extend the annealed megaprimer. It is notable that none of the PCR products synthesised by the megaprimer method were found to contain any base changes within the region of overlap.

3.1.2.3 Optimisation of megaprimer PCR conditions

Following synthesis of the megaprimers, PCR conditions were optimised for the synthesis of full length product. The first step in the optimisation was to vary the concentration of the 1:1 molar mix of megaprimers added to the PCR reaction.

Fig 3.4 **Optimisation of PCR conditions for *paxP* and *paxQ* megaprimer synthesis.**



Two main products were amplified for both *paxP* and *paxQ*, megaprimer PCR.

Megaprimer PCR of *paxP and paxQ* using Roche HiFi DNA polymerase with primers P1RL3/P1RL5 and P2RL1/P2RL6 respectively. 0.7% agarose gel. Annealing temperatures set as a gradient between 63 °C and 54°C for *paxP* and 70°C and 59°C for *paxQ*. Lane M, 1kb+ ladder. Expected size of PCR product is 1.6 kb for all lanes. See Table 2.3 for incubation conditions.

Template was diluted 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4} fold and amplified using an annealing temperature of 57°C with Roche HiFi DNA polymerase (Section 2.3.1.1). The 10^{-3} dilution was found to be optimal for yield of both *paxP* and *paxQ* PCR products (data not shown).

To determine the optimal annealing temperature for the second round megaprimer PCR synthesis of *paxP* and *paxQ*, gradient PCR (Section 2.3.3) was performed. Annealing temperatures from 63°C to 54°C were trialed for amplification of *paxP* (Figure 3.4). At 63°C the dominant product was a 600 bp band while the required product of 1.6 kb was present as a faint band. At 54°C the 1.6 kb and 0.6 kb products were observed in equal intensities. Annealing temperatures of 70°C to 59°C were tested for the amplification of *paxQ*. The dominant product across the entire temperature range was a 1.6 kb product corresponding in size to full-length *paxQ*. A fainter 700 bp product was also observed in all reactions. A temperature of 68°C (marked with an asterisk) was chosen as the optimal annealing temperature for amplification.

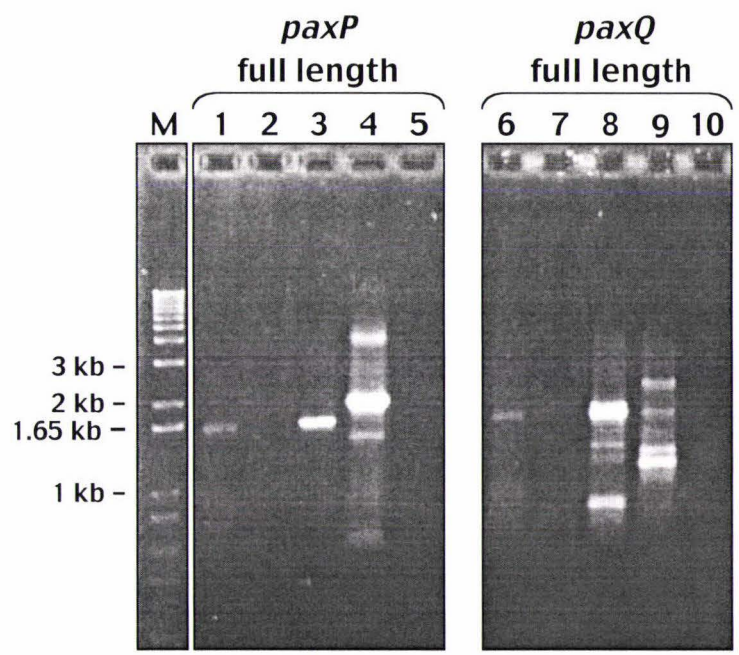
3.1.2.4 A paxQ clone synthesised with megaprimer PCR

The products of the *paxQ* PCR carried out at 68°C (Figure 3.4, marked with an asterisk) were ligated (Section 2.5) into the pGEM-T-Easy vector (Table 2.1) and *E. coli* XL1-blue competent cells transformed (Section 2.6) with the resulting plasmid. Clones were analysed by PCR amplification (Section 2.3) using primer combinations of P2RL1 and P2RL6, P2RL1 and PaxP2P4, and Pax60 and P2RL6 (Table 2.2). Clones producing PCR products of 1.6 kb, 411 bp and 439 bp were assumed to contain a full-length *paxQ* cDNA. One such clone was sequenced (Section 2.10) using primers M13(*lacZ*)Forward, Pax20, Pax34 and M13(*lacZ*)Reverse. Comparison of the cloned cDNA sequence with the known *paxQ* genomic sequence revealed two differences, corresponding to a C to T (Proline to Serine; Appendix 6.2.1) change at position 150 and a T to C (Methionine to Threonine; Appendix 6.2.2) change at position 854 of the appendix map. The positions of both changes are shaded in grey on the sequence map. On the basis of this result the strategy was abandoned. It appeared that two round megaprimer PCR amplification would make it difficult to synthesise an error free cDNA copy of either *paxP* or *paxQ*.

3.1.3 Comparison of three different methods of cDNA synthesis

Since the megaprimer PCR approach resulted in products with misincorporated bases, a decision was made to return to the original strategy (Figure 3.1) but to include cDNA synthesised using an oligo dT primer (Table 2.2). Templates generated from oligo dT, *paxP* and *paxQ* specific and random hexamer primed cDNA synthesis were tested for full length amplifications of *paxP* and *paxQ* (Figure 3.5). New cDNA stocks were prepared for this experiment using the same RNA as used previously (Section 2.3.6).

Fig 3.5 **Comparison of gene specific, random hexamer and oligo dT primers for cDNA synthesis**

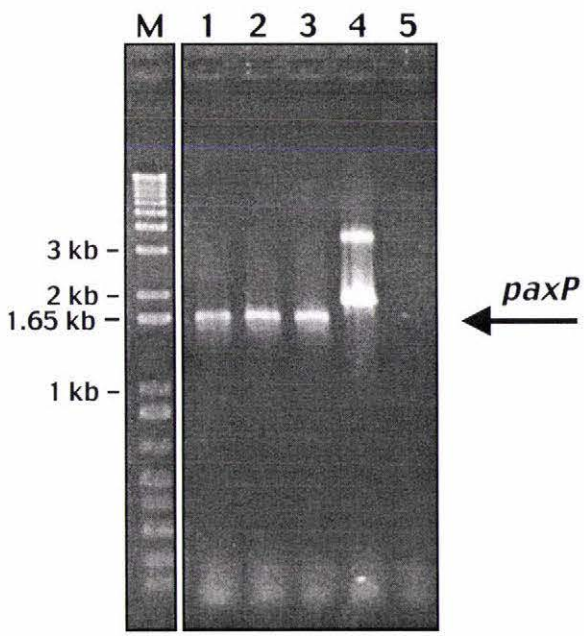


Gene specific cDNA template results in a greater yield of PCR product from full length amplification.

PCR amplification of full length *paxP* (primers P1RL3+P1RL5) and full length *paxQ* (primers P2RL1+P2RL6) from 1:10 diluted cDNA and genomic DNA (gDNA) templates. 0.7% agarose gel.
Lane M, 1kb+ ladder. Lanes 1 and 6, oligo-dT-primed cDNA template. Lanes 2 and 7, random hexamer-primed cDNA template. Lanes 3 and 8, *paxP* and *paxQ* cDNA template, respectively. Lanes 4 and 9, gDNA template. Lanes 5 and 10, no template. Expected sizes of PCR products are as follows: Lanes 1-3 and 6-8, 1.6 kb. Lane 4, 2 kb. Lane 9, 2.2 kb. Lanes 5 and 10, no product. See Table 2.3 for incubation conditions.

In this experiment the best yield of *paxP* and *paxQ* cDNA PCR products was obtained using template synthesised from *paxP* (lane 3) and *paxQ* (lane 8) primed template, while products were obtained for oligo-dT primed template, the yields were low (lanes 1 and 6). In the *paxP* genomic PCR reaction (lane 4) a product of the expected size of 1.9 kb was amplified but in addition three other non-specific products of 4 kb, 1.5 kb and 0.7 kb were also present. This lack of specificity was probably due to a combination of the low (45°C) annealing temperature and the higher concentration of template used. The *paxP* specific cDNA was chosen as template for further optimisation of PCR conditions. The improved yield of *paxP* after amplification was probably due to both the new cDNA preparation being of a higher quality when compared to previous templates and the use of lower stringency PCR conditions.

Fig 3.6 **PCR amplification of *paxP* from gene specific cDNA**



Full length *paxP* amplified from gene specific cDNA.

Full length *paxP* amplified with primers P1RL3+P1RL5 from gene specific cDNA or gDNA. Lanes 1-3 were pooled, the products gel purified and cloned into pGEM-T-easy. 0.7% agarose gel. Lane M, 1kb+ ladder. Lanes 1-3, 1:10 *paxP* specific cDNA template. Lane 4, gDNA template. Lane 5, no template. Expected sizes of PCR products are as follows: Lanes 1-3, 1.6 kb. Lane 4, 2 kb. Lane 5, no product. See Table 2.3 for incubation conditions.

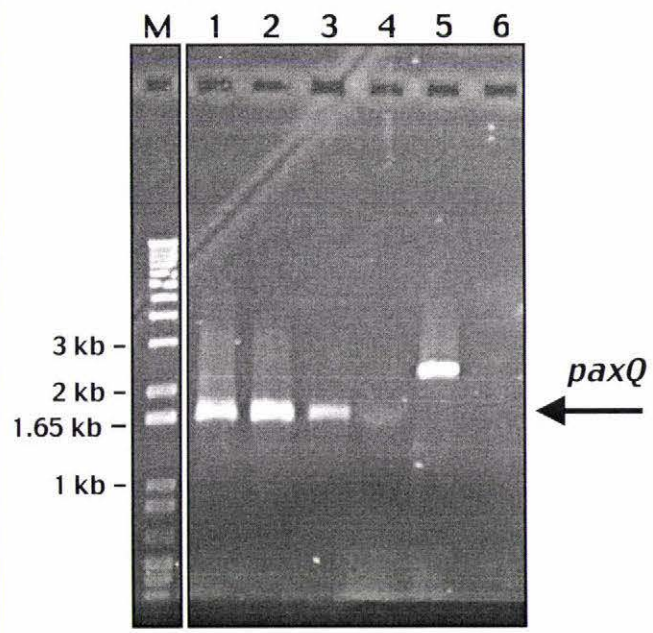
Amplification of *paxQ* showed that while the *paxQ* specific cDNA template produced a good yield of the full-length *paxQ* PCR product a number of non-specific products of 1.6 kb, 1.4 kb and 1.3 kb were also present in the genomic sample in addition to the expected 1.7 kb product. This lack of specificity was presumably the result of the low annealing temperatures used. Random hexamer primed cDNA did not amplify at all. However, under these conditions a relatively

specific product of small yield, was obtained using oligo dT primed template. Oligo-dT primed cDNA conditions were optimised for further experiments considering its specificity when compared to the *paxQ* specific cDNA.

3.1.3.1 Cloning of *paxP* cDNA

A *paxP* specific cDNA template was diluted 1:10 and amplified using primers P1RL3 and P1RL5 and Roche HiFi DNA polymerase (Section 2.3.1.1, Table 2.2). Three 50 µL PCR reactions (Figure 3.6) were pooled and gel purified before ligation into the pGEM-T Easy vector (Table 2.1). The products of the ligation were transformed (Section 2.6) into *E. coli* XL1-blue competent cells. The resulting transformants were screened for a cloned insert 1.6 kb in size using colony PCR (Section 2.3.4) with primers M13-(lacZ)Forward and M13-(lacZ)Reverse (Table 2.2). Three colonies were identified that had an insert of the correct size. These clones were analysed by PCR using primers Pax30 and Pax29 (Table 2.2). The *paxP* specific PCR indicated that all three clones contained a *paxP* insert. A clone was selected for sequencing (Section 2.10) with primers T7, CYP-8, Pax29 and SP6. The sequence data showed the clone did not contain any base changes when compared to genomic *paxP* sequence. This *paxP*-pGEM-T Easy clone was named pRL1.

Fig 3.7 **PCR amplification of *paxQ* from oligo dT primed cDNA**



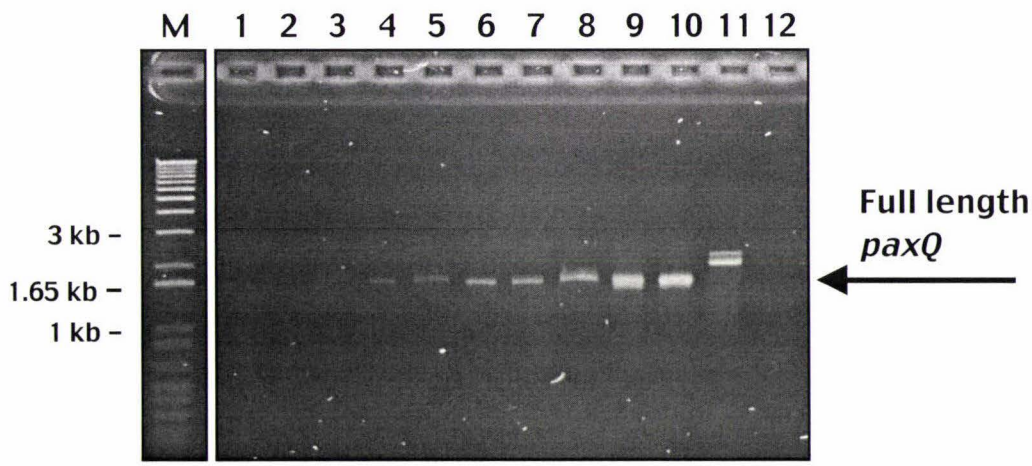
Full length *paxQ* amplified from oligo dT primed cDNA.

Full length *paxQ* amplified with primers P2RL1+P2RL6 from oligo dT primed cDNA or genomic DNA (gDNA) templates. 0.7% agarose gel Lane M, 1kb+ ladder. Lanes 1-3 represent 1:10, 1:100 and 1:1000 dilutions of cDNA template respectively. Lane 4, No RNA during cDNA synthesis template. Lane 5, gDNA. Lane 6, no template. Expected sizes of PCR products are as follows: Lanes 1-3, 1.6 kb. Lane 4, 2.2 kb. Lane 5, no product. See Table 2.3 for incubation conditions.

The *paxP* insert was digested (Section 2.4) from pRL1 with restriction enzymes *Eco*RI and *Not*I, unique sites introduced with the primers P1RL3 and

P1RL5 respectively. Because there are *Eco*RI and *Not*I restriction sites present in the vector MCS as well as in the primers used to amplify the insert, a single enzyme digestion would remove the insert from the vector. Single enzyme digests were therefore used to ensure each digest was complete. The digestion products were analysed by agarose gel electrophoresis (Section 2.9) to check each digest (data not shown). Once each digest was confirmed complete, *Eco*RI was added to the *Not*I digest, and *Not*I added to the *Eco*RI digest to complete the digestion of the *paxP* insert. The *Eco*RI/*Not*I *paxP* insert was gel purified (Section 2.7.1) and ligated (Section 2.5) into pGEX-6P-3 digested with *Eco*RI and *Not*I. The *paxP*-pGEX-6P-3 ligation was then transformed (Section 2.6) into *E. coli* XL1-blue competent cells. Colony PCR (Section 2.3.4) analysis of resulting transformants with primers pGEX 5' and pGEX 3' (Table 2.2) identified clones that contained an insert 1.6 kb in size. One clone was selected and the insert DNA sequenced (Section 2.10) with primers pGEX 5', CYP-8, Pax29 and pGEX 3' to check the accuracy of the sequence. The sequence data showed the clone was ligated in the correct orientation and did not contain any base alterations when compared to *paxP* genomic sequence. This *paxP*-pGEX-6P-3 clone was named pRL2.

Fig 3.8 **PCR cycle optimisation for *paxQ* amplification**



The optimal number of cycles for PCR amplification of *paxQ* from oligo dT primed cDNA is 33 cycles.

Oligo dT primed cDNA and genomic DNA amplified with primers P2RL1 and P2RL6. 0.7% agarose gel. Lane M, 1kb+ ladder. Lanes 1-10, 1:100 oligo dT cDNA amplified for 25-34 cycles respectively. Lane 11, genomic DNA amplified for 34 cycles. Lane 12, control no template amplified for 34 extension cycles. Expected sizes of PCR products are as follows: Lanes 1-10, 1.6 kb. Lane 11, 2.2 kb. Lane 12, no product. See Table 2.3 for incubation conditions.

3.1.3.2 Cloning of paxQ cDNA

Primers P2RL1 and P2RL6 were used to test newly prepared oligo dT primed cDNA template at three different concentrations, 10^{-1}, 10^{-2} and 10^{-3} dilutions of the stock. The new oligo dT-primed cDNA preparation proved to be of high quality, with amplification achieved in all three template dilutions tested (Figure 3.7). To minimise the misincorporation of nucleotides during PCR the yield of *paxQ* from the 10^{-2} dilution of oligo dT cDNA was monitored between cycles 25 and 34 of the PCR. Ten PCR reactions were set up and run, with one tube removed after the extension phase for each representative cycle between 25 and 34 (Figure 3.8). The trial showed that the number of PCR cycles could be reduced to 33 and still obtain reasonable yields of *paxQ*. Amplification of *paxQ* was repeated using the 10^{-2} dilution of oligo dT cDNA with 33 PCR cycles (Figure 3.9).

Fig 3.9 **PCR amplification of *paxQ* for cloning.**

Full length *pax*Q amplified from oligo dT primed cDNA.

Full length *pax*Q amplified with primers P2RL1+P2RL6 from oligo dT primed cDNA or genomic DNA (gDNA) templates. 0.7% agarose gel.
Lane M, 1kb+ ladder. Lanes 1and 2 represent 1:100, dilutions of oligo dT cDNA template. Lane 3, gDNA. Lane 4, no template. Expected sizes of PCR products are as follows: Lanes 1 and 2, 1.6 kb. Lane 3, 2.2 kb. Lane 4, no product. See Table 2.3 for incubation conditions.

Two 50 µL reactions were pooled and the products cloned (Section 2.5) without purification into pGEM-T Easy (Table 2.1). The ligation products were transformed (Section 2.6) into *E. coli* XL1-blue competent cells and the resulting transformants screened using colony PCR with primers M13(lacZ)Forward and M13(lacZ)-Reverse. Two clones were identified as containing the correct sized insert DNA. These two plasmids were digested with *Eco*RI to further characterise the insert DNA. The products of the digest were a 3.0 kb band, a 0.5 kb band and a 1.1 kb band, corresponding to the predicted linear pGEM-T-easy, 5' half of *paxQ* and 3' half of *paxQ* respectively. One clone was selected for sequencing with primers T7, Pax20, Pax34 and SP6. The sequencing results showed the clone

contained one base change to that of the genomic sequence located 35 bp 3' to the stop codon (Appendix 6.2.3). Given that this error is in the 3' non-translated region of the gene it would have no effect on the polypeptide sequence of PaxQ. The *paxQ*-pGEM-T Easy clone was named pRL3 (Table 2.1).

The *paxQ* insert was digested (Section 2.4) from pRL3 with restriction enzymes *Bam*HI and *Xho*I, unique restriction sites introduced in primers P2RL1 and P2RL6 respectively. The digest was performed as for *paxP* (Section 3.1.3.1) except for the enzymes used. The *Bam*HI/*Xho*I *paxQ* insert was ligated into pGEX-6P-3 expression vector digested with *Bam*HI and *Xho*I. Colony PCR (Section 2.3.4) analysis of resulting transformants with primers pGEX 5' and pGEX 3' identified clones with correctly sized *paxQ* inserts. The insert DNA of one such clone was sequenced (Section 2.10) with primers pGEX 5', Pax20, Pax34, PaxP2P1 and Pax60 to check the accuracy of the PCR. The sequence data of the clone corresponded to that of the genomic coding sequence and also confirmed it was ligated in the correct orientation. The clone did retain the base change outside of the coding region that was identified in pRL3 (Appendix 6.2.4). The *paxQ*-pGEX-6P-3 clone was named pRL4.

3.2 Recombinant expression of GST-*paxP* and GST-*paxQ* in *E. coli*

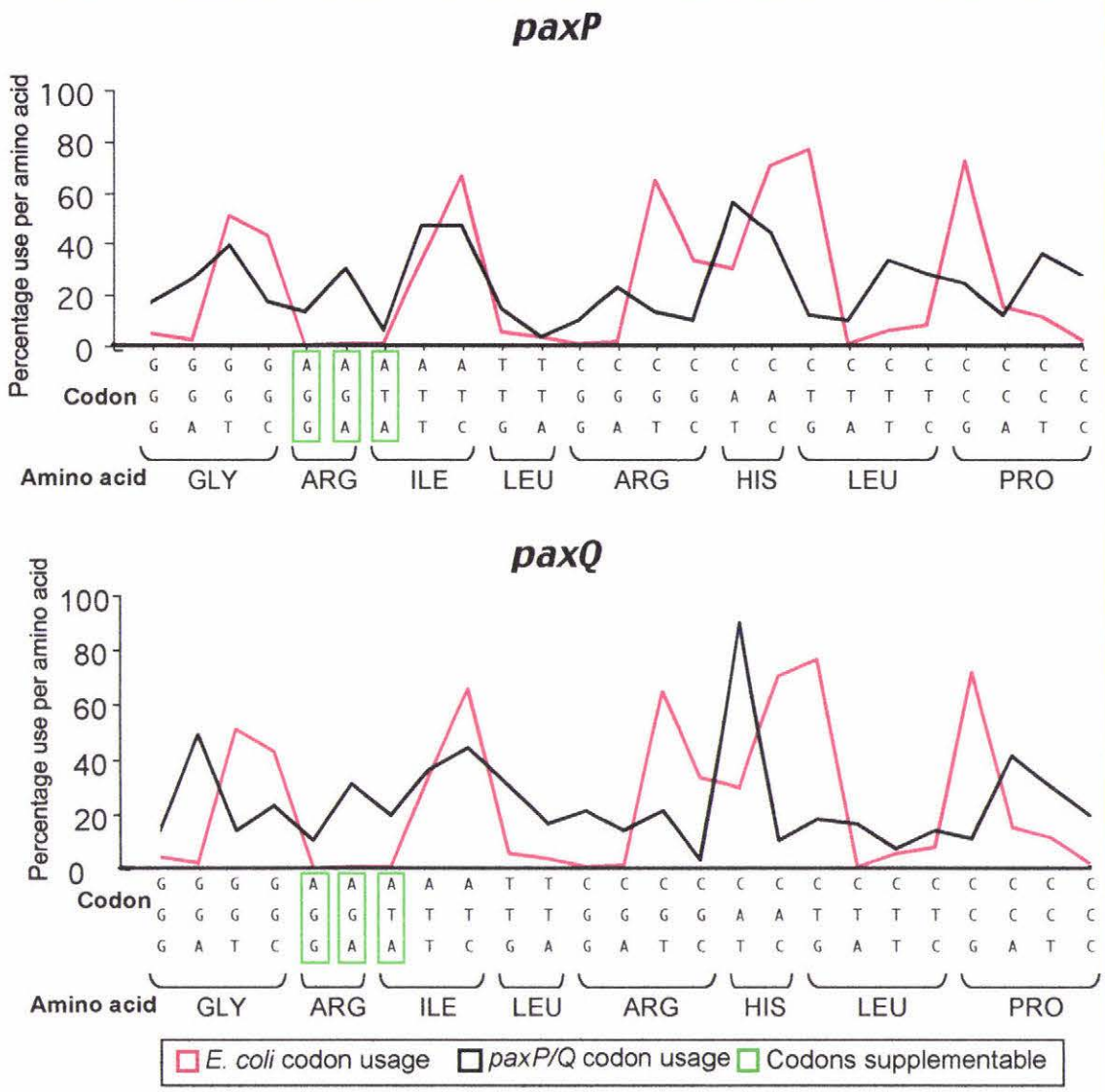
3.2.1 Expression of GST-PaxP from pRL2

Genes of eukaryotic origin often contain codons which are rarely used in *E. coli* (Hénaut and Danchin, 1996). This codon bias can limit the expression of eukaryotic genes in *E. coli* especially during the exponential phase of growth. To predict whether supplementation of tRNA genes for rare codons would be of benefit during heterologous expression, codon usage in a *paxP* and *paxQ* was compared to that of *E. coli* (Fig 3.10).

Codon usage in *paxP* and *paxQ* was calculated as percent usage per amino acid and compared to codon usage in *E. coli* genes that are highly expressed during exponential growth. The analysis revealed five codons in *paxP* cDNA which were infrequently

used in *E. coli* genes that are highly expressed during exponential growth, GGA (glycine), AGG, AGA, CGA, (arginine) and CCC (proline). Rare *E. coli* codons contained in *paxQ* cDNA consisted of GGA (glycine), AGG, AGA, CGG (arginine), ATA (isoleucine), and CCC (proline). A vector, pSJS1240 (Table 2.1), is availible which will supplement transformed cells with tRNAs for arginine codons AGG and AGA, and the isoleucine codon ATA (highlighted in green, Fig 3.10). It was predicted that cotransformation (Section 2.6) of the cDNA expression vector with pSJS1240 could potentially increase the expression of GST-PaxP and GST-PaxQ in BL21 *E. coli* cells.

Fig 3.10 **Rare codon usage in *paxP* and *paxQ***

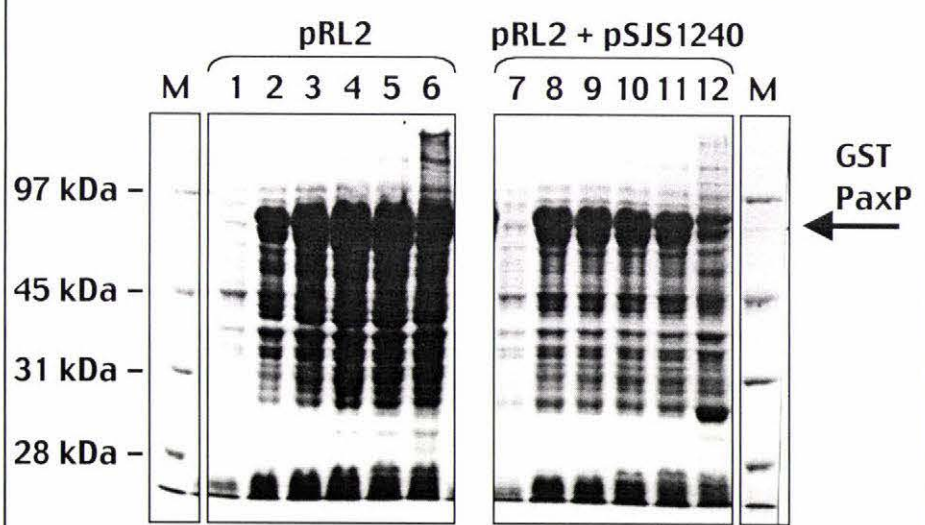


paxP and *paxQ* contain codons which are rarely used in *E. coli* during exponential growth.

Codon frequency for *paxP* was calculated using the CODON FREQUENCY program from the Genetics Computer Group (GCG) "Wisconsin Package Version 9.1. This data was then compared to the codon usage of genes in *E. coli* that are highly expressed during exponential growth (Hénaut and Danchin, 1996). Codons included in this figure were manually selected.

Yield of GST-PaxP expressed from BL21/pRL2 cells was compared with that from BL21 co-transformed with pRL2 and pSJS1240 (Figure 3.11). The two plasmids contained mutually compatible origins of replication and could be simultaneously maintained in *E. coli*. Co-transformants were selected for by plating on LB agar containing both amplicillin and spectinomycin antibiotic. BL21 *E. coli* was selected as the expression strain because it contains two mutations which make it deficient in lon and ompT proteases. Cells were harvested at 1, 2, 3 and 4 hours post-induction and also after an overnight induction period (Section 2.11). Analysis of these samples by SDS/PAGE (Section 2.13) showed that a protein of the predicted size of GST-PaxP (87 kDa) was highly expressed in all induced samples (henceforth assumed to be GST-PaxP). High yields of the predicted GST-PaxP fusion protein were visible by SDS/PAGE after a one hour induction period in both cultures (lanes 2 and 8). The yield of GST-PaxP increased post-induction reaching a maximum after three hours (lanes 3, 4 and 5). The yield of high molecular weight proteins was greatest in the overnight

Fig 3.11 **GST-PaxP expression supplemented with tRNA genes for rare *E. coli* codons.**



Relative expression levels of GST-PaxP in BL21 *E. coli* is enhanced by supplementation with tRNAs for rare codons.

Expression analysis of BL21/pRL2 (lanes 1-6) and BL21/pRL2+pSJS1240 (lanes 7-12) *E. coli* cultures grown at 37°C, after induction with 1mM IPTG, 10% SDS/PAGE gel stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lanes 1 and 7, no induction. Lanes 2-6 and 8-12 represent 1, 2, 3, 4 hour and overnight incubations post-induction respectively. Expected size of GST-PaxP is 87 kDa.

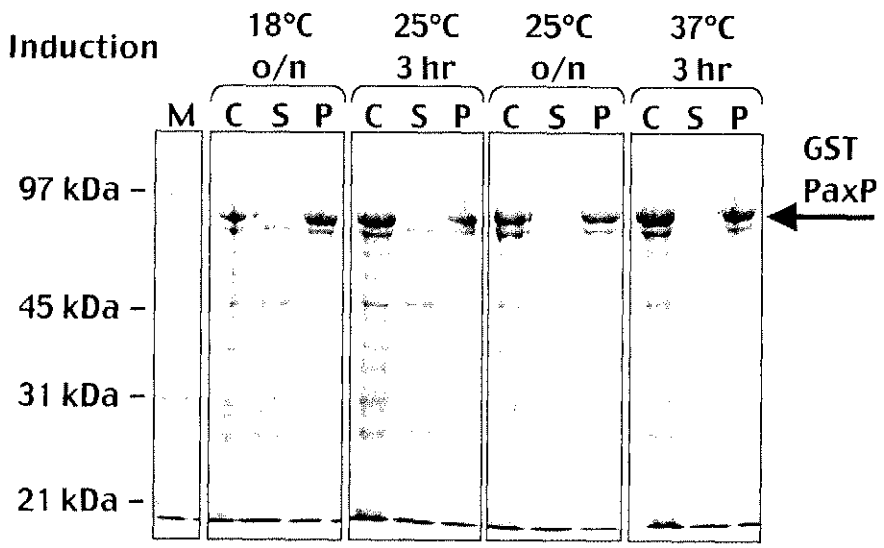
sample (lane 6). Yield of GST-PaxP expressed in the presence of pSJS1240 (Table 2.1) appeared to stay constant in samples representing 2, 3 and 4 hours post-induction (lanes 9, 10 and 11). The sample incubated overnight had a reduced yield of GST-PaxP. Overnight induction resulted in reduced yield of GST-PaxP (lane 12). Qualitative analysis suggests that pSJS1240 increases the relative amount of the GST-PaxP compared to other cellular proteins. Given that expression levels of GST-PaxP with or without cotransformation of pSJS1240 were acceptably high it was decided the use of tRNA supplementation was not required.

Both pRL2 and pRL4 expressions vectors contain 3' eukaryotic trailer sequence. The pRL2 construct contains 30 bp of *paxP* 3' trailer sequence while the pRL4 construct contains 58 bp of *paxQ* 3' trailer sequence. These sequences will not be translated as they are 3' to the stop codon. It is possible that the presence of non-translated eukaryotic sequences may affect the expression of *paxP* and *paxQ* in *E. coli*, however, the pRL2 and pRL4 vectors both generated acceptable expression levels.

3.2.2 Analysis of the solubility of GST-PaxP and GST-PaxQ expressed in *E. coli*.

BL21/pRL2 and BL21/pRL4 cultures were used to express GST-PaxP and GST-PaxQ respectively. Lysed cultures were separated into supernatant and pellet fractions by centrifugation (Section 2.11.3). GST-PaxP was visible as a distinct band corresponding to its predicted mass of 86 kDa. GST-PaxP was expressed as insoluble protein at 18°C, 25°C and 37°C for all induction periods tested (Figure 3.12). Unexpectedly, another highly expressed protein was visible with a mass of approximately 70-80 kDa. This is possibly the heat shock response protein, DnaK (Sherman and Goldberg, 1992). The common bands present in both the supernatant and pellet fractions indicate cell lysis was incomplete. A BL21/pRL4 culture was used to express GST-PaxQ in *E. coli*. Cells were processed the same as with the BL21/pRL2 culture (see above). GST-PaxQ was present in the pellet fraction of cells grown at 25°C and 37°C for induction periods of 3 hours and overnight (Figure 3.13). GST-PaxQ was visible as a distinct band at the predicted size of 85 kDa. Similar to GST-PaxP overexpression, a band thought to be DnaK was seen slightly below GST-PaxP when visualised by SDS/PAGE (Section 2.13).

Fig 3.12 **Solubility characteristics of GST-PaxP when expressed at 18°C, 25°C and 37°C**

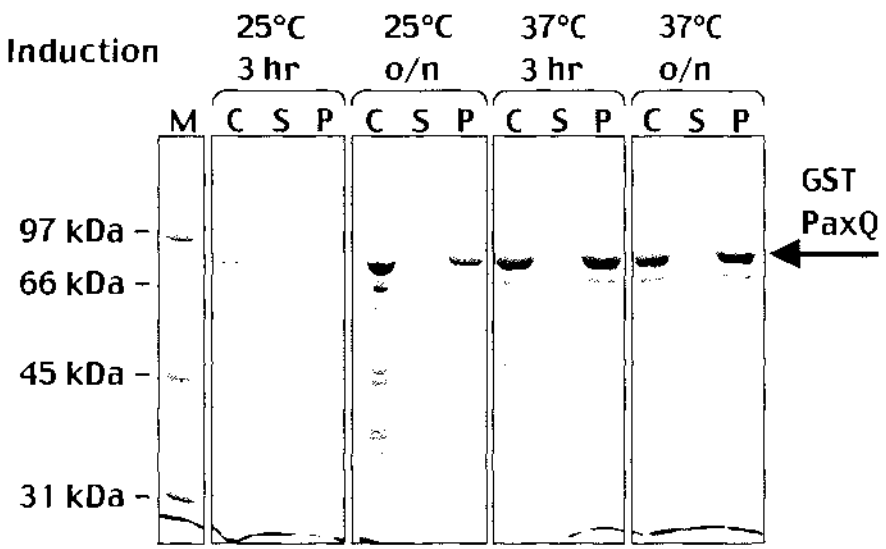


GST-PaxP is expressed as insoluble protein in *E. coli* at 18°C, 25°C and 37°C.

Solubility analysis of BL21/pRL2 *E. coli* cultures induced with 0.2 mM IPTG at 18°C, 25°C or 37°C for 3 hours or overnight. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of GST-PaxP is 87 kDa.

In order to promote the synthesis of a soluble fusion protein, GST-PaxP was co-expressed with the *E. coli* GroES and GroEL chaperone proteins (Figure 3.14). Overexpression of these chaperones can enhance the solubility of recombinant proteins expressed in *E. coli* (Goenka and Rao, 2001; Imamura *et al.*, 1999; Lee and Olins, 1992). Induced cultures of BL21/pRL2 and BL21/pRL2+pGroESL (Table 2.1) were incubated at 18°C, 25°C and 37°C. For all the conditions tested GST-PaxP was found as an insoluble product in the pellet fraction after cell lysis and centrifugation (Section 2.11.3). The GroES and GroEL proteins were visible as bands corresponding to the predicted masses of 10 kDa and 57 kDa respectively. Both proteins were present in the supernatant fraction at all temperatures and induction periods tested. The solubility of GroES and GroEL indicates they were correctly folded.

Fig 3.13 **Solubility characteristics of GST-PaxQ when expressed at 25°C and 37°C**



GST-PaxQ is expressed as insoluble protein in *E. coli* at 25°C and 37°C.

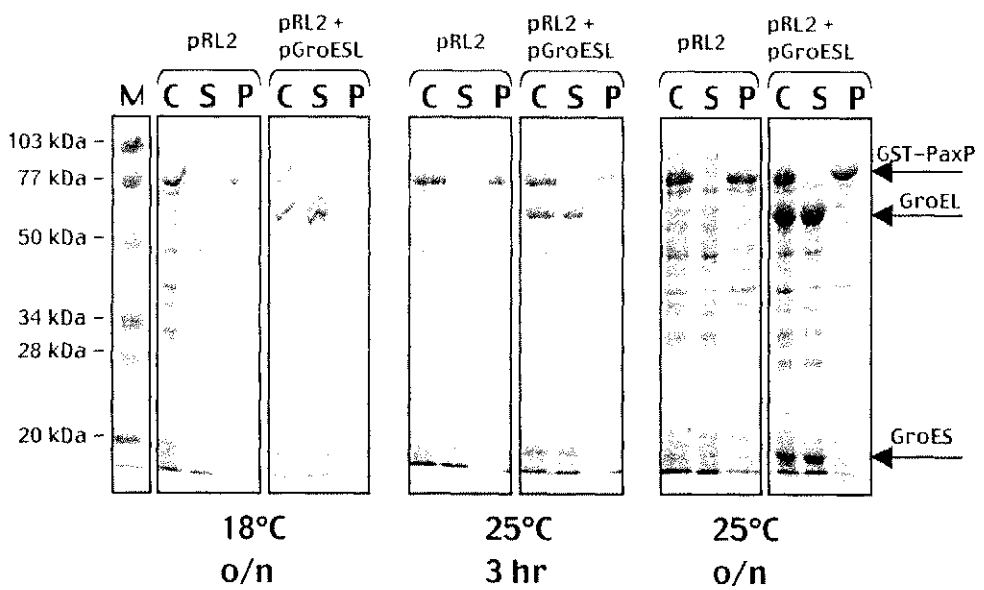
Solubility analysis of BL21/pRL4 *E. coli* cultures induced with 0.2 mM IPTG at 25°C or 37°C for 3 hours or overnight. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of GST-PaxQ is 85 kDa.

Formation of disulphide bonds is also thought to be a contributing factor towards expression of soluble proteins in *E. coli* (Prinz *et al.*, 1997). The Origami B strain of *E. coli* (Section 2.1.4.2) contains mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes which results in an increase in the efficiency of disulphide bond formation in the cytoplasm (Stewart *et al.*, 1998). Since both PaxP and PaxQ were thought to contain disulphide bonds, expression of GST-PaxP was examined in the Origami B host strain background (Figure 3.15). Cultures of Origami B/pRL2 were induced with IPTG at 18°C and 25°C for either 3 hours or overnight (Section 2.11.1). GST-PaxP was expressed as insoluble protein, present entirely in pellet fractions across all conditions tested. For these experiments a concentration of

0.1 mM IPTG was used for the induction as the *lacY* (permease) gene is deleted in Origami B resulting in more uniform uptake of IPTG compared to BL21 *E. coli*. Although Origami B usage did not alter the solubility of GST-PaxP, it was used for all further expression studies because it responds to IPTG induction in a true concentration dependant manner.

Since GST-PaxP and GST-PaxQ were expressed as insoluble protein under all conditions tested, a different approach involving detergent mediated solubilisation was adopted.

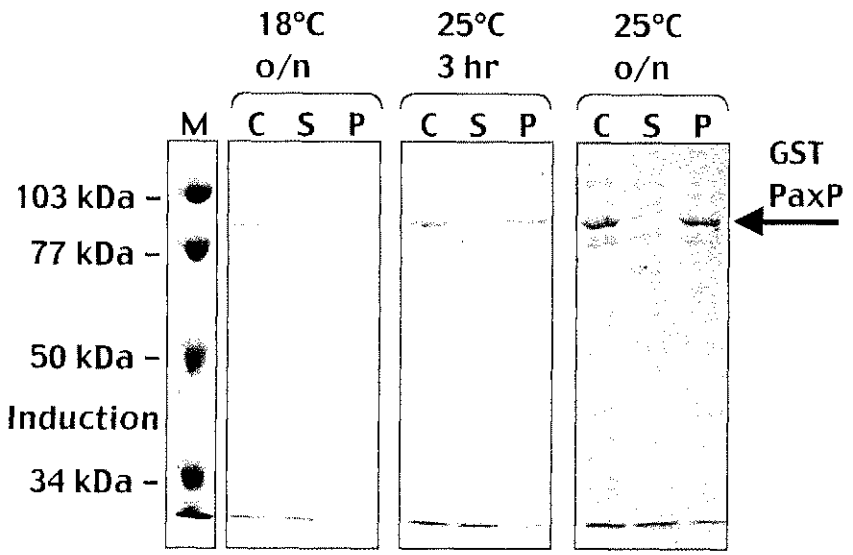
Fig 3.14 **Co-overexpression of GroES and GroEL *E. coli* chaperones with GST-PaxP**



Co-overexpression of GroES and GroEL chaperones does not alter the solubility of GST-PaxP

Expression analysis of BL21/pRL2 and BL21/pRL2+pGroESL. *E. coli* cultures induced with 0.2mM IPTG at 18°C and 25°C for 3 hours or overnight. 10% SDS/PAGE gels stained with *coomassie brilliant blue*.
Lane M, Pre-stained low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected sizes of polypeptides are as follows: GST-PaxP, 87 kDa. GroES, 10 kDa. GroEL, 57 kDa.

Fig 3.15 **Solubility characteristics of GST-PaxP when expressed in Origami B *E. coli* at 18°C and 25°C**



GST-PaxP is expressed as insoluble protein in Origami B *E. coli* at 18°C and 25°C.

Solubility analysis of Origami B/pRL2 *E. coli* cultures induced with 0.1 mM IPTG at 18°C and 25°C for 3 hours or overnight. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of GST-PaxP is 87 kDa.

3.3 Solubilisation of GST-PaxP with *N*-lauryl-sarkosine

3.3.1 Overview of the solubilisation of insoluble proteins with *N*-laurylsarkosine

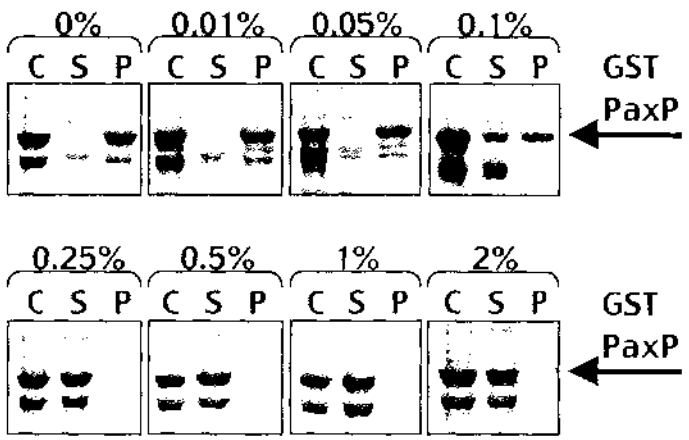
Affinity chromatography is normally undertaken to purify the protein of interest from a crude lysate, but the protein must be in a soluble state for this to occur. A method for the solubilisation of GST-PaxP involves the addition of *N*-laurylsarkosine to the lysis buffer in order to solubilise the protein of interest. The procedure for the solubilisation of GST-PaxP and GST-PaxQ follows that of Frangioni and Neel (1993) with minor alterations. Affinity chromatography of GST fusion proteins is inhibited by the presence of *N*-laurylsarkosine (Frangioni and Neel, 1993) and so purification is performed in the presence of Triton X-100 to sequester the detergent away from the protein into mixed micelles. The method was shown to be effective in the purification of catalytically active tyrosine phosphatase 1B (Frangioni and Neel, 1993) and as such appeared suited to GST-PaxP and GST-PaxQ isolation.

3.3.2 Determination of the minimum concentration of *N*-laurylsarkosine required to solubilise GST-PaxP

The minium concentration of *N*-laurylsarkosine required to solubilise GST-PaxP must be determined empirically as each protein has different solubility characteristics. The concentration of *N*-laurylsarkosine in the lysis buffer was varied between 0-2% (w/v) (Figure 3.16). Origami B/pRL2 cultures were grown at 37°C, shifted to 25°C for an overnight induction and lysed in the presence of *N*-laurylsarkosine. When *N*-laurylsarkosine was absent, GST-PaxP was present entirely in the pellet fraction. At 0.01% (w/v) and 0.05% (w/v) *N*-laurylsarkosine, GST-PaxP was still insoluble. However, at 0.1% (w/v), GST-PaxP was split evenly between supernatant and pellet fractions. At 0.25% (w/v) to 2% (w/v) *N*-laurylsarkosine concentrations, GST-PaxP was entirely soluble. A concentration of 0.25% (w/v) was selected for further

experiments as it was the minimum concentration required for full solubilisation of GST-PaxP. Higher concentrations could potentially denature the protein.

Fig 3.16 **The effect of *N*-laurylsarkosine on the solubility of GST-PaxP**



GST-PaxP is soluble in the presence of 0.1-0.25% *N*-laurylsarkosine.

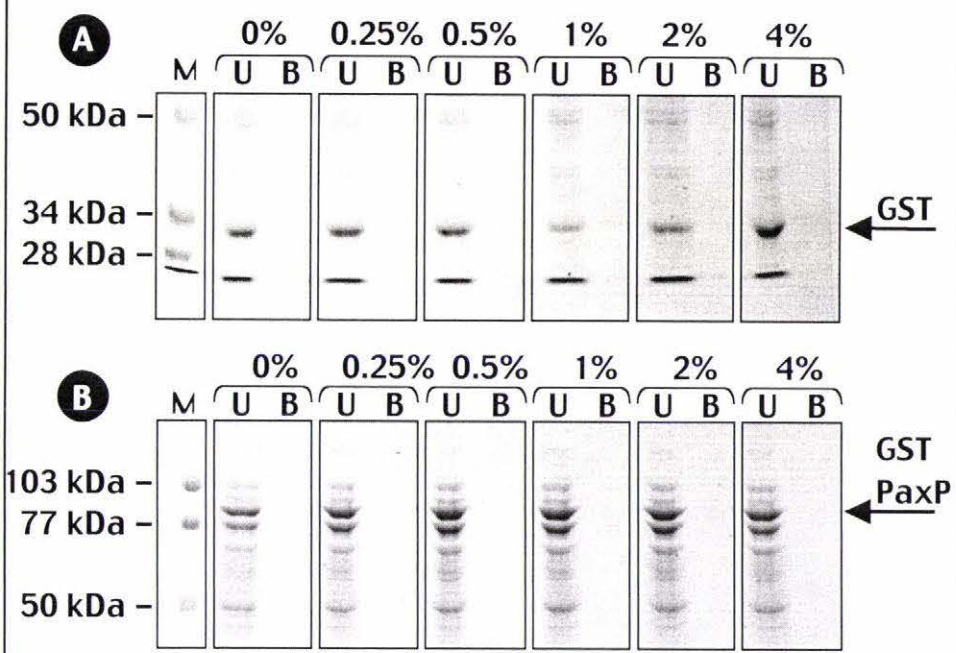
Solubility analysis of Origami B/pRL2 *E. coli* cultures induced with 0.1 mM IPTG at 25°C overnight. Cells were lysed in the presence of 0-2% (w/v) *N*-laurylsarkosine. 10% SDS/PAGE gel stained with coomassie brilliant blue.
Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction.

3.3.3 Determination of the minimum concentration of Triton X-100 required for affinity chromatography of solubilised GST-PaxP.

Once the protein is solubilised (Section 2.11.4), *N*-laurylsarkosine must be removed before affinity chromatography can take place. The *N*-laurylsarkosine is sequestered in mixed micelles after the addition of Triton X-100 to the affinity chromatography solution. The optimal concentration of Triton X-100 must be determined empirically. GST-PaxP was solubilised with 0.25% *N*-laurylsarkosine and affinity chromatography (Section 2.12) was performed in the presence of 0-4% (v/v) Triton X-100 (Figure 3.17). Affinity chromatography was performed using glutathione-sepharose beads which

specifically bind the GST protein of the GST-PaxP fusion. As a positive control for binding affinity, GST alone (part A) was compared to GST-PaxP (part B). The GST protein was subjected to the same detergent concentrations as GST-PaxP. Binding of both GST and GST-PaxP to the affinity beads was performed at 4°C for 30 minutes. None of the tested Triton X-100 concentrations allowed binding of either GST or GST-PaxP.

Fig 3.17 **The effect of Triton X-100 on affinity chromatography of solubilised GST and GST-PaxP**



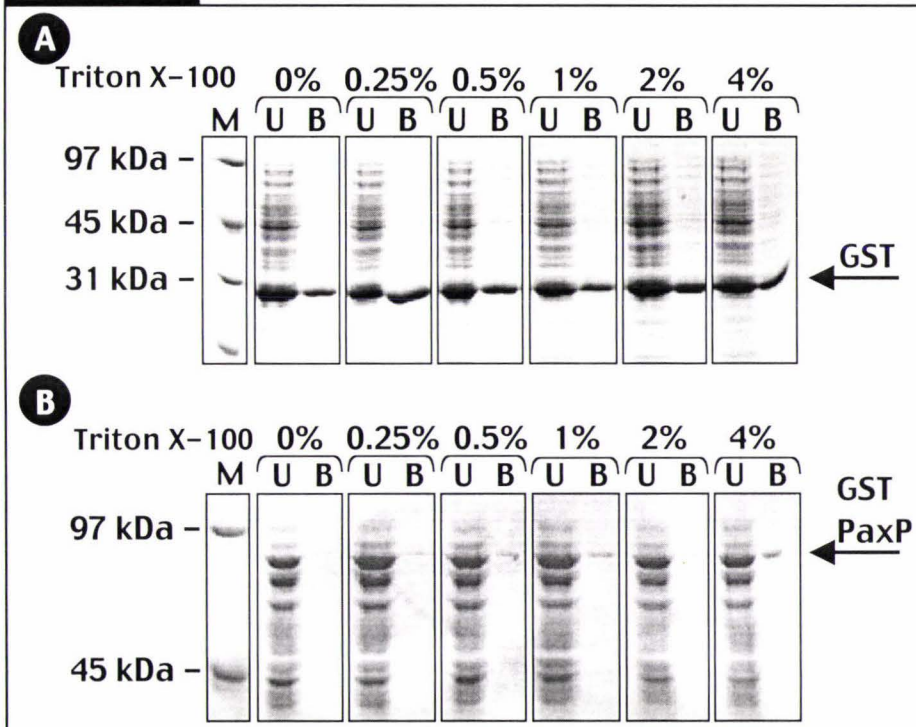
The addition of Triton X-100 has no effect on affinity chromatography of GST and GST-PaxP

Affinity chromatography of solubilised GST and GST-PaxP expressed from Origami B/pGEX-6P-3 and Origami B/pRL2 *E. coli* cultures respectively. Cultures were induced with 0.1 mM IPTG at 25°C overnight. Cells were lysed in the presence of 0.25% (w/v) *N*-laurylsarkosine, and soluble fractions incubated with glutathione-sepharose affinity beads. Affinity chromatography was performed in the presence of 0%, 0.25%, 0.5%, 1%, 2% and 4% (v/v) Triton X-100. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, prestained low molecular weight range ladder. Lane U, unbound fraction. Lane B, bound fraction. Expected sizes of polypeptides are as follows: GST, 25 kDa. GST-PaxP, 87 kDa.

Given the lack of binding activity of either GST or GST-PaxP, the affinity chromatography protocol was modified to try and increase the yield of bound protein. The affinity bead incubation period was identified as a possible reason for the lack of

binding affinity. The Triton X-100 concentration optimisation was repeated with the same conditions except that the incubation period was increased from 30 minutes to overnight at 4°C. The increased binding period greatly improved the yield of bound GST during affinity chromatography (Figure 3.18). Across all concentrations of Triton X-100 roughly 30%-40% of the available GST was bound (part A). GST-PaxP bound with a much lower affinity. No binding was visible in the absence of Triton X-100 while a small band was visible in the bound fraction when 4% (v/v) Triton X-100 was used. A final concentration of 4% (v/v) Triton X-100 was chosen as the concentration that would be used for all future experiments.

Fig 3.18 **The effect of Triton X-100 on modified affinity chromatography of solubilised GST and GST-PaxP**



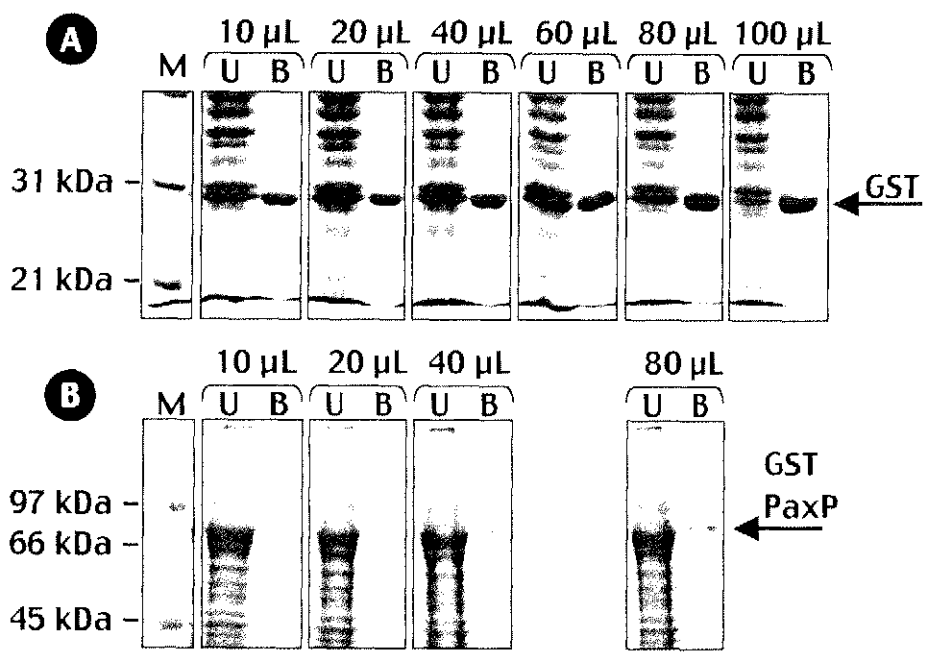
The addition of Triton X-100 has no effect on affinity chromatography of GST and GST-PaxP

Affinity chromatography of solubilised GST (A) and GST-PaxP (B) expressed from Origami B/pGEX-6P-3 and Origami B/pRL2 *E. coli* cultures respectively. Cultures were induced with 0.1 mM IPTG at 25°C overnight. Cells were lysed in the presence of 0.25% (w/v) *N*-laurylsarkosine, and soluble fractions incubated with glutathione-sepharose affinity beads. Affinity chromatography was performed in the presence of 0%, 0.25%, 0.5%, 1%, 2% and 4% (v/v) Triton X-100. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, low range ladder. Lane U, unbound fraction. Lane B, bound fraction. Expected sizes of polypeptides are as follows: GST, 25 kDa. GST-PaxP, 87 kDa

Further modifications to the affinity chromatography method (Section 2.12) were still required as yields were generally too low.

Two other variables were identified that could affect the efficiency of binding. These were, the volume of affinity beads used during the chromatography and the centrifugation speed used to pellet the beads during the purification procedure (Figure 3.19).

Fig 3.19 **The effect of affinity bead volume on affinity chromatography of solubilised GST and GST-PaxP**



10 µL of affinity beads is sufficient to perform affinity chromatography of GST and GST-PaxP.

Affinity chromatography of solubilised GST (A) and GST-PaxP (B) expressed from Origami B/pGEX-6P-3 and Origami B/pRL2 *E. coli* cultures respectively. Cultures were induced with 0.1 mM IPTG at 25°C overnight. Cells were lysed in the presence of 0.25% (w/v) *N*-laurylsarkosine and resulting soluble fractions incubated with glutathione-sepharose affinity beads. Affinity chromatography was performed in the presence of 4% (v/v) Triton X-100. The 50% bead slurry volume added to a soluble protein fraction was varied from 10 µL to 100 µL. 10% SDS/PAGE stained with coomassie brilliant blue.
Lane M, low range ladder. Lane U, unbound fraction. Lane B, bound fraction. Expected sizes of polypeptides are as follows: GST, 25 kDa. GST-PaxP, 87 kDa

Previously, 40 µL of a 50% affinity bead slurry was added to each sample and pulse centrifuged at 18000 x **g** (Section 2.12). Since the amount of affinity beads required for each purification is determined by the available protein concentration, 10 µL to 100 µL aliquots of affinity beads were tested to determine whether the bead volume was the limiting factor in the binding assay. In these experiments the centrifugation speed was reduced to five minutes at 500 x **g**. GST alone was tested with 10 µL to 100 µL of bead suspension (Figure 3.19, part A). A 10 µL volume of bead suspension was found to be sufficient to bind roughly 50% of the available GST protein. A bead volume of 100 µL bound approximately 95% of available GST.

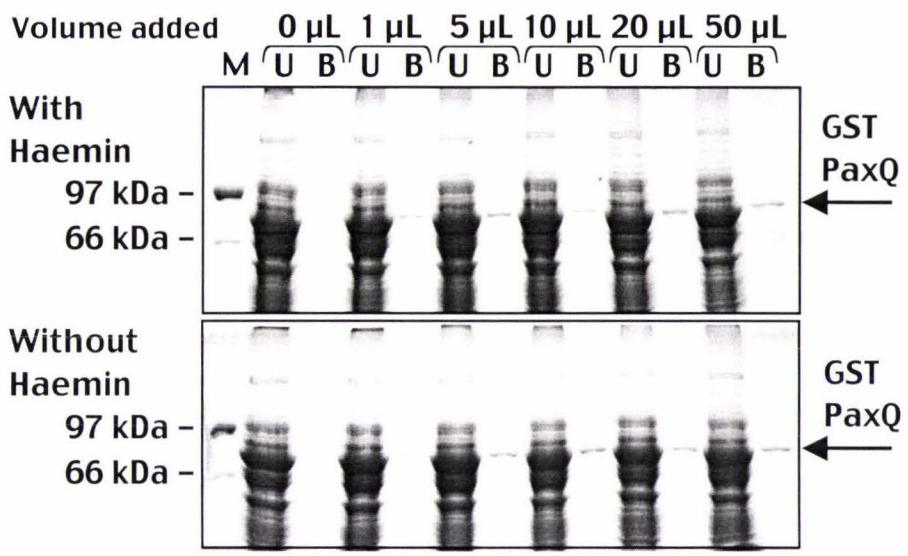
When GST-PaxP affinity chromatography was performed under the same conditions, very low levels of protein were present in the bound fraction (Figure 3.19, part B). A 10 µL volume of beads bound only small quantities of the GST-PaxP. Increasing the volume of bead slurry to 80 µL increased the yield of bound GST-PaxP but only by a small amount. The optimal volume of bead slurry for GST-PaxP binding under these conditions was determined to be 40 µL. Although the changes improved the binding efficiency of GST, the amounts of GST-PaxP that bound were still too low for a viable purification method. Further efforts were made to improve the binding efficiency of GST-PaxP.

3.3.4 Addition of the haemin ligand to enhance affinity chromatography of GST-PaxQ

It was hypothesised that a reason for the poor binding efficiency was that the majority of the solubilised protein was denatured. This denatured protein might refold under affinity chromatography conditions if the cytochrome P450 haem cofactor was added to the binding reaction, providing a nucleus for the protein to fold around. A 10 µM solution of haemin chloride was added to the GST-PaxQ affinity chromatography mixture in volumes ranging from 0 µL to 50 µL (Figure 3.20). A control solution lacking haemin chloride was also prepared and added in an identical manner. The addition of haemin chloride did not significantly alter the binding affinity of GST-PaxQ. A small change in the yield of bound GST-PaxQ was noticed as the volume of haemin chloride increased, however this was mimicked by the control solution which contained no haemin chloride. The small change in yield may be an

effect of the increased volume or possibly a pH dependent change. In order to further analyse the activity of the protein bound during affinity chromatography, recovery of bound fusion protein was attempted.

Fig 3.20 **The effect of haemin on affinity chromatography of solubilised GST-PaxQ**



The addition of haemin has no effect on affinity chromatography of GST-PaxQ

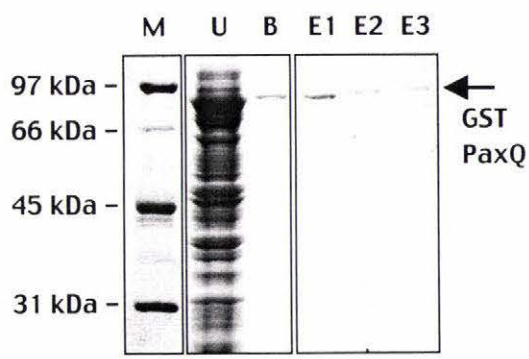
Affinity chromatography of solubilised GST-PaxQ expressed from Origami B/pGEX-6P-3 and Origami B/pRL4 *E. coli* cultures respectively. Cultures were induced with 0.1 mM IPTG at 25°C overnight. Cells were lysed in the presence of 0.25% (w/v) *N*-laurylsarkosine, and soluble fractions incubated with glutathione-sepharose affinity beads. Affinity chromatography was performed in the presence of 4% (v/v) Triton X-100 and 0 µL to 50 µL of a 10 µM haemin or control solution. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, low range molecular weight ladder. Lane U, unbound fraction. Lane B, bound fraction. Expected size of GST-PaxQ is 85 kDa.

3.3.5 Recovery of GST-PaxQ from glutathione-affinity beads

GST fusion proteins can be recovered from affinity beads by two methods, elution of the fusion protein from the beads or digestion of the protein of interest from the bound GST with a site specific protease.

GST-PaxQ was eluted from glutathione-sepharose beads upon the addition of 10 mM glutathione (Section 2.12.1.1) (Figure 3.21). The elution was performed three times and the proteins present in each of the subsequent supernatant fraction are shown in lanes E1, E2 and E3. Because of the low concentration of the eluates, they were concentrated almost two-fold before loading on the gel. GST-PaxQ was eluted from the affinity beads with low efficiency.

Fig 3.21 **Elution of GST-PaxQ from glutathione-sepharose affinity beads**

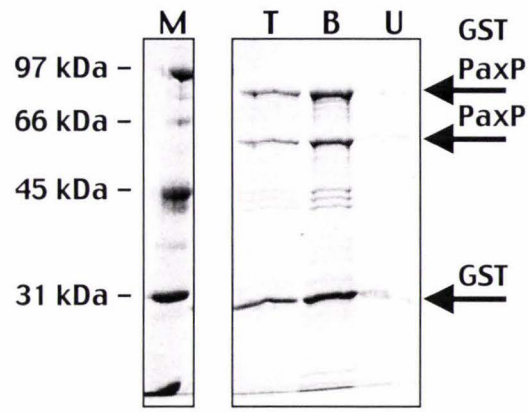


GST-PaxQ is eluted with low efficiency from glutathione-sepharose affinity beads.

Elution of solubilised GST-PaxQ bound to glutathione-sepharose affinity beads using 10 mM glutathione. 10% SDS/PAGE stained with coomassie brilliant blue.
Lane M, Low mass ladder. Lane U, unbound fraction. Lane B, bound fraction. Lanes E1,2,3, serial eluates from affinity beads. Lanes E1, E2 and E3 are 1.6 x concentrated compared to Lanes U and B. Expected size of GST-PaxQ is 85 kDa.

Digestion of GST-PaxP bound to affinity beads was also undertaken to determine if soluble PaxP protein could be released from the affinity beads (Figure 3.22). GST-PaxP was digested overnight with Prescission protease (Section 2.12.4) which recognises and cleaves a unique sequence in the linker region between the GST and PaxP components of the fusion protein. After digestion the solution was separated into bound and unbound fractions. Cleavage of GST-PaxP fusion protein was predicted to release PaxP into the supernatant. Using this method the digestion of GST-PaxP was found to be incomplete, with three major polypeptides visible in digested samples. Polypeptides corresponded in size to those of GST-PaxP (86 kDa),

Fig 3.22 **Digestion of GST-PaxP bound to affinity chromatography beads**



Digestion of GST-PaxP with Prescission protease does not release bound PaxP from the affinity beads.

Digestion of solubilised GST-PaxP bound to glutathione-sepharose affinity beads using Prescission protease. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low mass ladder. All lanes contain digested protein. Lane T, total protein fraction. Lane B, bound fraction. Lane U, unbound fraction. Expected sizes of polypeptides are as follows: GST-PaxP, 87 kDa. PaxP, 57 kDa. GST, 25 kDa.

PaxP (58 kDa) and GST (28 kDa) after the digestion (Lane T). Upon fractionation, PaxP remained bound to the affinity beads (Lane B) and was not present in significant amounts in the supernatant (Lane U). Faint protein bands corresponding in size to GST-PaxP, PaxP and GST are visible in the unbound fraction, but are likely to be due to incomplete separation of the affinity beads from the supernatant. Three bands corresponding to a mass of 40-45 kDa are visible in lanes T and B. These are proposed to originate from the Prescission protease, which has a mass of 45 kDA. The three bands seen in this size range may be due to partial disulphide bond reduction or proteolysis of the protease. Prescission protease is predicted to remain bound to the affinity beads since it is provided by the manufacturer as a GST fusion protein lacking the Prescission protease site.

The results of both elution and digestion protocols suggested that GST-PaxP and GST-PaxQ could not be efficiently purified when expressed as full length clones. These results indicated that modifications to PaxP and PaxQ would be necessary to alter their inherent insolubility.

3.4 Construction of N-terminal deletion PaxP and PaxQ expression vectors

3.4.1 The redesign of PaxP and PaxQ expression constructs

In order to increase the solubility of PaxP and PaxQ when expressed in *E. coli*, the N-terminal region of both proteins was identified as an area for modification. Cytochrome P450 enzymes targeted to microsomal regions in eukaryotes contain a highly hydrophobic region at their N-terminals which is thought to act as a membrane interacting region (Graham-Lorence and Peterson, 1996). Removal of this hydrophobic region from several eukaryotic P450 enzymes has been shown to increase solubility (when expressed in *E. coli*) with negligible effect on the catalytic activity of the modified protein (Larson *et al.*, 1991; Li and Chiang, 1991; Pernecky and Coon, 1996).

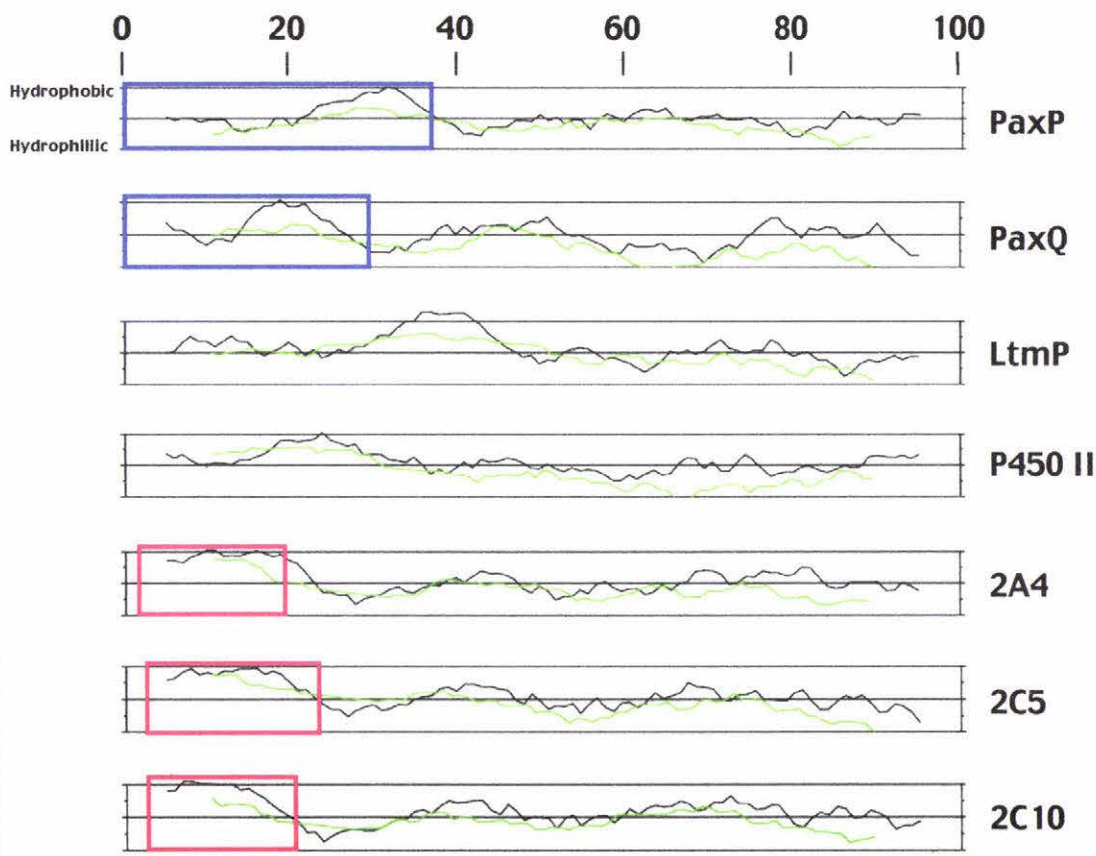
To identify the corresponding regions in the N-terminus of PaxP and PaxQ, hydropathy analysis was performed using the PEPPLOT program in the GCG bioinformatics package (Section 3.1.2). The first 100 amino acids of PaxP and PaxQ were compared to other cytochrome P450 enzymes that were either closely related, or published examples of modification (Figure 3.23). LtmP and P450II were selected on the basis of functional, and sequence similarity. Mammalian cytochromes 2A4 (Sueyoshi *et al.*, 1995), 2C5 (Cosme and Johnson, 2000) and 2C10 (Sandhu *et al.*, 1993) are published examples where the listed enzyme has been sucessfully expressed in *E. coli* and purified following N-terminal modification. Regions deleted from 2A4, 2C5 and 2C10 that were important for the expression of soluble protein in *E. coli* are highlighted with a red box.

The analysis reveals a hydrophobic region of roughly 20 amino acids in all the P450 sequences shown but the position of this region varies. It is this region which is thought to be important for membrane binding and was deleted in the published examples of N-terminal modification. Based on this analysis the hydrophobic regions identified in PaxP and PaxQ (residues 1-39 and 1-29 respectively) were targeted for removal. Fusions were prepared with both GST and thioredoxin (*trx*A), the latter is proposed to

offer enhanced solubility to fusion protein partners (Holmgren, 1985; LaVallie *et al.*, 1993).

Regions deleted from PaxP and PaxQ are highlighted in a pink box, involving residues 1-39 and 1-29 respectively. It was decided to express Δ1-39PaxP and Δ1-29PaxQ as both GST and Thioredoxin fusion proteins.

Fig 3.23 **Hydrophobicity analysis of the N-terminal regions of PaxP and PaxQ**



The N-termini of PaxP and PaxQ contain a hydrophobic region similar to other cytochrome P450 enzymes.

Hydrophobicity analysis of amino acids 1 to 100 from PaxP (*Penicillium paxilli*, accession # AAK11528), PaxQ (*P. paxilli*, Accession # AAK11527), LtmP (*Neotyphodium lolii*), P450II (*Gibberella fujikuroi*, Accession # CAA75566), P450-2C5 (Oryctolagus *cuniculus*, Accession # 10835506), P450-2A4 (*Mus musculus*, Accession # NP_034127), P450-2C10 (*Homo sapiens*, Accession # NP_000762).
Black Plot, Kyte-Doolittle scale for hydrophobicity (Kyte and Doolittle 1982). **Green plot**, GES scale for hydrophobicity (Engelman *et al.* 1986).
The **blue box** represents amino acids deleted from PaxP (1-36) and PaxQ (1-29). The **red box** represents regions that when removed resulted in increased protein solubility when expressed in *E. coli*.

The thioredoxin enzyme has been shown to confer a high degree of solubility to a fusion-partner enzyme. In addition, the thioredoxin enzyme (10 kDa) is smaller in size compared to glutathione-S-transferase (25 kDa) and would reduce the size of the resultant fusion protein. These two factors provided the basis of the decision to use the pThioHisA expression vector.

3.4.2 Construction of N-terminal modified PaxP and PaxQ GST fusion expression vectors

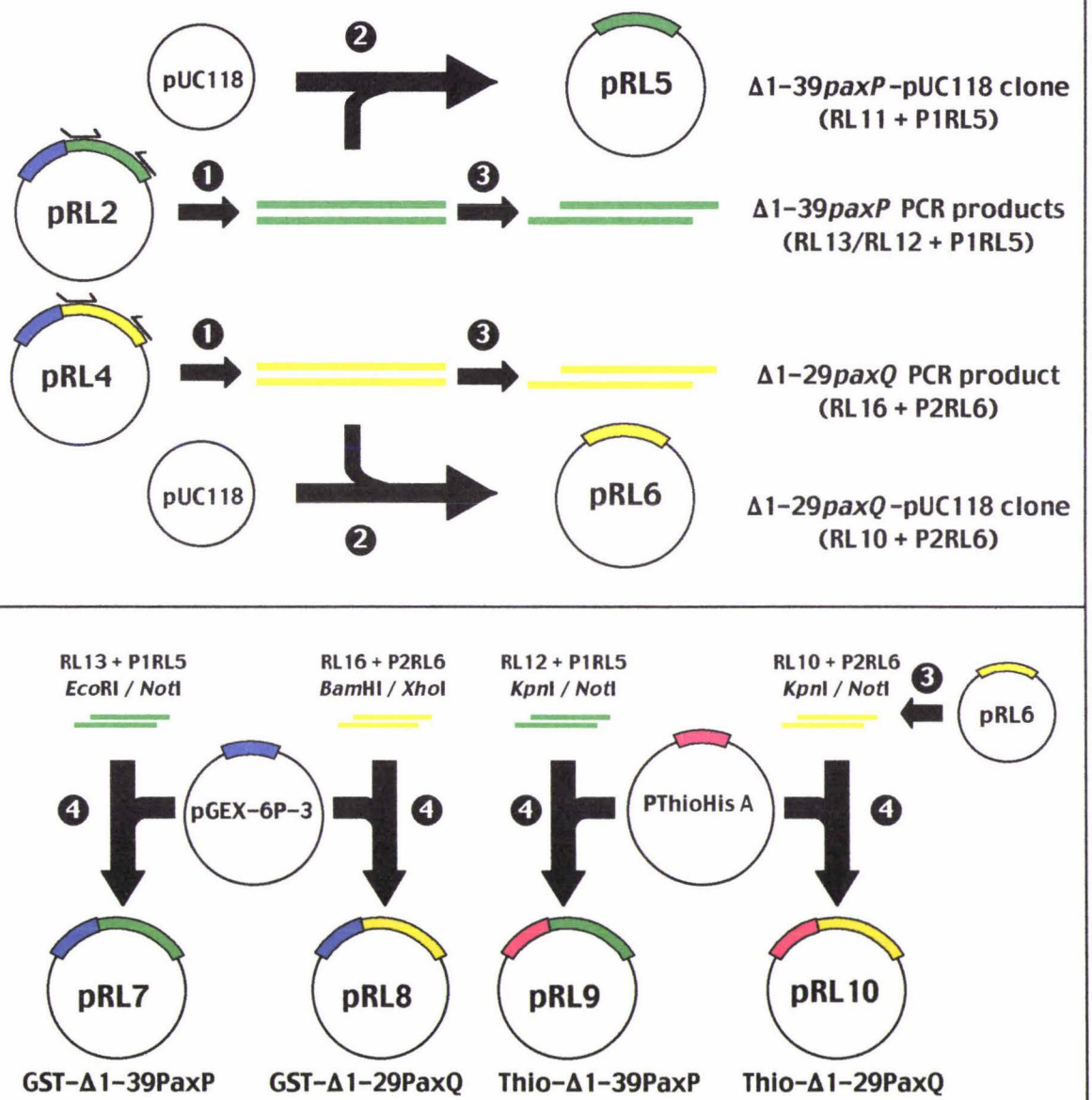
The strategy to construct N-terminal deletions of PaxP (Δ1-39*PaxP*) and PaxQ (Δ1-29*PaxQ*) expression vectors was altered compared to that previously used for constructing PaxP and PaxQ expression vectors (Section 3.1). With the availibility of cloned cDNA copies of both *paxP* (pRL2) and *paxQ* (pRL4) (Table 2.1), deletion constructs could be prepared directly from these templates by PCR (Figure 3.24) amplification of the targeted regions.

Primers RL13 and P1RL5 (Table 2.2) were used to amplify Δ1-39*paxP* with Stratagene *Pfu* turbo DNA polymerase (Section 2.3.1.1). Stratagene *Pfu* Turbo DNA polymerase was used for these PCR reactions because it is claimed to synthesise DNA with greater accuracy than Roche Expand High Fidelity DNA polymerase. The PCR product was gel purified (Section 2.7.1), digested with *Eco*RI and *Not*I restriction enzymes (Section 2.4) and then ligated (Section 2.5) into pGEX-6P-3 (Table 2.1). The products of the ligation were transformed into *E. coli* Origami B cells (Section 2.6). Ampicillin resistant colonies were tested for the presence of a Δ1-39*paxP* insert by inducing protein expression and screening for the presence of Δ1-39PaxP-GST fusion proteins of the expected size (81 kDa). One clone was selected and sequenced (Section 2.10) using pGEX 5', Cyp-8, Pax29 and pGEX-3' primers. The sequence data showed the clone contained no base changes relative to genomic sequence. The Δ1-39*paxP*-pGEX-6P-3 clone was named pRL7 (Table 2.1).

Primers RL16 and P2RL6 amplified Δ1-29*PaxQ* from the plasmid pRL4 using *Pfu* turbo DNA polymerase (Section 2.3.1.1). The PCR product was gel purified (Section 2.7.1) and digested with *Bam*HI and *Xho*I restriction enzymes (Section 2.4). The Δ1-29*paxQ* PCR product was then ligated into pGEX-6P-3 and the products

transformed into *E. coli* Origami B cells (Section 2.6). Transformants were screened by colony PCR (Section 2.3.4) with primers pGEX5' and pGEX3' (Table 2.2) to identify clones with Δ1-29*paxQ* inserts. Selected clones were further analysed by inducing Δ1-29PaxQ-GST expression and screening for a polypeptide of 82 kDA.

Fig 3.24 **Strategy to construct N-terminal deleted PaxP and PaxQ expression vectors**



Step 1: Synthesis of N-terminal deleted versions of *paxP* and *paxQ* using PCR amplification from pRL2 and pRL4 expression vectors. Primers contain unique restriction sites as hanging tails or mismatching bases.

Step 2: Blunt-end ligation of PCR product into pUC118 digested with *Sma*I.

Step 3: Digestion of PCR product or plasmid with restriction enzymes to produce cohesive-ended PCR product.

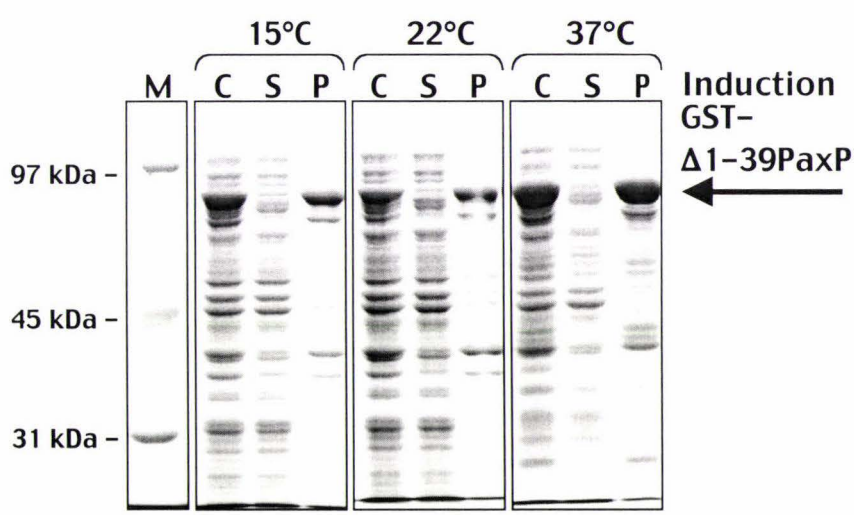
Step 4: Ligate Δ1-39*paxP* or Δ1-29*paxQ* into pGEX-6P-3 or pThioHis-A

One clone that expressed a polypeptide of this size was sequenced (Section 2.10) using primers pGEX5', Pax60, Pax34 and PaxP2P1 (Table 2.2). Sequence data revealed two changes within the coding region of Δ1-29*paxQ*. One change was positioned 458 bp from the *paxQ* ATG changed a TTC (phenylalanine) codon to a TTA (leucine) codon (Appendix 6.2.5). The second change was positioned 810 bp from the *paxQ* ATG (Appendix 6.2.6). This change introduced a silent mutation, changing an ACC to a ACA codon, both coding for threonine. Since the TTC to TTA mutation resulted in a conserved amino acid substitution, the clone was still retained for expression analysis. The 1-29*paxQ*-pGEX-6P-3 clone was named pRL8 (Table 2.1).

3.4.3 Solubility characteristics of N-terminal deleted PaxP and PaxQ as GST fusion proteins.

Origami B/pRL7 and Origami B/pRL8 cultures were used to express GST-Δ1-39PaxP and GST-Δ1-29PaxQ respectively. Cultures were induced at 15°C, 22°C and 37°C and processed using the standard method for solubility analysis.

Fig 3.25 **Solubility characteristics of GST-Δ1-39PaxP when expressed at 15°C, 22°C and 37°C**



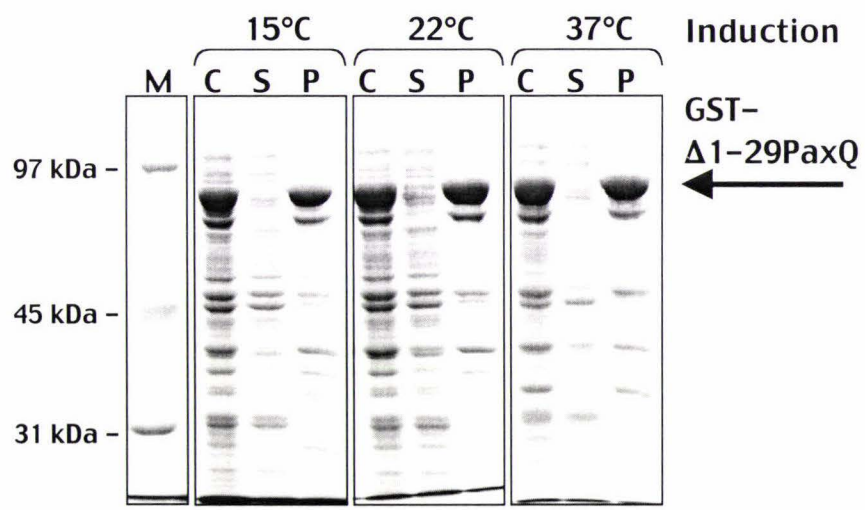
GST-Δ1-39PaxP is expressed as insoluble protein in *E. coli* at 15°C, 22°C and 37°C.

Solubility analysis of Origami B/pRL7 *E. coli* cultures induced with 0.1 mM IPTG at 15°C, 22°C or 37°C until an A_{600} cell density of 1.5-1.7 was reached. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of GST-Δ1-39PaxP is 82 kDa.

Cultures were harvested at an A_{600} cell density of 1.5-1.7 to normalise cellular protein levels. This was particulary important for the 15°C cultures which required one to two days to reach the required cell density. A polypeptide of the predicted size (82 kDa) was found entirely in the pellet fraction. This polypeptide was distinctly visible as a roughly 80 kDa protein in the pellet for samples taken at all three temperatures (Figure 3.25). What is thought to be the 70 kDa DnaK heat shock response protein is also visible just below GST-Δ1-39PaxP on the gel as a highly expressed insoluble endogenous protein.

GST-Δ1-29PaxQ (82 kDA) was also expressed as insoluble protein when expressed at temperatures of 15°C, 22°C and 37°C (Section 2.11.3) (Figure 3.26). DnaK expression was also visible in response to the overexpression of GST-Δ1-29PaxQ.

Fig 3.26 **Solubility characteristics of GST-Δ1-29PaxQ when expressed at 15°C, 22°C and 37°C**



GST-Δ1-29PaxQ is expressed as insoluble protein in *E. coli* at 15°C, 22°C and 37°C.

Solubility analysis of Origami B/pRL8 *E. coli* cultures induced with 0.1 mM IPTG at 15°C, 22°C and 37°C until a culture A_{600} of 1.5-1.7 was reached. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of GST-Δ1-29PaxQ is 82 kDa.

3.4.4 Construction of N-terminal modified PaxP and PaxQ Thioredoxin fusion expression vectors

Primers RL12 and P1RL5 were used to amplify Δ1-39*paxP* from pRL2 template for insertion into the pThioHis A expression vector (Figure 3.24). The PCR product was gel purified and single enzyme digestions performed with *Kpn*I and *Not*I restriction enzymes. The digested PCR product was ligated into pThioHis A and the products transformed into *E. coli* Origami B cells. Colony PCR (Section 2.3.4) was used to screen ampicillin resistant colonies with primers Trx fwd and Trx rev, Trx fwd and P1RL9 and finally P1RL7 and Trx rev (Table 2.2). A clone with a Δ1-39*paxP* insert was selected for sequencing with primers Trx fwd, Cyp-8, Pax29 and Trx rev (Table 2.2). Sequence data revealed a mutation 1049 bp from the *paxP* ATG start codon, converting a TGG (tryptophan) codon to a TTG (leucine) codon (Appendix 6.2.7). Since both tryptophan and leucine are hydrophobic residues, the mutation was considered conserved in nature and the construct was used for Thioredoxin-Δ1-39PaxP expression analysis. The vector was named pRL9.

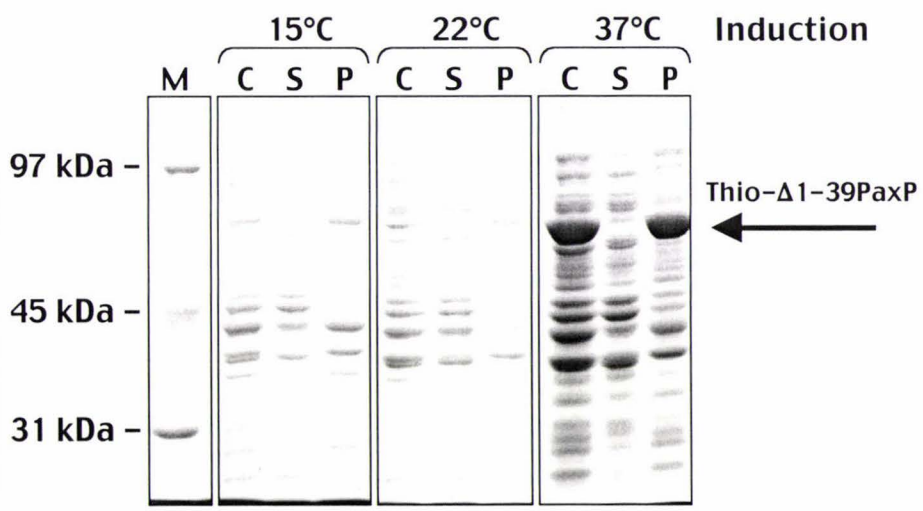
Primers RL10 and P2RL6 (Table 2.2) were used to PCR amplify Δ1-29*paxQ* from the pRL4 vector (Figure 3.24). The positioning of the *Kpn*I site in RL10 meant that Δ1-29*paxQ* could not be digested as a PCR product. Taking into account the fact *Pfu* turbo DNA polymerase (Section 2.3.1.1) synthesises blunt ended PCR products meant the Δ1-29*paxQ* PCR product had to be cloned using blunt-end ligation into pUC118 as an intermediate step before digestion. The products of the Δ1-29*paxQ*-pUC118 ligation were transformed into *E. coli* Top10 competent cells (Section 2.1.4.3), and transformants screened with colony PCR (Section 2.3.4) using primers M13(lacZ)Forward and M13(lacZ)Reverse for clones with an insert 1.6 kb in size. One clone was selected for sequencing (Section 2.10) with primers M13(lacZ)Forward, Pax34, PaxP2P1 and Pax60. Sequence data confirmed the Δ1-29*paxQ* insert contained no mutations within the coding region but did contain a change outside of the coding region which was carried over from pRL3 and pRL4 (Section 3.1.3.2). The Δ1-29*paxQ*-pUC118 clone was named pRL6 (Table 2.1). The Δ1-29*paxQ* insert was then digested from pRL6 with *Kpn*I and *Xho*I restriction enzymes (Section 2.4), gel purified (Section 2.7.1) and ligated into the pThioHis-A expression vector. The ligation products were transformed into *E. coli* Origami B cells (Section 2.6) and resulting

ampicillin resistant colonies were screened for insert identity using colony PCR (Section 2.3.4) with primers RL10 and PaxP2P1. Clones with Δ1-29*paxQ* inserts were induced for protein expression (Section 2.11.1) to check the ligation occurred as predicted. A clone that expressed the predicted 82 kDa polypeptide was chosen and sequenced (Section 2.10) with Trx fwd, confirming the insert had been digested and ligated as expected. The Δ1-29*paxQ*-pThioHis A clone was named pRL10.

3.4.5 Solubility characteristics of N-terminal deleted PaxP and PaxQ as Thioredoxin fusion proteins.

Origami B/pRL9 and Origami B/pRL10 cultures were used to express Thioredoxin-Δ1-39PaxP and Thioredoxin-Δ1-29PaxQ respectively. Cultures were induced at 15°C, 22°C and 37°C and processed using the standard method for solubility analysis (Section 2.11.3). Cultures were harvested when a A_{600} cell density of 1.5 –1.7 was reached.

Fig 3.27 **Solubility characteristics of Thio-Δ1-39PaxP when expressed at 15°C, 22°C and 37°C**



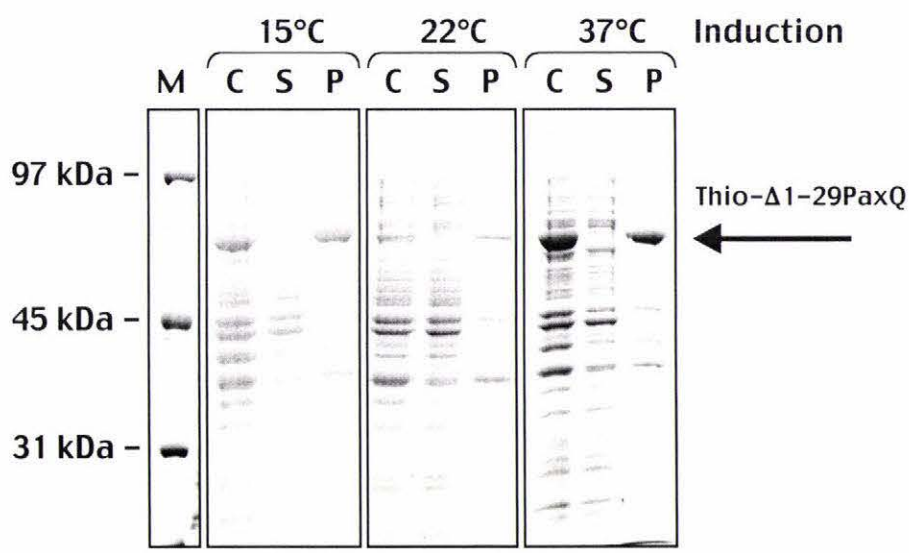
Thioredoxin-Δ1-39PaxP is expressed as insoluble protein in *E. coli* at 15°C, 22°C and 37°C.

Solubility analysis of Origami B/pRL9 *E. coli* cultures induced with 0.1 mM IPTG at 15°C, 22°C and 37°C until a culture A_{600} of 1.5-1.7 was reached. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of Thioredoxin-Δ1-39PaxP is 67 kDa.

Thioredoxin-Δ1-39PaxP was still expressed as an insoluble protein at all temperatures tested (Figure 3.27). Expression of Δ1-39PaxP as a Thioredoxin fusion was visible as a distinct band corresponding to the predicted size of 67 kDa. Expression levels were generally lower than that achieved for GST fusion proteins. The protein thought to be DnaK was present after Thioredoxin-Δ1-39PaxP overexpression, but at a reduced level when compared to Figures 3.25 and 3.26.

Expression of Δ1-29PaxQ as a Thioredoxin fusion was visible as a distinct band corresponding to the predicted size of 67 kDa. Thioredoxin-Δ1-29PaxQ was expressed as an insoluble protein at 15°C, 22°C and 37°C induction temperatures (Figure 3.28). The expression levels of Thioredoxin-Δ1-29PaxQ were similar to those achieved with Thioredoxin-Δ1-39PaxP expression.

Fig 3.28 **Solubility characteristics of Thio-Δ1-29PaxQ when expressed at 15°C, 22°C and 37°C**



Thioredoxin-Δ1-29PaxQ is expressed as insoluble protein in *E. coli* at 15°C, 22°C and 37°C.

Solubility analysis of Origami B/pRL10 *E. coli* cultures induced with 0.1 mM IPTG at 15°C, 22°C and 37°C until a culture A_{600} of 1.5-1.7 was reached. 10% SDS/PAGE gels stained with coomassie brilliant blue.
Lane M, Low molecular weight ladder. Lane C, total cell fraction. Lane S, supernatant fraction. Lane P, pellet fraction. Expected size of Thioredoxin-Δ1-29PaxQ is 67 kDa.

Chapter 4 Discussion

The aim of this project was to express PaxP and PaxQ in sufficient quantities to determine the substrates and products of their reactions. The strategy adopted was to overexpress PaxP and PaxQ in *E. coli* as GST fusion proteins. Recombinant expression of proteins as GST fusions is widely used to enable affinity purification of the protein of interest (Smith and Johnson, 1988). However, many proteins while highly expressed by this system, form insoluble inclusion bodies. Strategies to overcome this problem include refolding protocols or solubilisation with detergents. Alternatively, other expression systems have been used such as *in vitro* translation, expression in yeast species or insect-baculovirus based systems. The first step in the heterologous expression of PaxP and PaxQ was to check the expression level and solubility of each enzyme in *E. coli*.

4.1 Expression of PaxP and PaxQ in *E. coli*

High level expression of eukaryotic genes in *E. coli* can be problematic due to the limited codon usage of endogenous genes and subsequent tRNA bias (Hénaut and Danchin, 1996). Certain codons present in *paxP* and *paxQ* are used infrequently in *E. coli* and were predicted to limit the expression of the gene products (Figure 3.10). The co-expression of two *E. coli* tRNAs for AGG, AGA (arginine) and AUA (isoleucine) resulted in slightly improved expression levels of PaxP, but was not essential for the expression of GST-PaxP (Figure 3.11). Cellular fractionation revealed that both GST-PaxP and GST-PaxQ were expressed as insoluble protein (Figure 3.12 and 3.13). Several strategies were adopted to resolve this problem.

The expression of recombinant proteins in *E. coli* often results in insoluble protein aggregates. The reasons for the synthesis of insoluble proteins in *E. coli* are not defined but many factors have been analysed. The length of the polypeptide is not thought to be a definitive indicator for inclusion body formation. The 100 kDa T7 RNA polymerase enzyme has been expressed at 37°C to high levels in *E. coli* in a soluble state, while an 11 kDa fragment of bovine prochymosin formed inclusion bodies at 32°C (Schein, 1989). The overexpression of endogenous proteins in *E. coli* can lead to inclusion body formation, which discounts the concept of inclusion body formation being an active process which removes "foreign" proteins (Schein, 1989). Inclusion body formation is probably caused by the high expression rates associated with the use of heterologous

expression vectors. If the folding rate of a particular protein is relatively slow, it will be exposed to a high concentration of unfolded polypeptides before it has time to form the correct tertiary structure. The aggregation of hydrophobic polypeptides is a thermodynamically efficient way to minimise contact with hydrophilic molecules in the cytosol. This aggregation of polypeptides occurs in competition with the correct folding of the polypeptide.

4.1.1 The GroES/EL chaperones

The aggregation of highly expressed polypeptides into inclusion bodies could be due to saturation of the endogenous chaperone supply. The correct structural folding of recently synthesised polypeptides is assisted by chaperone proteins (Ellis, 1987). The two most highly characterised examples in *E. coli* are the DnaK and GroES/EL chaperones (Kusmierczyk and Martin, 2001). The DnaK chaperone functions by binding unfolded polypeptides and preventing their aggregation. The GroES/EL chaperones oligomerise into a protein complex which forms a sealed cavity. GroEL (57 kDa) heptamerises to form a ring shaped cavity which is sealed by a dome-like heptamer of GroES (10 kDa). The complex binds unfolded polypeptides and encloses them in a sealed interior cavity. This isolation allows folding to occur without interference from other polypeptides. GroEL binds the majority of endogenous *E. coli* proteins as soon as translation begins, and then processes them at the rate of synthesis. The majority of proteins bound by GroEL require more than one folding cycle to form a correct tertiary structure (Ewalt *et al.*, 1997). The overexpression of GroES/EL in concert with the expression of recombinant proteins has been shown in some cases to be beneficial. The co-overexpression of GroES/EL with mammalian Zeta-crystallin in *E. coli* resulted in a significant reduction in the formation of inclusion bodies (Goenka and Rao, 2001). However, in this study the insolubility of PaxP expressed in *E. coli* was not altered by the over-expression of GroES/EL (Figure 3.14). This could be due to incompatibility between GST-PaxP and the chaperone assembly. The folding cavity of the GroES/EL complex is large enough to enclose a polypeptide of roughly 60 kDa (Kusmierczyk and Martin, 2001). This is probably not large enough to enclose the 86 kDa GST-PaxP fusion protein. Larger proteins of up to 150 kDa have been shown to interact with GroEL, but were not released in the time-dependent manner found for smaller proteins (Ewalt *et al.*, 1997). Another factor complicating the folding of

GST-PaxP is the presence of the GST on the N-terminus of the fusion protein. GST is a highly soluble 26 kDa enzyme and as such this location may inhibit the recognition and processing of the PaxP region by chaperones. A fusion at the C-terminus would not interfere with catalytic or membrane binding functions and would leave the hydrophobic N-terminus free to be bound by chaperones as synthesis occured. A poly-histidine affinity tag could be used in place of a GST affinity tag. Since GroES/EL processing has been demonstrated to be size dependent (Ewalt *et al.*, 1997) a 6 amino acid affinity tag would allow PaxP (56 kDa) to be enclosed in the chaperone cavity.

4.1.2 Disulphide bond formation

The cytosol of *E. coli* is a strongly reducing environment. This reducing characteristic will inhibit the formation of disulphide bonds between cysteine residues in expressed proteins. The thioredoxin and glutaredoxin systems generate the reducing potential of the *E. coli* cytoplasm. Both groups of enzymes take electrons from NADPH and can catalyse the reduction of disulphide bonds in the cytoplasm. The *E. coli* Origami B cell line has been engineered to minimise the reducing potential of its cytoplasm. *E. coli* Origami B contains mutations in both the thioredoxin reductase (*trxB*) and glutathione-oxidoreductase (*gor*) genes (Stewart *et al.*, 1998). These mutations create a block in the reducing pathway, preventing electron flow from NADPH to glutathione, thioredoxin, and glutaredoxin 1, 2 and 3 enzymes. However, no difference in the solubility of expressed GST-PaxP was observed between BL21 and Origami B cells (Figure 3.15). Although PaxP and PaxQ contain four and nine cysteine residues respectively, covalent bond formation between these cysteine residues is not likely to occur in their native state as disulphide bonds are rarely found in intracellular proteins (Thornton, 1981). The modification of the genetic background and growth conditions of *E. coli* failed to identify conditions that allowed GST-PaxP/Q to be synthesised in a soluble state, therefore an investigation of protein solubilisation methods was conducted.

4.2 Protein solubilisation and refolding

The heterologous over-expression of proteins in *E. coli* often leads to the production of insoluble inclusion bodies. The frequency with which this problem arises has lead to the development of numerous methods to solubilise insoluble proteins and refold them into active protein.

The high density of inclusion bodies means they can be easily separated from endogenous protein with moderate centrifugation. Separated inclusion bodies are almost totally pure and only contain small amounts of contaminating proteins which tend to co-purify due to their hydrophobic nature. These minor contaminants can be removed by washing inclusion body preparations with detergents or very low concentrations of chaotrophic agents such as guanidine hydrochloride. If too high a concentration of detergent or chaotroph is used, the inclusion body itself will be solubilised. A benefit of protein aggregation is the inherent protection from proteolysis rendered on proteins which are trapped in the inclusion body (Lilie *et al.*, 1998).

Once the inclusion bodies have been harvested they must be solubilised so that refolding of the proteins can occur. Solubilisation of these highly aggregated polypeptides can be achieved *in vitro* by the addition of chaotropic agents such as urea or guanidine hydrochloride at concentrations of 6-8 M concentrations (Rudolph and Lilie, 1996).

The denaturant used to solubilise the proteins must be removed so that renaturation can occur. This is usually achieved through dilution or dialysis of the protein solution. Due to the tendancy of polypeptides to aggregate as the denaturant is removed, these procedures are usually performed with very low concentrations of protein. This can be problematic as excessive volumes of solution will be required to renature proteins which readily aggregate. Continuous or "pulse" renaturation provides a partial solution to this problem (Rudolph, 1990). The addition of unfolded polypeptides to the refolding buffer in a stepwise manner will allow added polypeptides time to refold before the next aliquot is added. The aggregation of polypeptides is dependant on the concentration of unfolded protein, and as such, renatured examples will not compete

with the folding process . This stepwise refolding process allows refolding to occur at higher total concentrations of protein.

The refolding process is poorly understood and requires the empiracal determination of correct conditions for optimal folding. The pH, temperature and ionic strength of the folding solution are important determinants of folding efficiency. Addition of low molecular weight molecules can have a significant effect on the folding process. Zn^{2+} or Ca^{2+} ions have been shown to help stabilise partially folded polypeptides and prevent their aggregation (Rudolph and Lilie, 1996). L-arginine has also been shown to help fold some polypeptides. It is thought that it increases the solubility of hydrophobic regions of the polypeptide in solution. In a simmilar manner, detergents can be used at low concentrations to increase the solubility of folding intermediates.

An alternative method has been developed in which application of extreme hydrostatic pressures of one to two kbar, with non-denaturing concentrations of guanidine chloride (0.25 M), will solubilise inclusion bodies and produce high yields of refolded protein (St John *et al.*, 1999).

The successful renaturation of proteins is still a poorly understood process that depends on the empirical application of various techniques. An alternative and milder treatment of insoluble proteins involving detergent-mediated solubilisation was used here.

4.2.1 Detergent mediated solubilisation

Detergents are able to solubilise proteins without denaturation. Frangioni and Neel (1993) studied the detergent mediated solubilisation of chicken muscle pyruvate kinase expressed as a GST fusion protein in *E. coli*. They tested several different detergents alone and in combination. These included, Triton X-100, *N*-octyl glucoside, CHAPSO, RIPA buffer and *N*-laurylsarkosine. *N*-laurylsarkosine was shown to be the only treatment capable of shifting the insoluble protein into solution. *N*-laurylsarkosine (0.2%) had previously been shown to solubilise actin that had been overexpressed in *E. coli* (McNally *et al.*, 1991), while Frangioni and Neel (1993) used 1.5% *N*-laurylsarkosine to solubilise pyruvate kinase. The improved performance of N-laurylsarkosine compared to Triton X-100 was attributed to its ability to inhibit the

coaggregation of proteins with components of the bacterial outer membrane. Use of this method with chicken pyruvate kinase GST fusion protein gave enzyme with fully recovered kinase activity following purification by using glutathione affinity chromatography (Frangioni and Neel, 1993). In the experiments described here, the optimal concentration of *N*-laurylsarkosine for the solubilisation of GST-PaxP was 0.25% (Figure 3.17). This concentration is very similar to that used by McNally *et al.*, (1991) to solubilise actin but is less than that used by Frangioni and Neel (1993). The use of *N*-laurylsarkosine requires the addition of Triton X-100 to sequester *N*-laurylsarkosine into mixed micelles before affinity purification can occur. Frangioni and Neel (1993) found a final concentration of 2% Triton X-100 was optimal during affinity chromatography of most solubilised GST fusion proteins they tested. However, in the case of the highly insoluble protein tyrosine phosphatase 1B protein a concentration of 4% Triton X-100 was used. The purification GST-PaxP produced the greatest yield when 4% Triton X-100 was used (Figure 3.19). This concentration (4%) is higher than that used by Frangioni and Neel (1993) to purify chicken pyruvate kinase, but the same as that required for purification of the highly insoluble protein tyrosine phosphatase 1B. All experiments with GST-PaxP/Q included GST as a positive control. The high recovery rate of active GST by these methods confirmed that the lack of binding from GST-PaxP protein was not due to a deficiency in the protocol. It does not however, prove that the PaxP member of the GST-PaxP fusion protein was still folded after treatment.

4.2.2 Affinity chromatography of solubilsed protein

The binding affinity of solubilised GST-PaxP was very low, with only a small fraction of the available protein binding the affinity beads. This could be due to the size of the GST-PaxP fusion protein retarding the yield of protein bound by the affinity beads. Frangioni and Neel (1993) purified several proteins ranging in size from 29 kDA to 83 kDa in order to analyse the relationship between mass of the protein and the saturation of glutathione affinity beads. They found that a greater amount of the smaller mass protein (29 kDa) bound, compared to the yields achieved with proteins of greater mass (83 kDa). A similar relationship was observed in this study where, even taking into account the mass differences, the amount of GST-PaxP bound was low. GST-PaxQ was also shown to exhibit the same binding characteristic as GST-PaxP

when purified under identical conditions (Figure 3.21). One possibility for this reduced binding affinity was the requirement of haem for the correct folding of GST-PaxP/Q. The addition of haem to the chromatography solution was tested but did not improve the binding affinity. The absence of the haem group would result in a cavity in the centre of the enzyme, which could result in a destabilised tertiary structure. GST-PaxP/Q expressed in *E. coli* may not contain a bound haem due to the inability of the cell to synthesise haem at a sufficient rate to keep pace with the rapid expression of the P450. The addition of haem to the affinity chromatography solution did not enhance the binding of GST-PaxQ. An alternative explanation for the reduced binding affinity was that solubilised GST-PaxP and GST-PaxQ was in a denatured state. The binding observed could be the result of non-specific reactions. In support of this hypothesis is the fact PaxQ was not released from the affinity beads after digestion with Prescission protease (Figure 3.22). The protease recognises a unique amino acid sequence in the linker region between the two fusion proteins; cleavage of this site should release PaxQ into the supernatant. The fact that PaxQ remained in the pellet fraction suggests that it was attached to the beads by non-specific interations. It is also possible that although correctly folded, the PaxP and PaxQ proteins were interacting hydrophobically with the beads. These results lead to the conclusion that detergent mediated solubilisation was not effective in the purification of GST-PaxP or GST-PaxQ. A new strategy was formed to modify the inherent insolubility characteristics of these two proteins.

4.3 N-terminal modification

A key reason for the insolubility of GST-PaxP and GST-PaxQ in *E. coli* may have been the presence of an N-terminal trans-membrane region. This is a common feature of microsomal cytochrome P450 enzymes. The trans-membrane domain is characterised by the presence of 17 to 25 hydrophobic residues at the N-terminus, which form a single membrane-spanning alpha helix. Hydrophobicity analysis using both the Kyte-Doolittle (Kyte and Doolittle, 1982) and Engelman (Engelman *et al.*, 1986) methods revealed a region in both PaxP and PaxQ which matched the criteria for a transmembrane membrane interacting region (Figure 3.24). Removal of this hydrophobic region from several members of the P450 2C family overcame inclusion body formation during expression in *E. coli* (Cosme and Johnson, 2000; Sandhu *et al.*, 1993; Sueyoshi *et al.*,

1995). Deletion of the N-terminus had little effect on the activity of the enzyme. Based on the length of the hydrophobic region and the regions deleted in P450s 2A4, 2C5 and 2C10 (Figure 3.24) it was decided to remove amino acids 1-39 from PaxP and amino acids 1-29 from PaxQ and express the modified enzymes as a fusion proteins in *E. coli*. In addition to the use of the GST expression vector, fusions were prepared with the thioredoxin fusion system. The thioredoxin gene fusion system will express a cloned gene as a fusion protein with an N-terminal thioredoxin. The thioredoxin system was chosen because it confers a high degree of solubility and a histidine affinity tag to proteins expressed as thioredoxin fusions (LaVallie *et al.*, 1993). The thioredoxin system has been used sucessfully to express plant taxadiene synthase in *E. coli* (Huang *et al.*, 1998). This enzyme is involved in the synthesis of the diterpenoid anti-cancer drug, taxol. Expression of unmodified taxadiene synthase in *E. coli* resulted in the production of inclusion bodies. Attempts to reconstitute active protein after solubilisation of the inclusion bodies was unsucessful. Expression of the same gene as a thioredoxin fusion protein yielded 15-20% soluble protein. Expression of Δ1-39PaxP and Δ1-29PaxQ as both GST and thioredoxin fusion proteins in *E. coli* yielded only insoluble protein.

4.3.1 Insolubility characteristics of Δ1-39PaxP and Δ1-29PaxQ

The insolubility of both PaxP and PaxQ when expressed without N-terminal trans-membrane regions suggests they have a high degree of inherent insolubility. This insolubility could be caused by regions of the enzyme that interact with the membrane but are encoded some distance from the N-terminal of the polypeptide. The 3D structure and N-terminal modification of cytochrome P450 2C5 have shown that the F-G, A-B and B-C loops of P450 2C5 are also involved in membrane interaction. The N-terminal modified P450 2C5 was shown to interact with *E. coli* cell membranes in a salt-dependent manner. This reversible interaction was thought to be due to the additional membrane binding regions (Williams *et al.*, 2000). It is highly likely that PaxP and PaxQ contain similar regions, which remain in the Δ1-39PaxP and Δ1-29PaxQ constructs. Alignment of PaxP, PaxQ and 2C5 polypeptide sequences revealed that the regions of PaxP and PaxQ which align with the P450 2C5 F-G loop, A-B loop and ß2 sheet are rich in hydrophobic residues. The presence of these hydrophobic regions on the surface of the enzyme could have contributed to the

insolubility encountered during high level expression in *E. coli*. The presence of an N-terminal fusion protein would also prevent the insertion of these regions into a membrane. This situation would increase the probability of hydrophobic regions in the polypeptides aggregating during folding.

4.4 Sequence alignment of PaxP and PaxQ

Comparison of available 3D structures has shown that although cytochrome P450 enzymes can share as little as 15% sequence identity, all have a conserved structural fold which is evolutionarily preserved from bacteria to mammals. The availibility of 3D structures for several cytochrome P450 enzymes means that uncharacterised P450 enzymes can be analysed at a primary sequence level and compared and contrasted with sequences of known structure. Because the sequence similarity between cytochrome P450 enzymes is so variable, a prediction of structural function must be made on the basis of secondary structure as well as sequence conservation. The alignment of PaxP and PaxQ with four related fungal P450s and two structurally determined P450s has highlighted the regions important for substrate binding, redox partner interaction and membrane attachment (Figure 1.10).

4.4.1 Cytochrome P450 secondary structure

The alignment of PaxP and PaxQ with related fungal P450 enzymes has revealed an extremely high degree of conservation among the six fungal enzymes. Almost all elements of secondary structure are conserved across the fungal sequences, exceptions being the C, D and G helices. The G helix is a recognised substrate recognition site and as such can be expected to be highly variable in order to accommodate the varied substrates of the different enzymes. Both C and D helices are exposed on the outer surface of the enzyme, with the C helix involved in redox partner binding. Eukaryotic and prokaryotic P450 enzymes would interact with quite different redox partner enzymes. The diferences between the prokaryotic iron sulphur electron donor and the eukaryotic CPR enzyme would require quite different redox binding sites, and hence divergent sequences in those regions. The fungal enzymes share conserved residues in the C helix, with hydrophobic and positively charged residues most common. These

two characteristics are predicted to be necessary for CPR-P450 interactions (Sevrioukova *et al.*, 1999).

The I helix has a high degree of sequence conservation across all the aligned sequences, particularly at the C-terminal end. An interesting feature of the I helix is the conserved distortion adjacent to the haem group. This distortion has been proposed to be the path of protons required during oxygenation (Ravichandran *et al.*, 1993). After dioxygen is split during catalysis, one atom is incorporated into the substrate and the other is combined with two protons to form water. The groove in the I helix binds a water molecule which would create a proton donating pathway from the solvent exposed Glu 268 to the water molecule, to Thr 269 which could donate the proton to an iron-bound oxygen atom. The Glu 268 residue in P450BM-3 is present in all the fungal sequences as a conserved histidine residue. Histidine can be present in either an uncharged or positively charged state and could act as a proton donor in the same manner as glutamate.

The C helix has been shown to be involved in the interaction of P450BM-3 with the BM-3 flavin domain. This helix is often characterised by the presence of a conserved tryptophan at the N-terminal end and a conserved basic residue four amino acids distant from the tryptophan in the C-terminal direction. Lewis and Hlavica (2000) postulate that the tryptophan residue which occurred in the evolution of eukaryotic P450 systems provided a more tightly regulated electron transfer process. This altered regulation was considered relevant due to the use of an iron-sulphur protein as a redox partner in prokaryotic and mitochondrial systems, in comparison to the more complex CPR enzyme used in eukaryotic microsomal systems. However, this residue is found in PaxQ, but not in PaxP. As both PaxP and PaxQ are expected to bind the same CPR enzyme there would have to be another equally accessible electron donation pathway which PaxP could utilise. The C helix of PaxQ tends to share a greater homology with the P450BM-3 sequence than with the other four fungal sequences. The sequence of PaxP appears to segregate with the other fungal sequences in helix C, and contains a threonine residue in place of the tryptophan. Considering that the other four fungal sequences either contain a valine or isoleucine residue at this position, a hydrophobic protein-protein interaction may be more important than a posssible electron transfer function at this position.

4.4.2 Redox partner interactions

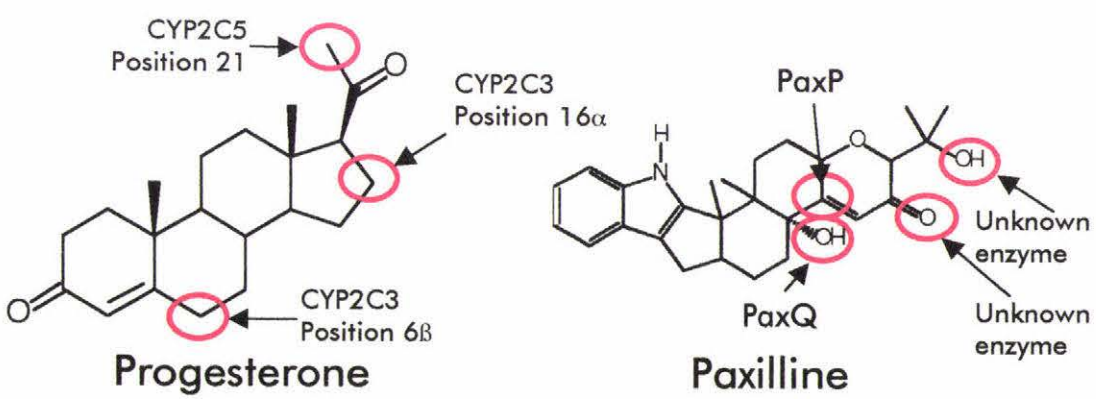
The crystalisation and structural solution of a complex between the BM-3 haem and flavin domain has provided a base from which other P450-redox partner interactions can be modeled. Sevrioukova *et al.* (1999) noted that there was a low number of direct contacts for the interface between the haem and flavin domains. This situation was explained by the fact that P450BM-3 encodes both haem and reductase (FAD/FMN) domains with a linker peptide joining the two enzymatic domains. This structurally enforced close proximity would reduce the requirement for strong binding contacts. Most other P450 systems have a separate reducing enzyme system for the haem protein, thereby necessitating the need for more numerous contacts between the enzymes.

The sequence conservation of the meander region of P450 enzymes is related to the important role it plays in both redox partner binding and electron transfer. The structural determination of the P450BM-3 electron donator/acceptor complex allowed a hypothesis for electron transfer between the FMN and haem groups to be formulated. Sevrioukova *et al.* (1999) suggested that electron transfer would occur between the FMN group and Pro 383. From Pro383 to Gln 388 and then directly to the haem iron through bonded orbitals via the essential cysteine linkage (Cys 401). An alternative path from Gln 388 was suggested which required through-space jumps to Pro 393, Gly 395 and/or Arg 399. These three residues are 3.4, 3.6 and 3.1 Å, respectively, distant from the porphyrin ring of the haem. The flavin group of eukaryotic CPR enzymes has been shown to be buried deeper in the enzyme than that in P450BM-3 (Wang *et al.*, 1997). This altered position means that a rearrangement of the FAD and FMN domains of CPR would be required to make the FMN acessible to the P450 enzyme. In addition to this altered FMN position, the flavin domain of P450BM-3 has a different environmnet surrounding the FMN group. P450BM-3 has mostly neutral and hydrophobic residues around the FMN, while in flavodoxin and the flavin domain of CPR the FMN is surrounded by negatively charged residues (Sevrioukova *et al.*, 1999). This alteration has effects on both protein-protein interactions and the redox properties of the FMN group.

4.4.3 The substrate binding site

The substrate binding site of CYP2C5 can be expected to be similar to that for PaxP and PaxQ. The substrate of CYP2C5, progesterone has a similar structure to that of paxilline (Figure 4.1). The specificity of the 2C5 active site is demonstrated by the alanine residue found at position 113. This position is held by a valine residue in most other 2C familly enzymes. If an alanine is used to replace the valine 113 found in CYP2C3, its activity changes from a 16α and 6ß hydroxylase to a 21 hydroxylase. In addition, the mutation of Phe 205 in CYP2C3 to a valine results in almost total abolishment of position 21 hydroxylation (Figure 4.1). PaxP contains a leucine residue at the corresponding position to Ala 113 in CYP2C5. This results in an extra carbon bond compared to a valine residue. If the relationship between residue size and the position of hydroxylation is correct, oxygenation should occur further towards the middle of a paspaline molecule. The predicted hydroxylation position of PaxP is shown in Figure 4.1. This position is similar to what would be expected from the mutagenesis of position 113 in Cyp2C family enzymes. This residue would be an ideal candidate for a mutagenesis study of PaxP. Two other residues involved in substrate positioning are Val 205 and Leu 208 in CYP2C5.

Fig 4.1 **Comparison of progesterone and paxilline biosynthesis**



Selected sites of hydroxylation are shown for progesterone and paxilline **Red circles** denote regions which are modified by the listed enzymes. The enzyme responsible for (or predicted to perform) the hydroxylation event is listed beside each region. The predicted activities of PaxP and PaxQ are shown in Figure 1.3.

If Val 205 is replaced with a phenylalanine (as found in CYP2C3) hydroxylation at position 21 is almost totally abolished. Mutation of Leu 208 has also been shown to alter the regioselectivity of CYP2C5 (Szklarz *et al.*, 1995). PaxP contains an isoleucine while PaxQ contains a valine at the corresponding position to Val 205. The reciprocal mutation of this position in PaxP and PaxQ could also be used to determine residues important for substrate positioning

Leucine residues located in SRS-5 of CYP2C5 have been implicated in the positioning of substrate hydroxylation. The leucine at position 363 of CYP2C5 could be an important determinant of catalytic specificity . This residue has been shown to alter the regiospecificity of CYP2B enzymes and the position aligns with a conserved hydrophobic or neutral residue in all the fungal sequences.

The hairpin turn associated with beta sheet 4 rests very close to the active site of cytochrome P450 enzymes. Residues present in the turn have also been implicated in the regiospecificity of hydroxylation. Hydrophobic residues are present in all the aligned sequences at the position corresponding to Phe473 in CYP2C5.

4.5 Alternate strategies

The heterologous expression of proteins can be achieved in many different systems other then *E. coli*. Popular systems include *in vitro* translation, yeast species, insect cell cultures, mammalian cell cultures and fungal species.

In vitro systems are often used because of their convenience and speed. The Roche Rapid Translation System 500 (RTS) is a commercial product designed for the *in vitro* synthesis of proteins. It utilises an *E. coli* lysate that is coupled to a feeding chamber which replenishes metabolites and dilutes inhibitory by-products of transcription/translation process. The system is capable of synthesising milligram quantities of protein in 24 hours (Martin *et al.*, 2001). Purification of the synthesised protein is simplified by the use of a cell-free system, negating the need for cell lysis. Use of the RTS 500 system permitted the synthesis of a folding-defective maltose binding protein (MalE31) in a soluble state after it had been previously shown to form inclusion bodies in *E. coli* (Betton and Hofnung, 1996; Betton, 2000). However, the

high cost and recent introduction of the system has limited its application. This system has considerable potential, with futher development and usage extending and defining the limitations of the system.

Eukaryotic expression systems are useful because of their increased polypeptide processing ability compared to prokaryote systems. Proteins expressed by eukaryotic systems can be post-translationally modified or correctly inserted into membranes depending on their particular motifs. Disadvantages of eukaryotic systems include incomplete post-translational modification, which can generate a heterogenous population of the protein of interest. Expression levels in eukaryotic systems are generally lower than that achieved in *E. coli* (Guengerich *et al.*, 1991).

Yeast expression systems are commonly used as a compromise between prokaryotic and eukaryotic expression systems. They offer rapid genetic modification and simple culture conditions while still containing more elaborate protein folding systems. Species often used for heterologous expression studies include *Saccharomyces cerevisiae, Pichia pastoris* and *Schizosaccharomyces pombe.* Expression in *P. pastoris* is driven by robust and highly regulated promoters isolated from genes required for methanol metabolism (Gellissen, 2000). Good vector systems using the methanol oxidase (AOX1) promoter are available such as the pPICZ vectors availible from Invitrogen Life Technologies. *P. pastoris* expression cultures can be easily scaled from one litre cultures up to commercial scale preparations. Disadvantages include, unpredictable expression levels and the over-glycosylation of proteins. Excess glycosylation will alter the mass of the protein and could alter its activity. Some examples of proteins successfully expressed in yeast are listed below. Geraniol 10-hydroxylase, a plant cytochrome P450 monooxygenase involved in terpenoid indole alkaloid biosynthesis has been successfully expressed and purified from *S. cerevisiae* (Collu *et al.*, 2001). *Ent*-kaurenoic acid oxidase (CYP88A) is a cytochrome P450 isolated from *Arabidopsis thaliana* which catalyses the monooxygenation of diterpenoids required for gibberellin synthesis. CYP88A was expressed in *S. cerevisiae*, and functionally characterised *in vivo,* by making use of the endogenous *S. cerevisiae* cytochrome P450 reductase (CPR) (Helliwell *et al.*, 2001). However, there are cases where eukaryotic expression has been problematic. Of the thirteen plant cytochrome P450 cDNA clones isolated from *Taxus cuspidata* and expressed in

S. cerevisiae, only eight were functional (Schoendorf *et al.*, 2001). The clones which were expressed as non-functional enzymes in *S. cerevisiae* were subsequently expressed in *Spodoptera fugiperda,* the insect-baculovirus-based system, with 100% success (Jennewein *et al.*, 2001). These results suggest that yeast and baculovirus expression systems may be better suited to the expression of functionally active PaxP and PaxQ.

An alternative approach would be expression of these two P450s in *P. paxilli.* A protocol for reliable integrative transformation has been established for *P. paxilli.* Expression under the control of native promoters would ensure high level expression under paxilline inducing conditions. *Aspergillus* cytoskeletal proteins, have been expressed from an autonomously replicating plasmid and sucessfully purified from *Aspergillus nidulans* (Efimov and Morris, 2000). The close taxonomic relationship between *Penicillium* and *Aspergillus* species could allow the direct usage of the *A. nidulans* expression vector in *P. paxilli.* Alternatively an integrative vector could be developed that utilises a strong native promoter such as that found in *paxM* (Figure 1.4). Sucha construct would result in high levels of expression of the transformed genes under conditions that induce paxilline synthesis.

4.5.1 Functional analysis of PaxP and PaxQ

Any protein synthesised by a heterologous expression system requires the presence of a cytochrome P450 reductase (CPR) if the activity of the purified enzyme is to be determined. The CPR in yeast species has been shown to reduce plant and mammalian P450 enzymes (Guengerich *et al.*, 1991; Helliwell *et al.*, 2001). Yeast like most eukaryotes has only one CPR which reduces all P450s expressed in the cell (Degtyarenko, 1995). Consequently the CPR binding site is highly conserved.

Another approach to the functional characterisation of cytochrome P450 enzymes has been to transform one member of a gene cluster into a mutant strain in which the gene cluster has been deleted. The mutant strain is then fed radiolabelled substrates and the products of the catalysis identified. This method has been used for the functional analysis of *G. fujikuroi* gibberellin biosynthetic enzymes P450-1 (Rojas *et al.*, 2001) and P450-4 (Tudzynski *et al.*, 2001). P450-1 is involved in the hydroxylation of *ent*-kaurenoic acid while P450-4 hydroxylates *ent*-kaurene. Recombinant strains of

G. fujikuroi lacking the gibberellin gene cluster were transformed with the gene of interest and cultured on media containing a radiolabeled substrate. Expression of P450-1 and P450-4 resulted in the conversion of the substrate to products that were identified by gas and liquid chromatography in conjunction with mass spectrophotometric analysis. Interestingly, multiple products were identified for both P450-1 and P450-4 suggesting that these enzymes catalyse multiple biosynthetic steps. The ability of these enzymes to catalyse several steps in a single pathway is an emerging theme for cytochrome P450 enzymes. Both *ent*-kaurene oxidase enzymes from *A. thaliana* and *G. fujikuroi* catalyse three successive oxidations of the same methyl group (Helliwell *et al.*, 1999; Tudzynski *et al.*, 2001). The two *ent*-kaurenoic acid oxidases from *A. thaliana* and *G. fujikuroi* have also been shown to catalyse three and four steps respectively (Helliwell *et al.*, 2001; Rojas *et al.*, 2001). The *ent*-kaurenoic acid oxidase P450-1 from *G fujikuroi* was also implicated in the formation of up to 12 different products, involving the multiple hydroxylation of four different positions on the gibberellin substrate. Among the range of reactions catalysed by cytochrome P450s is the demethylation of substrates. Lanosterol 14α-demethylase is a *S. cerevisiae* cytochrome P450 enzyme that demethylates lanosterol after three sucessive hydroxylations. Lanosterol is also a substrate of CYP51 as seen in Figure 1.7, part H. Each successive product during the triple oxidation process showed a higher affinity between it and the enzyme. This means the substrate may be held in the active site throughout the demethylation process (Aoyama *et al.*, 1989). A similar demethylation activity is shown by the aromatase cytochrome P450, which catalyses the conversion of androstenedione to estrone (Graham-Lorence *et al.*, 1995). P450arom catalyses three successive hydroxylations that result in the release of the targeted methyl group as formic acid, a very similar reaction to that catalysed by lanosterol demethylase.

The use of these approaches would be the logical way forward to identifying the substrates and products of the enzymes responsible for paxilline biosynthesis. The results described above highlight the importance of using heterologus expression systems such as those used for taxol and gibberellin research. The availability of deletions of the paxilline biosynthesis cluster indicates that the radiolabeled substrate feeding approach used for gibberellin reasearch could be applied to *P. paxilli*. Mutants

lacking the paxilline gene cluster could be transformed with individual genes followed by the identification of products formed from specific substrates. However, unlike gibberellin biosynthesis many of the substrates and products are not yet known and radiolabeled intermediates are not yet available. Deletion of *paxP* and *paxQ* results in greatly reduced paxilline production and the accumulation of paspaline and 13-desoxypaxilline respectively. These two mutants will be ideal systems for the overproduction of these radiolabeled intermediates.

While the deletion of *paxG*, *paxM* or *paxC* leads to a paxilline negative phenotype, there is no accompanying build up of indole-diterpenoid intermediates. This means that these three enzymes are all essential for paxilline biosynthesis and are likely to act before or be responsible for the covalent linkage of the indole group with GGPP. In light of the gibberellin research described above, the synthesis of paxilline from paspaline may require fewer enzymes than initially thought. It now seems possible that PaxP could catalyse multiple hydroxylations of paspaline. Previously, it was thought that paspaline B was the only product of PaxP. However, given that the position 30 methyl group (see Figure 1.3) present in paspaline is missing from paxilline, PaxP is now predicted to act as a demethylase, removing this CH_3 group from paspaline. The putative substrate of PaxQ (13-desoxypaxilline, as indicated by the *paxQ* deletion mutant) requires only a single hydroxyl group to be added to form paxilline. It is likely therefore that the conversion of paspaline to paxilline may involve just PaxP and PaxQ.

The creation of a ketone group located at position 10 of the paxilline molecule is the only other modification to paspaline required in order to synthesise paxilline. There is currently no candidate gene or gene product that is proposed to perform this modification. The ketone group could be created by a monooxygenase and a dehydrogenase, or by two sequential hydroxylations. Among the other, as yet uncharacterised, genes in the vicinity of the paxilline biosynthetic cluster are a dehydrogenase (*paxH*) and a monooxygenase (*paxN*) (Young *et al.*, 2001). The functions encoded by these two open reading frames are predicted from observed sequence similarity with other proteins. These two enzymes in concert with PaxP and PaxQ may be involved in the conversion of paspaline to paxilline. However, the "promiscuity" of monooxygenase enzymes such as P450-1 from *G. fujikuroi* suggest

that it is possible that PaxP and PaxQ complete all the required steps from paspaline onwards.

To determine the steps catalysed by PaxP and PaxQ functional enzymatic analysis of the enzymes could be performed *in vitro* or *in vivo*. Purified haem oxygenase (HemO) (Zhu *et al.*, 2000) and lanosterol 14α-demethylase (Aoyama *et al.*, 1984) are examples of haem-proteins which can be reconstituted into an active *in vitro* system upon the addition of a CPR enzyme, NADPH and phospholipid. If enzyme can be successfully expressed in a soluble and active form, *in vitro* analysis would be the most straightforward method of study. If heterologous expression and purification cannot be performed, *in vivo* studies in *P. paxilli* would be an alternative strategy. One of the limitations of these approaches is sufficient uptake of the radiolabeled substrates into the fungal cell. The precursors of GGPP are highly charged isoprenoid molecules and would not be expected to readily traverse a lipid bilayer and enter the cell (Figure 1.2). However, the predicted substrates of PaxP and PaxQ are less polar and might be more readily taken up by the cell (Figure 1.3). *P. paxilli* protoplasts, which lack a fungal outer cell wall, may be more efficient at taking up substrate into the cytoplasm. If the introduction of substrates into living cells was inefficient, crude cell extracts or microsomal preparations could be prepared. Alternatively, provision of substrates in the presence of membrane-permeabilisation agents, such as charged detergents, could increase their uptake into the cell (Heerklotz, 2001).

4.6 Summary

The expression of PaxP and PaxQ in *E. coli* has shown that prokaryotic systems are not suitable for the synthesis of soluble active forms of these enzymes. Attempts to solubilise these proteins in *E. coli* using detergent and N-terminal deletions were unsuccessful. In the light of these results future experiments should be carried out in a eukaryotic system with the aim of expressing and purifying the enzyme. Alternatively, proposed substrates could be radiolabeled and introduced to deletion mutants of *P. paxilli* in order to identify the products of catalysis. Recent studies on the roles of P450 enzymes in the biosynthesis of gibberellin in both *G. fujikuroi* and *A. thaliana* suggest that the synthesis of paxilline may require fewer biosynthetic enzymes than was first proposed (Young *et al.*, 2001). As proposed in Figure 1.3 as few as five enzymes

may be required for the conversion of IPP to paxilline. The availibility of gene deletions of *paxG, M, C, P* and *Q* provide the tools for *in vivo* dissection of this pathway. Use of eukaryotic expression systems such as yeast and baculovirus-insect cells will be the preferred systems for overexpression of these enzymes for *in vitro* catalytic studies.

Chapter 5

Bibliography

Aiyar, A., and Leis, J. (1993) Modification of the megaprimer method of PCR mutagenesis: improved amplification of the final product. *Biotechniques* **14**: 366-369.

Aoyama, Y., Yoshida, Y., and Sato, R. (1984) Yeast cytochrome P-450 catalyzing lanosterol 14 alpha-demethylation. II. Lanosterol metabolism by purified P-450(14)DM and by intact microsomes. *J Biol Chem* **259**: 1661-1666.

Aoyama, Y., Yoshida, Y., Sonoda, Y., and Sato, Y. (1989) Deformylation of 32-oxo-24,25-dihydrolanosterol by the purified cytochrome P-45014DM (lanosterol 14 alpha-demethylase) from yeast evidence confirming the intermediate step of lanosterol 14 alpha-demethylation. *J Biol Chem* **264**: 18502-18505.

Belofsky, G.N., Gloer, J.B., Wicklow, D.T., and Dowd, P.F. (1995) Antiinsectan alkaloids: shearinines A-C and a new paxilline derivative from the ascostromata of *Eupenicillium shearii*. *Tetrahedron Letts* **51**: 3959-3968.

Betton, J., and Hofnung, M. (1996) Folding of a mutant maltose-binding protein of *Escherichia coli* which forms inclusion bodies. *J Biol Chem* **271**: 8046-8052.

Betton, J.-M. (2000) Production of a soluble and active form of a defective folding protein with the rapid translation system. *Biochemica* **4**: 7-9.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987) XL1-Blue: A high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5**: 376-378.

Byrne, K.M., Ondeyka, J., and Smith, S.K. (2000) Biosynthesis of nodulosporic acid, precursor studies. In *Genetics and molecular biology of industrial microorganisms conference*. Vol. P4 Indiana university, Bloomington, IN, pp. 24.

Chappell, J. (1995) The biochemistry and molecular biology of isoprenoid metabolism. *Plant Physio* **107**: 1-6.

Cole, R., Kirksey, J., and Wells, J. (1974) A new tremorgenic metabolite from *Penicillium paxilli*. *Can J Microbiol* **20**: 1159-1162.

Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R., and Memelink, J. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* **508**: 215-220.

Cosme, J., and Johnson, E.F. (2000) Engineering microsomal cytocrome P450 2C5 to be a soluble, monomeric enzyme. *J Biol Chem* **275**: 2545-2553.

Cupp-Vickery, J.R., and Poulos, T.L. (1995) Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nat Struct Biol* **2**: 144-153.

de Jesus, A.E., Gorst-Allman, C.P., Steyn, P.S., van Heerden, F.R., Vleggar, R., Wessels, P.L., and Hull, W.E. (1983) Tremorgenic mycotoxins from *Penicillium crustosum*. Biosynthesis of Penitrem A. *J Chem Soc. Perkin Trans*: 1863-1868.

Degtyarenko, K.N. (1995) Structural domains of P450-containing monooxygenase systems. *Protein Eng* **8**: 737-747.

Del Tito, B.J., Ward, J.M., Hodgson, J., Gershater, C.J.L., Edwards, H., Wysocki, L.A., Watson, F.A., Sathe, G., and Kane, J.F. (1995) Effects of minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*. *J Bacteriol* **177**: 7086-7091.

Dorner, J., Cole, R., Cox, R., and Cunfer, B. (1984) Paspalitrem C, a new metabolite from sclerotia of *Claviceps paspali*. *J Ag Food Chem* **32**: 1069-1071.

Efimov, V.P., and Morris, N.R. (2000) The LIS1-related NUDF protein of *Aspergillus nidulans* interacts with the coiled-coil domain of the NUDE/RO11 protein. *J Cell Bio* **150**: 681-688.

Ellis, J. (1987) Proteins as molecular chaperones. *Nature* **328**: 378-379.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem* **15**: 321-353.

Ewalt, K.L., Hendrick, J.P., Houry, W.A., and Hartl, F.U. (1997) *In vivo* observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90**: 491-500.

Frangioni, J.V., and Neel, B.G. (1993) Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal Biochem* **210**: 179-187.

Gallagher, R., Latch, G., and Keogh, R. (1980) The janthitrems: fluorescent tremorgenic toxins produced by *Penicillium janthinellum* isolates from ryegrass pastures. *Appl Environ Microbiol* **39**: 272-273.

Gallagher, R.T., Hawkes, A.D., Steyn, P.S., and Vleggaar, R. (1984) Tremorgenic neurotoxins from perennial ryegrass causing ryegrass staggers disorder of livestock: structure elucidation of lolitrem B. *J Chem Soc. Chem Comm*: 614-616.

Gatenby, W.A., Munday-Finch, S.C., Wilkins, A.L., and Miles, C.O. (1999) Terpendole M, a novel indole-diterpenoid isolated from *Lolium perenne* infected with the endophytic fungus *Neotyphodium lolii*. *J Agric Food Chem* **47**: 1092-1097.

Gellissen, G. (2000) Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* **54**: 741-750.

Goenka, S., and Rao, C.M. (2001) Expression of recombinant zeta-crystallin in *Escherichia coli* with the help of GroEL/ES and its purification. *Protein Expr Purif* **21**: 260-267.

Goloubinoff, P., Gatenby, A.A., and Lorimer, G.H. (1989) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**: 44-47.

Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267**: 83-90.

Graham-Lorence, S., Amarneh, B., White, R.E., Peterson, J.A., and Simpson, E.R. (1995) A three-dimensional model of aromatase cytochrome P450. *Protein Sci* **4**: 1065-1080.

Graham-Lorence, S.E., and Peterson, J.A. (1996) Structural alignments of P450s and extrapolations to the unknown. *Methods Enzymol* **272**: 315-326.

Gribskov, M., Burgess, R.R., and Devereux, J. (1986) PEPPLOT, a protein secondary structure analysis program for the UWGCG sequence analysis software package. *Nucleic Acids Res* **14**: 327-334.

Guengerich, F.P., Brian, W.R., Sari, M.A., and Ross, J.T. (1991) Expression of mammalian cytochrome P450 enzymes using yeast-based vectors. *Methods Enzymol* **206**: 130-145.

Hasemann, C.A., Ravichandran, K.G., Peterson, J.A., and Deisenhofer, J. (1994) Crystal structure and refinement of cytochrome P450terp at 2.3 A resolution. *J Mol Biol* **236**: 1169-1185.

Heerklotz, H. (2001) Membrane stress and permeabilization induced by asymmetric incorporation of compounds. *Biophysics J* **81**: 184-195.

Helliwell, C.A., Poole, A., Peacock, W.J., and Dennis, E.S. (1999) *Arabidopsis ent*-kaurene oxidase catalyzes three steps of gibberellin biosynthesis. *Plant Physiol* **119**: 507-510.

Helliwell, C.A., Chandler, P.M., Poole, A., Dennis, E.S., and Peacock, W.J. (2001) The CYP88A cytochrome P450, *ent*-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proc Natl Acad Sci USA* **98**: 2065-2070.

Hénaut, A., and Danchin, A. (1996) Analysis and predictions from *Escherichia coli* sequences, or *E. coli in silico*. In *Escherichia coli and Salmonella typhimurium cellular and molecular biology*. Vol. 2. Neidhardt, F., Curtiss III, R., Ingraham, J.,

Lin, E., Brooks Low, K., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. and Umbarger, H. (eds). Washington, D.C: American Society for Microbiology, pp. 2047-2066.

Holmgren, A. (1985) Thioredoxin. *Annu Rev Biochem* **54**: 237-271.

Huang, K.X., Huang, Q.L., Wildung, M.R., Croteau, R., and Scott, A.I. (1998) Overproduction, in *Escherichia coli*, of soluble taxadiene synthase, a key enzyme in the Taxol biosynthetic pathway. *Protein Expr Purif* **13**: 90-96.

Huang, X.-H., Tomoda, H., Nishida, H., Masuma, R., and Omura, S. (1995) Terpendoles, novel ACAT inhibitors produced by *Albophoma yamanashiensis*. I. Production, isolation and biological properties. *J Antibiot* **48**: 1-4.

Ichinose, H., Wariishi, H., and Tanaka, H. (2002) Identification and characterization of novel cytochrome P450 genes from the white-rot basidiomycete, *Coriolus versicolor*. *Appl Microbiol Biotechnol* **58**: 97-105.

Imamura, H., Jeon, B., Wakagi, T., and Matsuzawa, H. (1999) High level expression of *Thermococcus litoralis* 4-alpha-glucanotransferase in a soluble form in *Escherichia coli* with a novel expression system involving minor arginine tRNAs and GroELS. *FEBS Lett* **457**: 393-396.

Jennewein, S., Rithner, C.D., Williams, R.M., and Croteau, R.B. (2001) Taxol biosynthesis: taxane 13 alpha-hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc Natl Acad Sci USA* **98**: 13595-13600.

Kasper, C.B. (1971) Biochemical distinctions between the nuclear and microsomal membranes from rat hepatocytes. *J Biol Chem* **246**: 577-581.

Kusmierczyk, A.R., and Martin, J. (2001) Chaperonins--keeping a lid on folding proteins. *FEBS Lett* **505**: 343-347.

Kyte, J., and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**: 105-132.

Laakso, J.A., Gloer, J.B., Wicklow, D.T., and Dowd, P.F. (1992) Sulpinines A-C and secopenitrem B: new antiinsectan metabolites from the sclerotia of *Aspergillus sulphureus*. *J Organic Chem* **57**: 2066-2071.

Lange, B.M., Rujan, T., Martin, W., and Croteau, R. (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci USA* **97**: 13172-13177.

Larson, J.R., Coon, M.J., and Porter, T.D. (1991) Purification and properties of a shortened form of cytochrome P-450 2E1: deletion of the NH2-terminal

membrane-insertion signal peptide does not alter the catalytic activities. *Proc Natl Acad Sci USA* **88**: 9141-9145.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCoy, J.M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *Escherichia coli* cytoplasm. *Biotechnology* **11**: 187-193.

Laws, I., and Mantle, P.G. (1989) Experimental constraints in the study of the biosynthesis of indole alkaloids in fungi. *J Gen Microbiol* **135**: 2679-2692.

Lee, S.C., and Olins, P.O. (1992) Effect of overproduction of heat shock chaperones GroESL and DnaK on human procollagenase production in *Escherichia coli*. *J Biol Chem* **267**: 2849-2852.

Lewis, D.F., and Hlavica, P. (2000) Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects. *Biochim Biophys Acta* **1460**: 353-374.

Li, Y.C., and Chiang, J.Y. (1991) The expression of a catalytically active cholesterol 7 alpha-hydroxylase cytochrome P450 in *Escherichia coli*. *J Biol Chem* **266**: 19186-19191.

Lilie, H., Schwarz, E., and Rudolph, R. (1998) Advances in refolding of proteins produced in *E. coli*. *Curr Opin Biotechnol* **9**: 497-501.

Mantle, P.G., and Weedon, C.M. (1994) Biosynthesis and transformation of tremorgenic indole-diterpenoids by *Penicillium paxilli* and *Acremonium lolii*. *Phytochemistry* **36**: 1209-1217.

Martin, G.A., Kawaguchi, R., Lam, Y., DeGiovanni, A., Fukushima, M., and Mutter, W. (2001) High-yield, *in vitro* protein expression using a continuous-exchange, coupled transcription/ translation system. *Biotechniques* **31**: 948-950, 952-943.

Masters, B.S., and Okita, R.T. (1980) The history, properties, and function of NADPH-cytochrome P-450 reductase. *Pharmacol Ther* **9**: 227-244.

McNally, E., Sohn, R., Frankel, S., and Leinwand, L. (1991) Expression of myosin and actin in *Escherichia coli*. *Methods Enzymol* **196**: 368-389.

Munday-Finch, S.C., Wilkins, A.L., and Miles, C.O. (1996) Isolation of paspaline B, an indole -diterpenoid from *Penicillium paxilli*. *Phytochemistry* **41**: 327-332.

Needleman, S.B., and Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Bio* **48**: 443-453.

Ohnuma, S., Nakazawa, T., Hemmi, H., Hallberg, A.-M., Koyama, T., Ogura, K., and Nishino, T. (1996) Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis. *J Biol Chem* **271**: 10087-10095.

Omura, T., and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. *J Biol Chem* **239**: 2370-2378.

Ortiz de Montellano, P.R. (1987) *Cytochrome P-450: structure, mechanism and biology*. New York: Plenum press.

Park, S.Y., Shimizu, H., Adachi, S., Nakagawa, A., Tanaka, I., Nakahara, K., Shoun, H., Obayashi, E., Nakamura, H., Iizuka, T., and Shiro, Y. (1997) Crystal structure of nitric oxide reductase from denitrifying fungus *Fusarium oxysporum*. *Nat Struct Biol* **4**: 827-832.

Pendurthi, U.R., Lamb, J.G., Nguyen, N., Johnson, E.F., and Tukey, R.H. (1990) Characterization of the CYP2C5 gene in 21L III/J rabbits. Allelic variation affects the expression of P450IIC5. *J Biol Chem* **265**: 14662-14668.

Penn, J., and Mantle, P.G. (1994) Biosynthetic intermediates of indole-diterpenoid mycotoxins from selected transformations at C-10 of paxilline. *Phytochemistry* **35**: 921-926.

Pernecky, S.J., and Coon, M.J. (1996) N-terminal modifications that alter P450 membrane targeting and function. *Methods Enzymol* **272**: 25-34.

Peterson, J.A., and Graham, S.E. (1998) A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure* **6**: 1079-1085.

Podust, L.M., Poulos, T.L., and Waterman, M.R. (2001) Crystal structure of cytochrome P450 14-alpha-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. *Proc Natl Acad Sci USA* **98**: 3068-3073.

Porter, T.D., and Kasper, C.B. (1986) NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. *Biochemistry* **25**: 1682-1687.

Porter, T.D., and Coon, M.J. (1991) Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J Biol Chem* **266**: 13469-13472.

Poulos, T.L., Finzel, B.C., and Howard, A.J. (1987) High-resolution crystal structure of cytochrome P450cam. *J Mol Biol* **195**: 687-700.

Poulos, T.L. (1995) Cytochrome P450. *Curr Opin Struct Biol* **5**: 767-774.

Prinz, W.A., Aslund, F., Holmgren, A., and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J Biol Chem* **272**: 15661-15667.

Ravichandran, K.G., Boddupalli, S.S., Hasermann, C.A., Peterson, J.A., and Deisenhofer, J. (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **261**: 731-736.

Rojas, M.C., Hedden, P., Gaskin, P., and Tudzynski, B. (2001) The P450-1 gene of *Gibberella fujikuroi* encodes a multifunctional enzyme in gibberellin biosynthesis. *Proc Natl Acad Sci USA* **98**: 5838-5843.

Rudolph, R. (1990) Renaturation of recombinant, disulfide-bonded proteins from inclusion bodies. In *Modern methods in protein and nucleic acid research*. Tschesche, H. (ed). New York: Walter de Gryter, pp. 149-172.

Rudolph, R., and Lilie, H. (1996) *In vitro* folding of inclusion body proteins. *FASEB J* **10**: 49-56.

Sandhu, P., Baba, T., and Guengerich, F.P. (1993) Expression of modified cytochrome P450 2C10 (2C9) in *Escherichia coli*, purification, and reconstitution of catalytic activity. *Arch Biochem Biophys* **306**: 443-450.

Sandmann, G., Misawa, N., Wiedemann, M., Vittorioso, P., Carattoli, A., Morelli, G., and Macino, G. (1993) Functional identification of *al-3* from *Neurospora crassa* as the gene for geranylgeranyl pyrophosphate synthase by complementation with *crt* genes, *in vitro* characterization of the gene product and mutant analysis. *J Photochem Photobiol* **18**: 245-251.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467.

Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *Bio/Technology* **7**: 1141-1149.

Schoendorf, A., Rithner, C.D., Williams, R.M., and Croteau, R.B. (2001) Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from Taxus and functional expression in yeast. *Proc Natl Acad Sci USA* **98**: 1501-1506.

Sevrioukova, I.F., Li, H., Zhang, H., Peterson, J.A., and Poulos, T.L. (1999) Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci USA* **96**: 1863-1868.

Sherman, M., and Goldberg, A.L. (1992) Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Escherichia coli*. *EMBO J* **11**: 71-77.

Shi, J., Blundell, T.L., and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* **310**: 243-257.

St John, R.J., Carpenter, J.F., and Randolph, T.W. (1999) High pressure fosters protein refolding from aggregates at high concentrations. *Proc Natl Acad Sci USA* **96**: 13029-13033.

Stewart, E.J., Aslund, F., and Beckwith, J. (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J* **17**: 5543-5550.

Steyn, P.S., and Vleggaar, R. (1985) Tremorgenic mycotoxins. *Prog Chem Organic Natural Products* **48**: 1-80.

Stryer, L. (1995) *Biochemistry*. New York: W. H. Freeman and company.

Sueyoshi, T., Park, L.J., Moore, R., Juvonen, R.O., and Negishi, M. (1995) Molecular engineering of microsomal P450 2a-4 to a stable, water-soluble enzyme. *Arch Biochem Biophys* **322**: 265-271.

Szczesna-Skorupa, E., Browne, N., Mead, D., and Kemper, B. (1988) Positive charges at the NH2 terminus convert the membrane-anchor signal peptide of cytochrome P-450 to a secretory signal peptide. *Proc Natl Acad Sci USA* **85**: 738-742.

Szklarz, G.D., He, Y.A., and Halpert, J.R. (1995) Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. *Biochemistry* **34**: 14312-14322.

Telfer, E.J. (2000) Regulation of paxilline biosynthesis in *Penicillium paxilli*. In *Institute of Molecular Biosciences* Palmerston North: Massey University, pp. 257.

Thornton, J.M. (1981) Disulphide bridges in globular proteins. *J Mol Biol* **151**: 261-287.

Thuring, R.W.J., Sanders, J.P.M., and Borst, P. (1975) A freeze-squeeze method for recovering long DNA from agarose gels. *Anal Biochem* **66**: 213-220.

Tomoda, H., Tabata, N., Yang, D., Takayanagi, H., and Omura, S. (1995) Terpendoles, novel ACAT inhibitors produced by *Albophoma yamanashiensis*. III. Production, isolation and structure elucidation of new components. *J Antibiot* **48**: 793-804.

Tudzynski, B., and Hölter, K. (1998) Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster. *Fungal Genet Biol* **25**: 157-170.

Tudzynski, B., Hedden, P., Carrera, E., and Gaskin, P. (2001) The P450-4 gene of *Gibberella fujikuroi* encodes *ent*-kaurene oxidase in the gibberellin biosynthesis pathway. *Appl Environ Microbiol* **67**: 3514-3522.

van den Brink, H.M., van Gorcom, R.F., van den Hondel, C.A., and Punt, P.J. (1998) Cytochrome P450 enzyme systems in fungi. *Fungal Genet Biol* **23**: 1-17.

Vieira, J., and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods Enzymol* **153**: 3-11.

Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S., and Kim, J.J. (1997) Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci USA* **94**: 8411-8416.

Weedon, C.M., and Mantle, P.G. (1987) Paxilline biosynthesis by *Acremonium loliae*; a step towards defining the origin of lolitrem neurotoxins. *Phytochemistry* **26**: 969-971.

Wiedemann, M., Misawa, N., and Sandmann, G. (1993) Purification and enzymatic characterization of the geranylgeranyl pyrophosphate synthase from *Erwinia uredovora* after expression in *Escherichia coli*. *Arch Biochem Biophys* **306**: 152-157.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **5**: 121-131.

Yabusaki, Y. (1995) Artificial P450/reductase fusion enzymes: what can we learn from their structures? *Biochimie* **77**: 594-603.

Yano, J.K., Koo, L.S., Schuller, D.J., Li, H., Ortiz de Montellano, P.R., and Poulos, T.L. (2000) Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* **275**: 31086-31092.

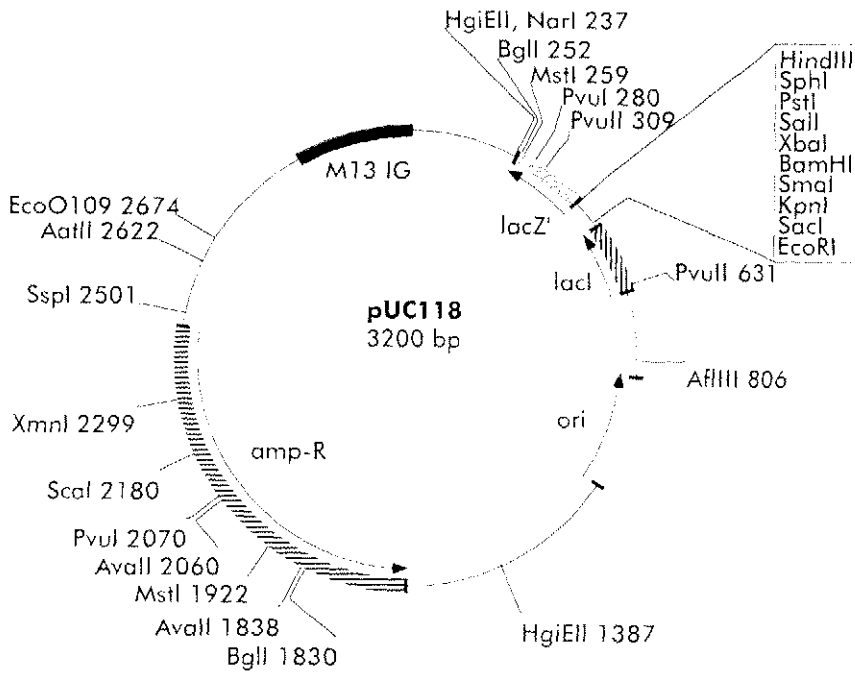
Young, C., McMillian, L., Telfer, E., and Scott, B. (2001) Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. *Mol Microbiol* **39**: 1-13.

Zhu, W., Wilks, A., and Stojiljkovic, I. (2000) Degradation of heme in gram-negative bacteria: the product of the *hemO* gene of *neisseriae* is a heme oxygenase. *J Bacteriol* **182**: 6783-6790.

Chapter 6 Appendix

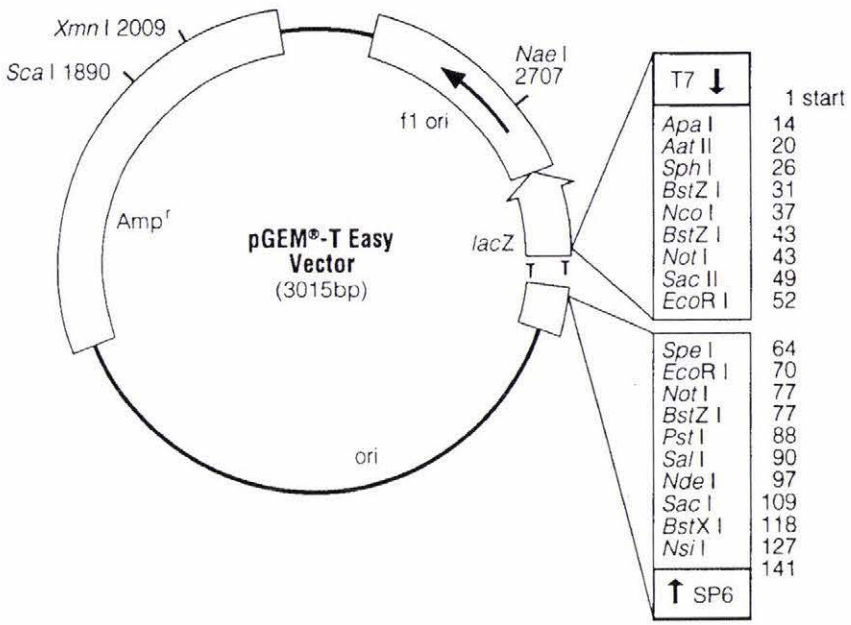
6.1 Vector maps

6.1.1 pUC118



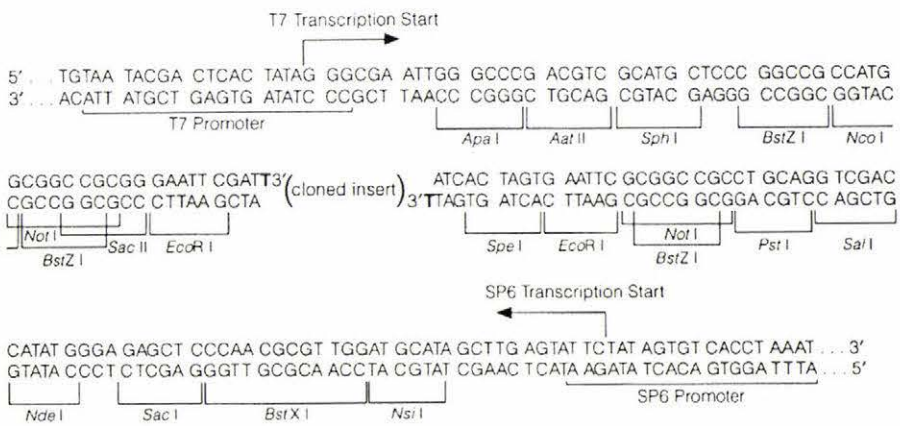
Plasmid name: pUC118
Plasmid size: 3200 bp
Constructed by: Vieira and Messing
Construction date: 1987
Comments/References: Vieira J. and Messing J., 1987. Production of single-stranded plasmid DNA. Methods Enzymol. 153. p 3

6.1.2 pGEM-T-Easy



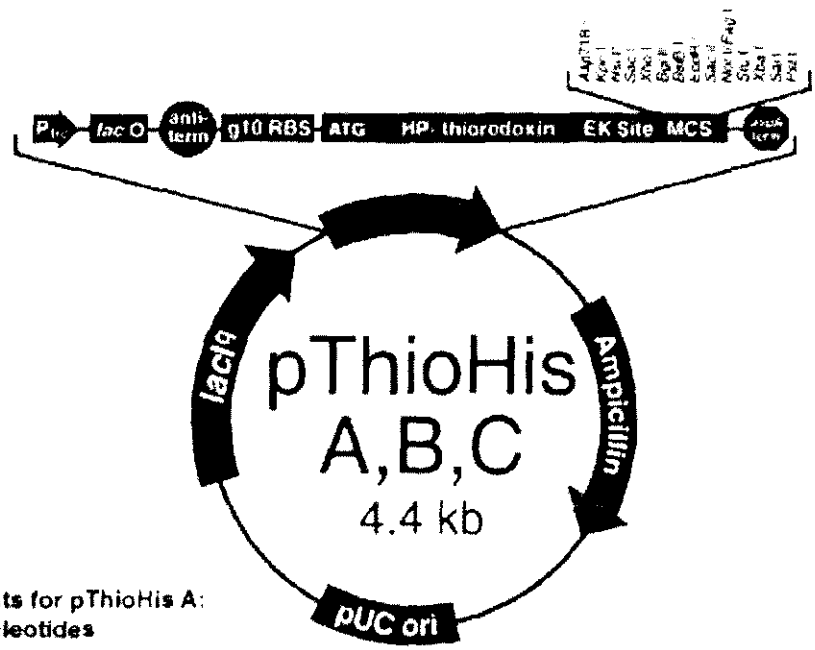
T7 RNA Polymerase transcription initiation site	1
SP6 RNA Polymerase transcription initiation site	141
T7 RNA Polymerase promoter (–17 to +3)	2999–3
SP6 RNA Polymerase promoter (–17 to +3)	139–158
multiple cloning region	10–128
lacZ start codon	180
lac operon sequences	2836–2996, 166–395
lac operator	200–216
β-lactamase coding region	1337–2197
phage f1 region	2380–2835
binding site of pUC/M13 Forward Sequencing Primer	2956–2972
binding site of pUC/M13 Reverse Sequencing Primer	176–192

pGEM-T-Easy Multiple cloning site



Images courtsey of Promega corp.

6.1.3 pThioHis A



Comments for pThioHis A:
4365 nucleotides

Ampicillin resistance ORF:	bases 201-1061
pUC origin:	bases 1206-1879
Lac Repressor (*lacI*^{q}) ORF:	bases 1967-3049
Trc Promotor Region:	bases 3261-3470
Thioredoxin ATG:	bases 3471-3473
Trx forward priming site:	bases 3777-3794
Enterokinase site:	bases 3813-3827
Multiple cloning site:	bases 3827-3892
aspA termination:	bases 3893-3959
Trx reverse priming site:	bases 3967-3986

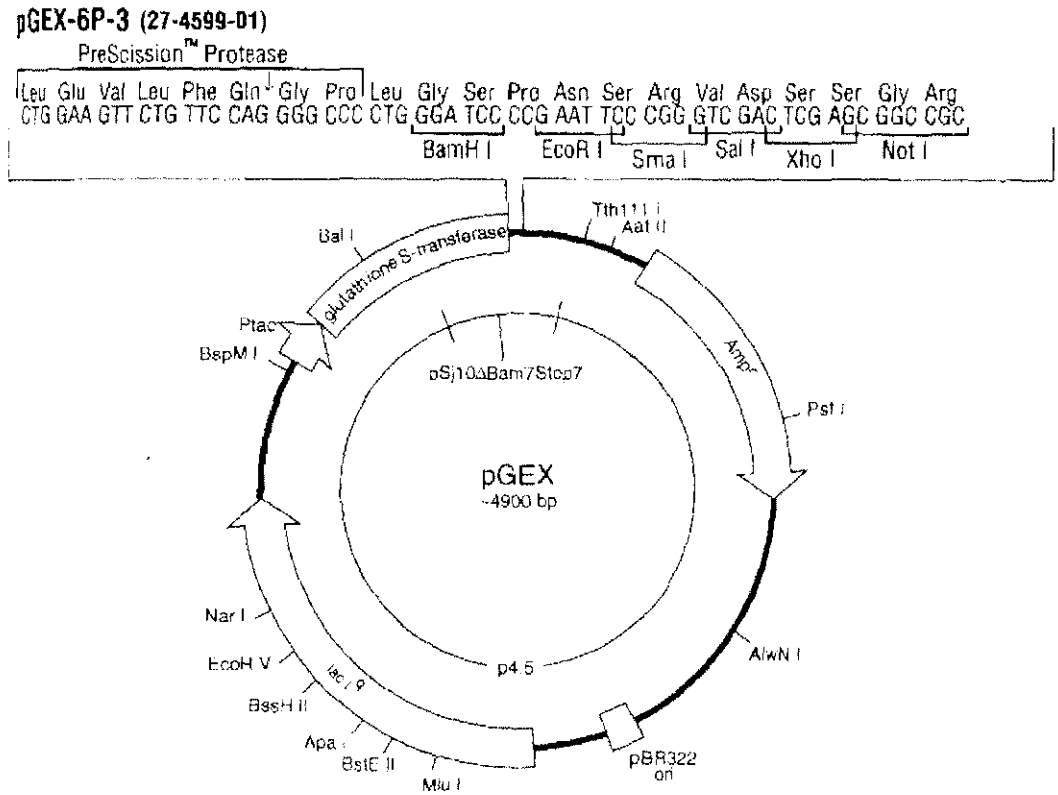
pThioHis A multiple cloning site

```
                                        Trx forward sequencing priming site
3756  TCT AAA GGT CAG TTG AAA GAG TTC CTC GAC GCT AAC CTG GCC GGC TCT
      Ser Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser

                  Enterokinase recognition sequence  Asp718 I  Kpn I    Not I      Xho I Sac I  Bgl II
3804  GGA TCC GGT GAT GAC GAT GAC AAG GTA CCT ATG CAT GAG CTC GAG ATC
      Gly Ser Gly Asp Asp Asp Asp Lys Val Pro Met His Glu Leu Glu Ile
                                      ▲ Enterokinase cleavage site
      BstB I  EcoR I      Sac II  Not I        Stu I  Xba I      Sal I              Pst I
3852  TTC GAA TTC CGC GGC CGC AGG CCT CTA GAG TCG ACC TGC AGT AAT CGT
      Phe Glu Phe Arg Gly Arg Arg Pro Leu Glu Ser Thr Cys Ser Asn Arg
                                  aspA transcriptional terminator
3900  ACA GGG TAG TACAAATA AAAAAGGCAC GTCAGATGAC GTGCCTTTTT TCTTGTGAGC
      Thr Gly ***
                  Trx reverse sequencing priming site
3957  AGTAAGCTTG GCACTGGCCGT CGTTTTACAA CGTCGTGACT GGGAAAA
```

Images courtesy of Invitrogen Life technologies

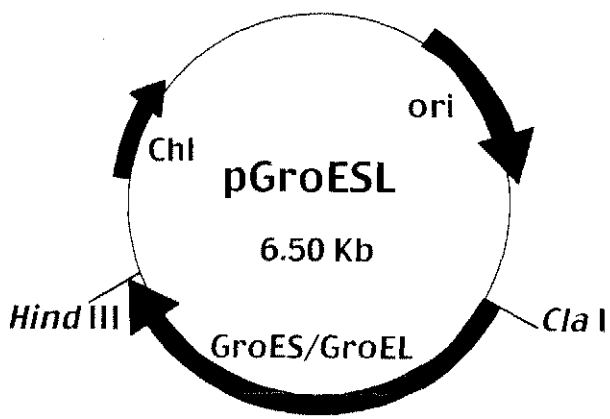
6.1.4 pGEX-6P-3



Ampicillin resistance ORF:	bases 1391-2249
Origin of replication:	bases 2316-3012
Lac Repressor (*lacI*^{q}) ORF:	bases 3332-4412
Start codon for GST (ATG):	bases 258-260
pGEX 5′ priming site:	bases 869-891
Coding for PreScission protease site:	bases 918-938
Multiple cloning site:	bases 945-980
pGEX 3′ priming site	bases 1055-1033

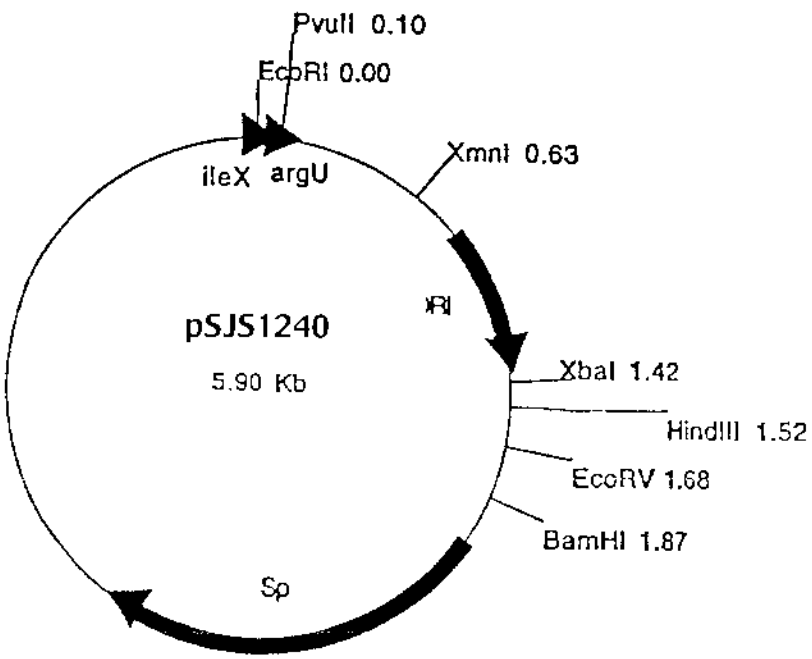
Image courtesy of APBiotech .

6.1.5 pGroESL



Plasmid name:	pGroESL
Derivitive of:	pACYC184
Plasmid size:	6.5 kb
Reference:	(Goloubinoff *et al.*, 1989)

6.1.6 pSJS1240



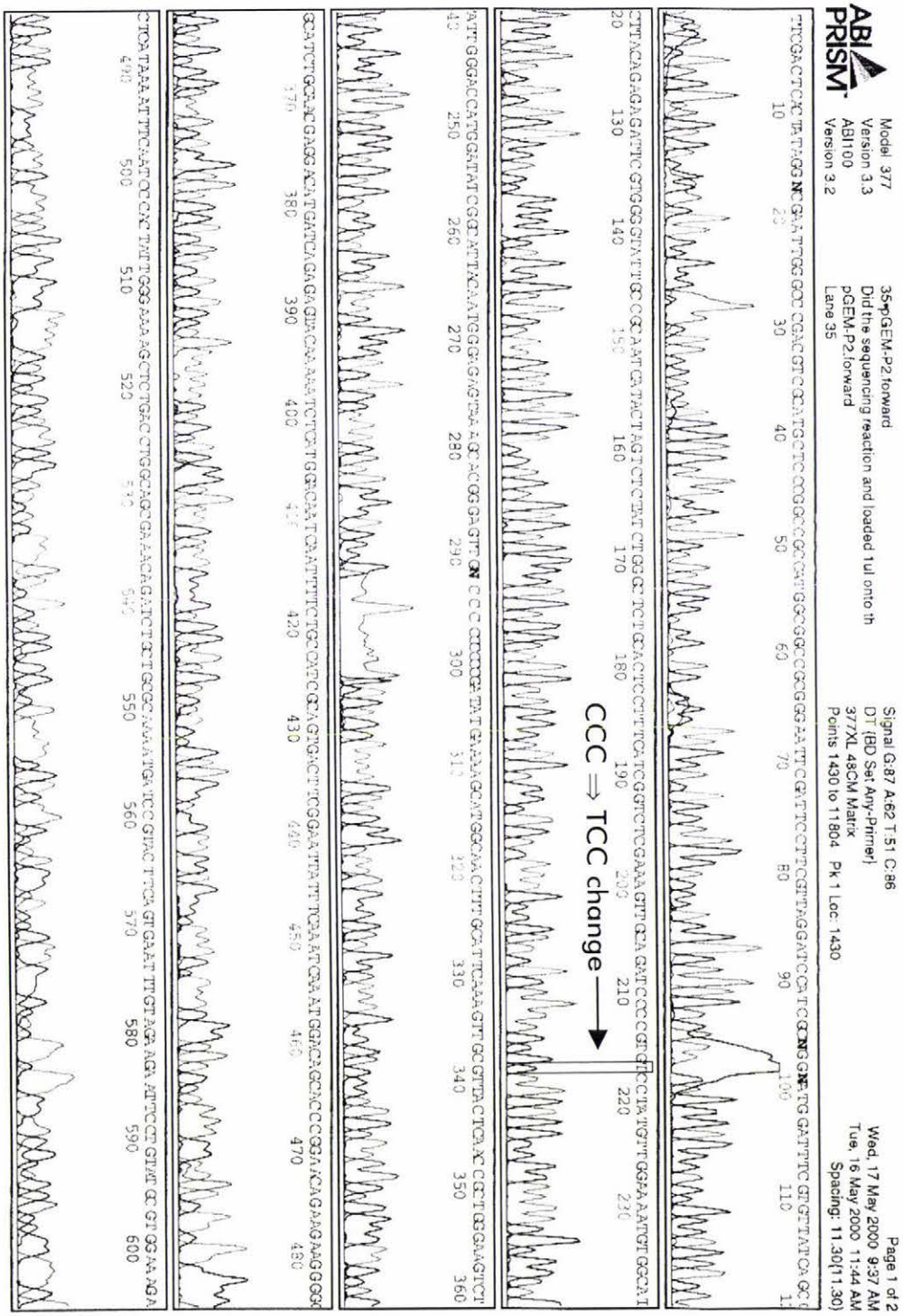
Plasmid name:	pSJS1240
Derivitive of:	pACYC184
Plasmid size:	5.9 kb
Reference:	(Del Tito *et al.*, 1995)

6.2 Automated sequencing chromatograms

6.2.1 Chromatogram: #1

Template: paxQ-pGEM-T-Easy

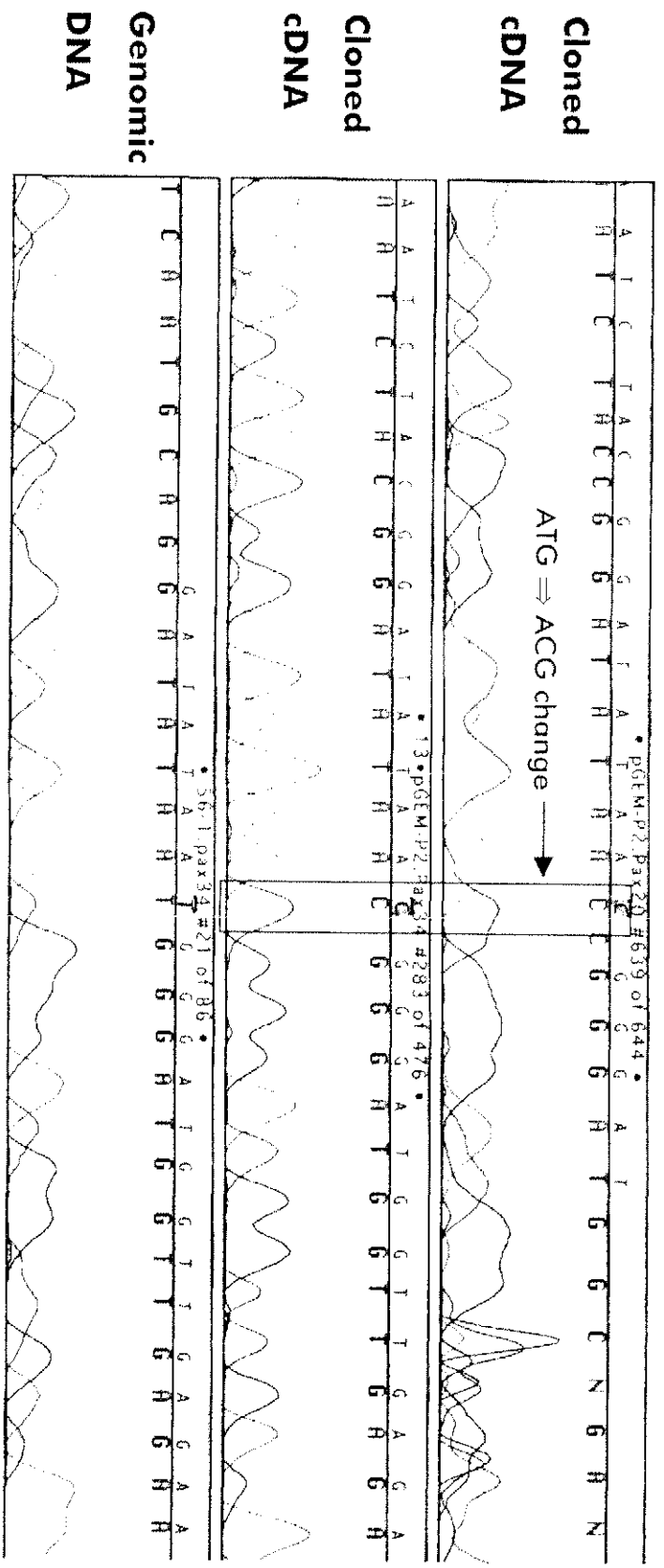
Primer: M13(*lacZ*)Forward (forward direction)



6.2.2 Chromatogram: #2

Template: paxQ-pGEM-T-Easy

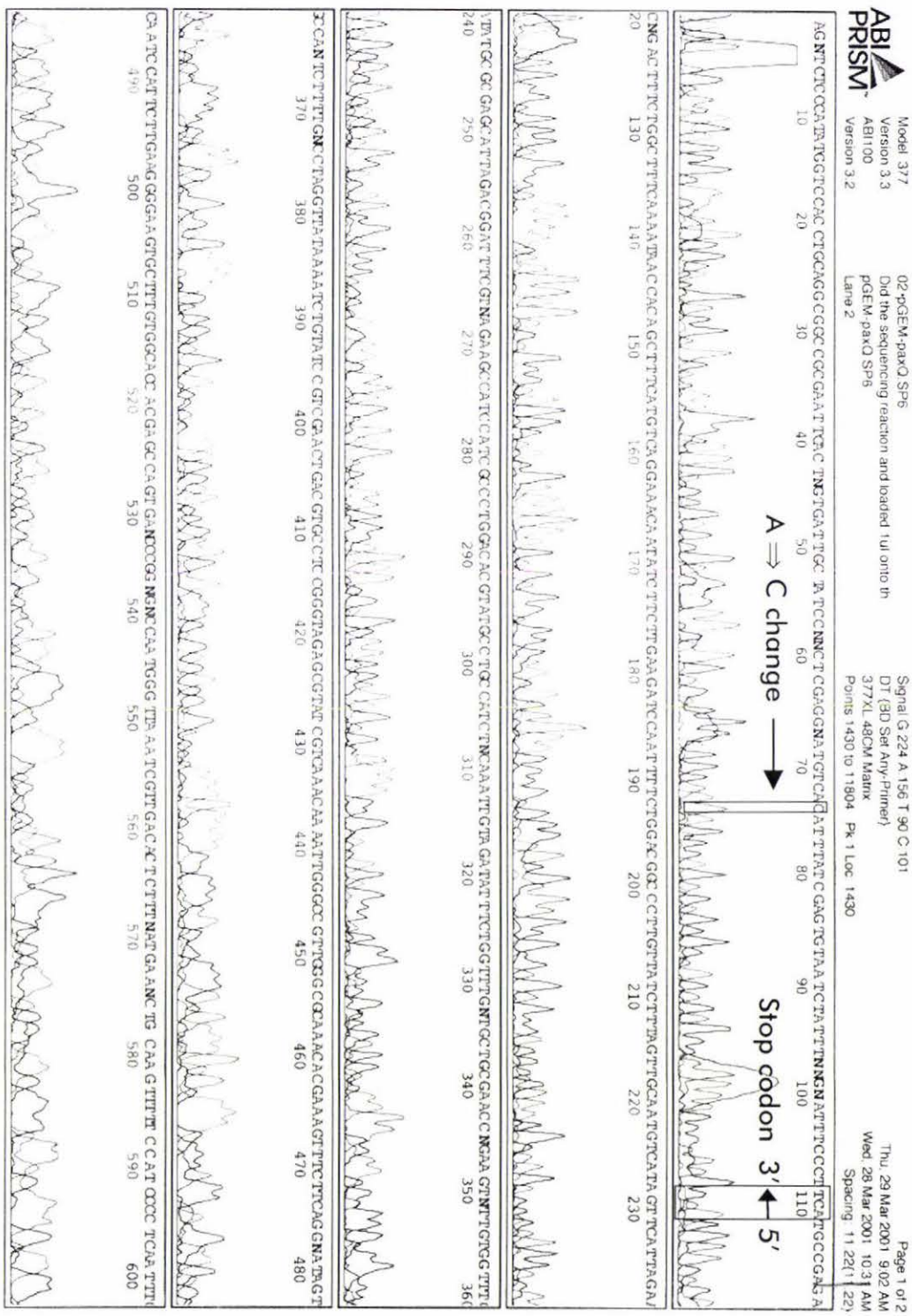
Primer: Pax20 and Pax34 (forward direction)



6.2.3 Chromatogram: #3

Template: pRL3

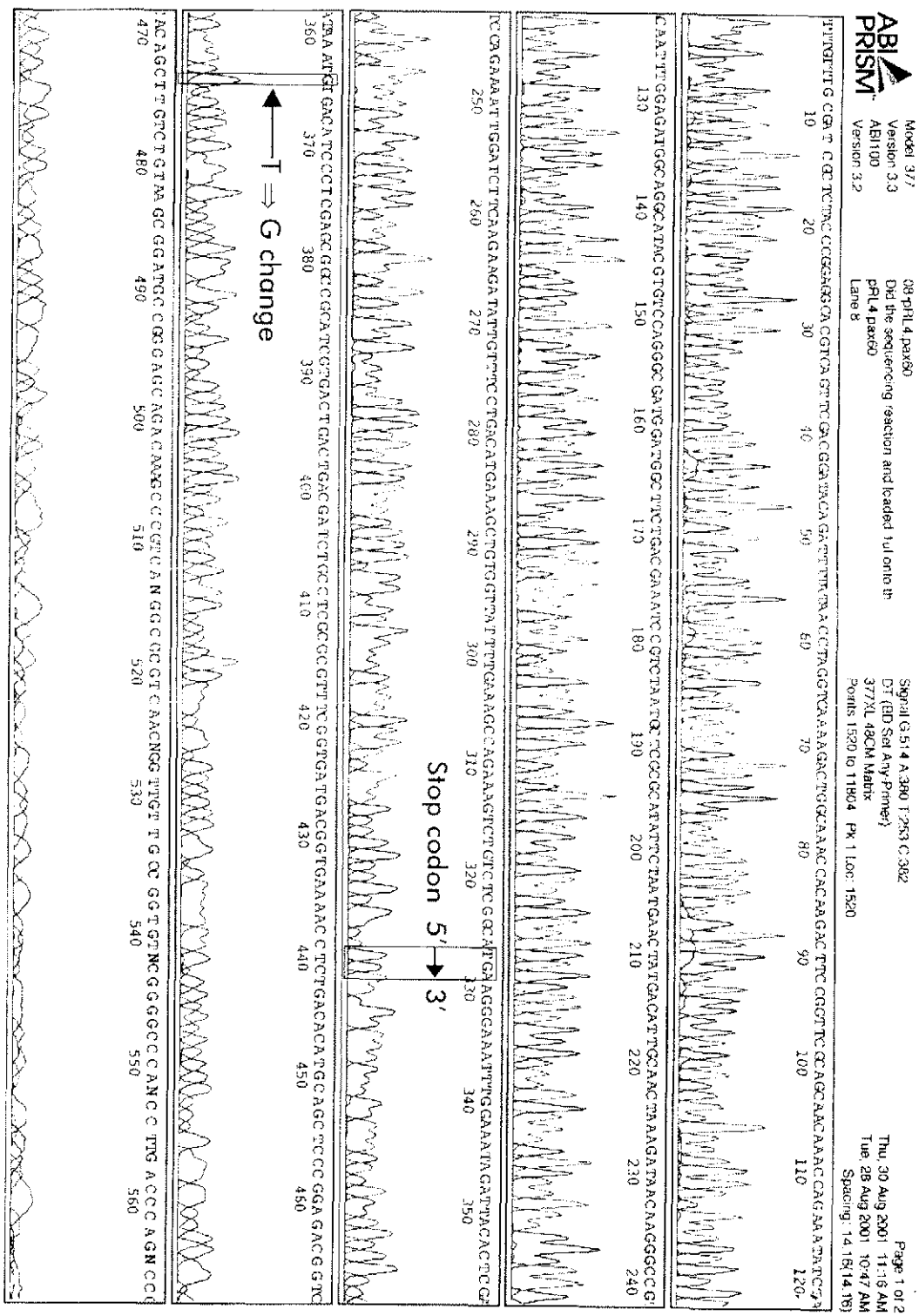
Primer: SP6 (reverse direction)



6.2.4 Chromatogram: #4

Template: pRL4

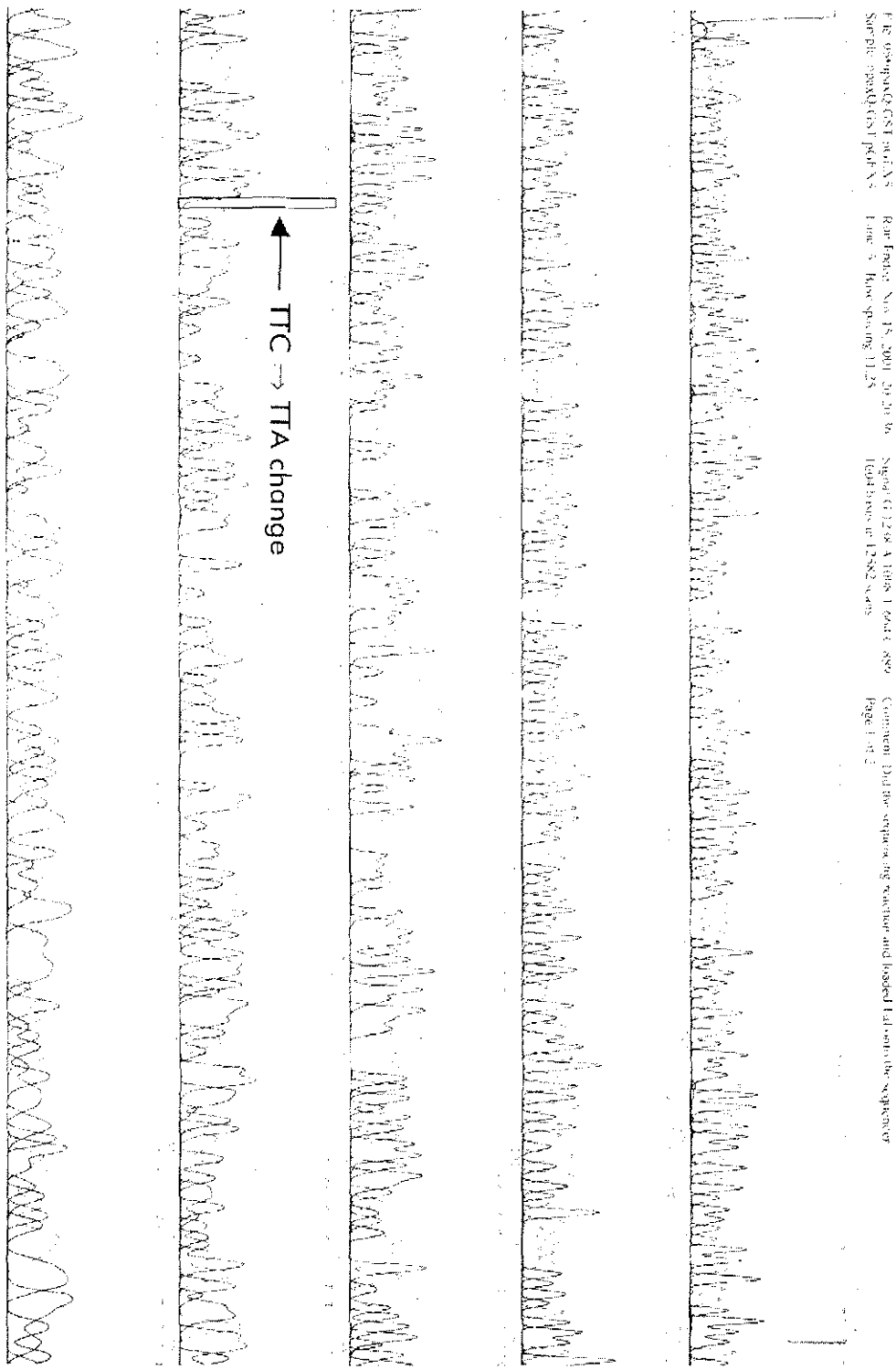
Primer: Pax60 (forward direction)



6.2.5 Chromatogram: #5

Template: pRL8

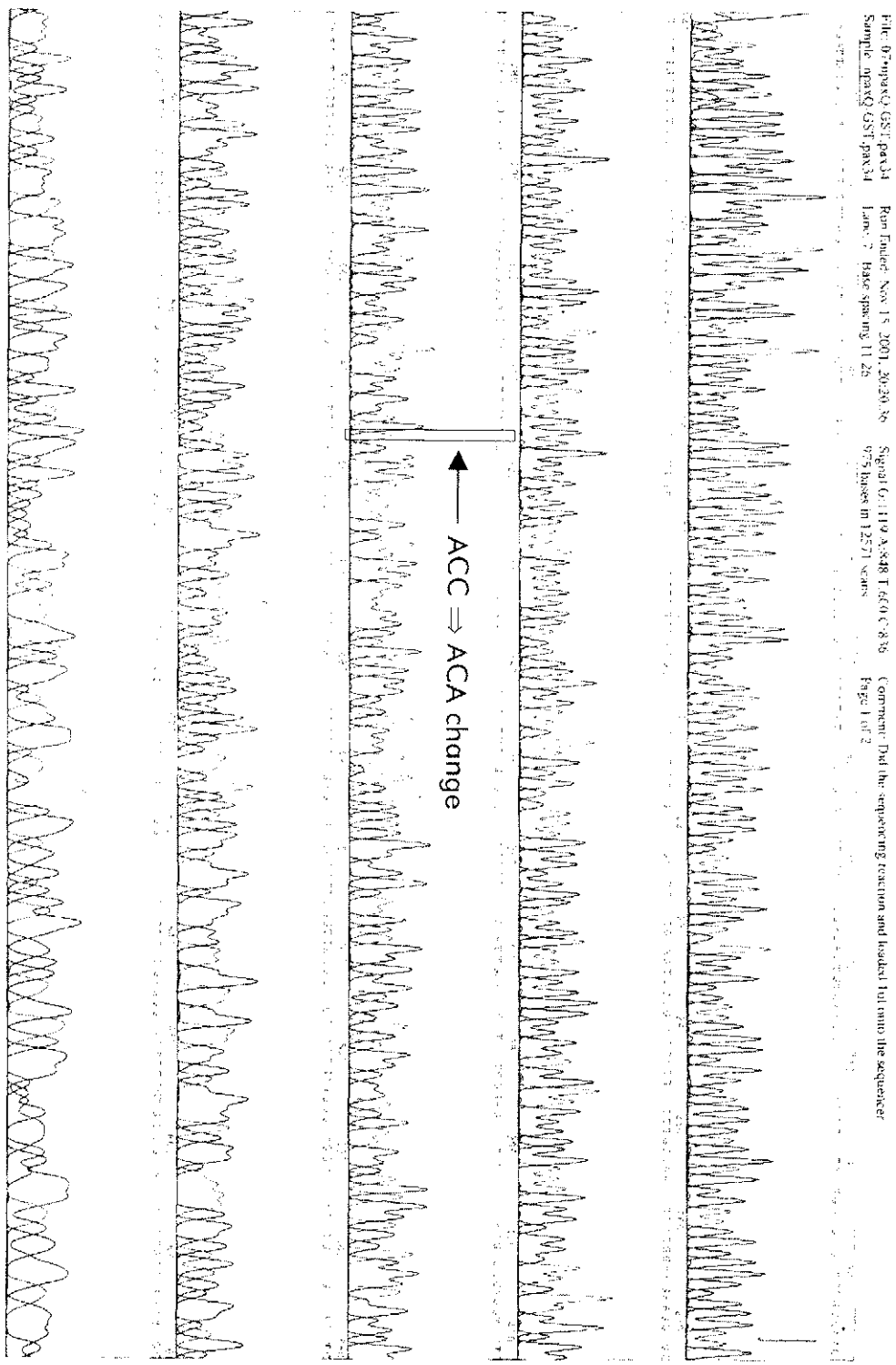
Primer: pGEX 5′ (forward direction)



6.2.6 Chromatogram: #6

Template: pRL8

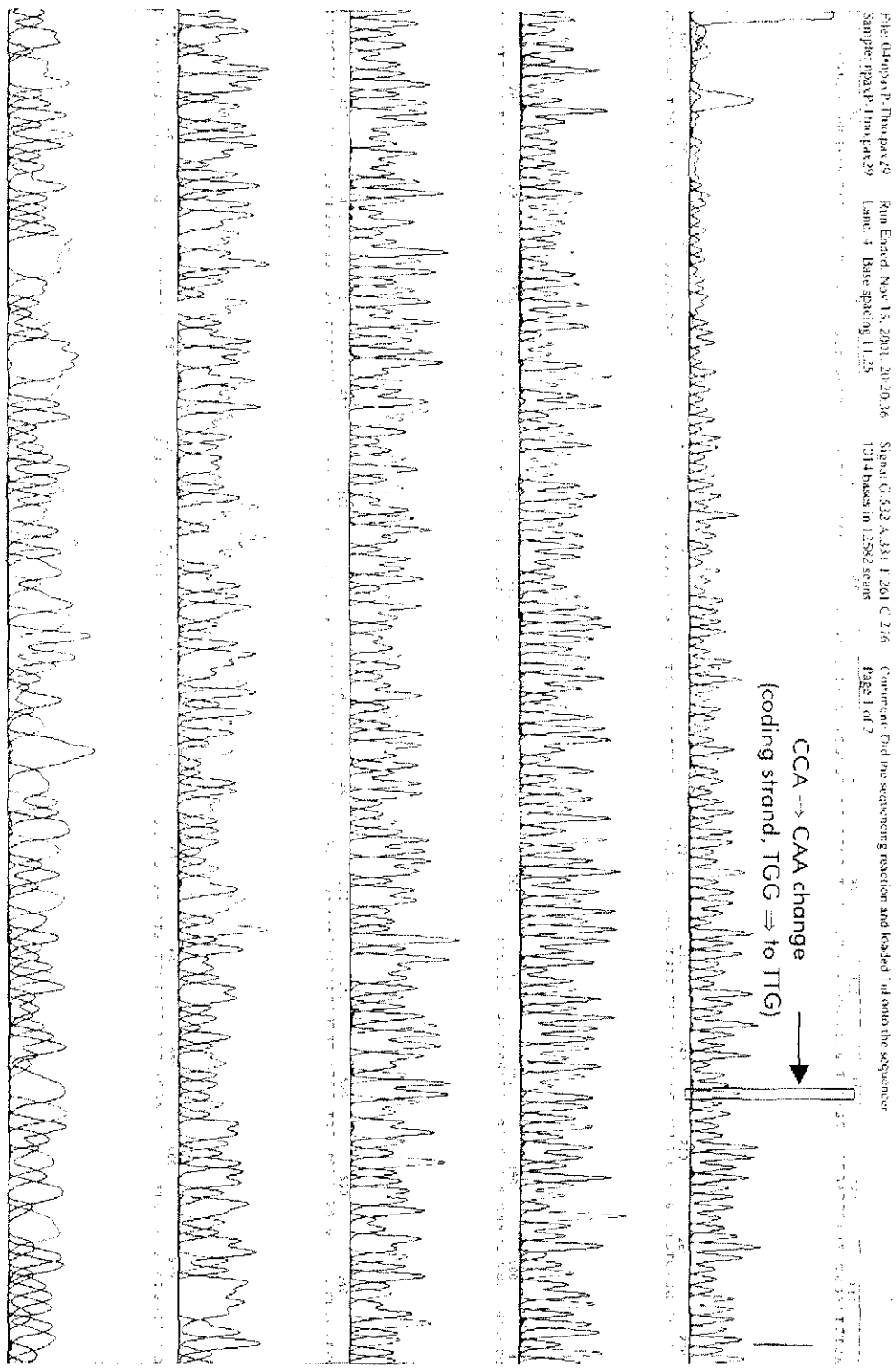
Primer: Pax34 (forward direction)



6.2.7 Chromatogram: #7

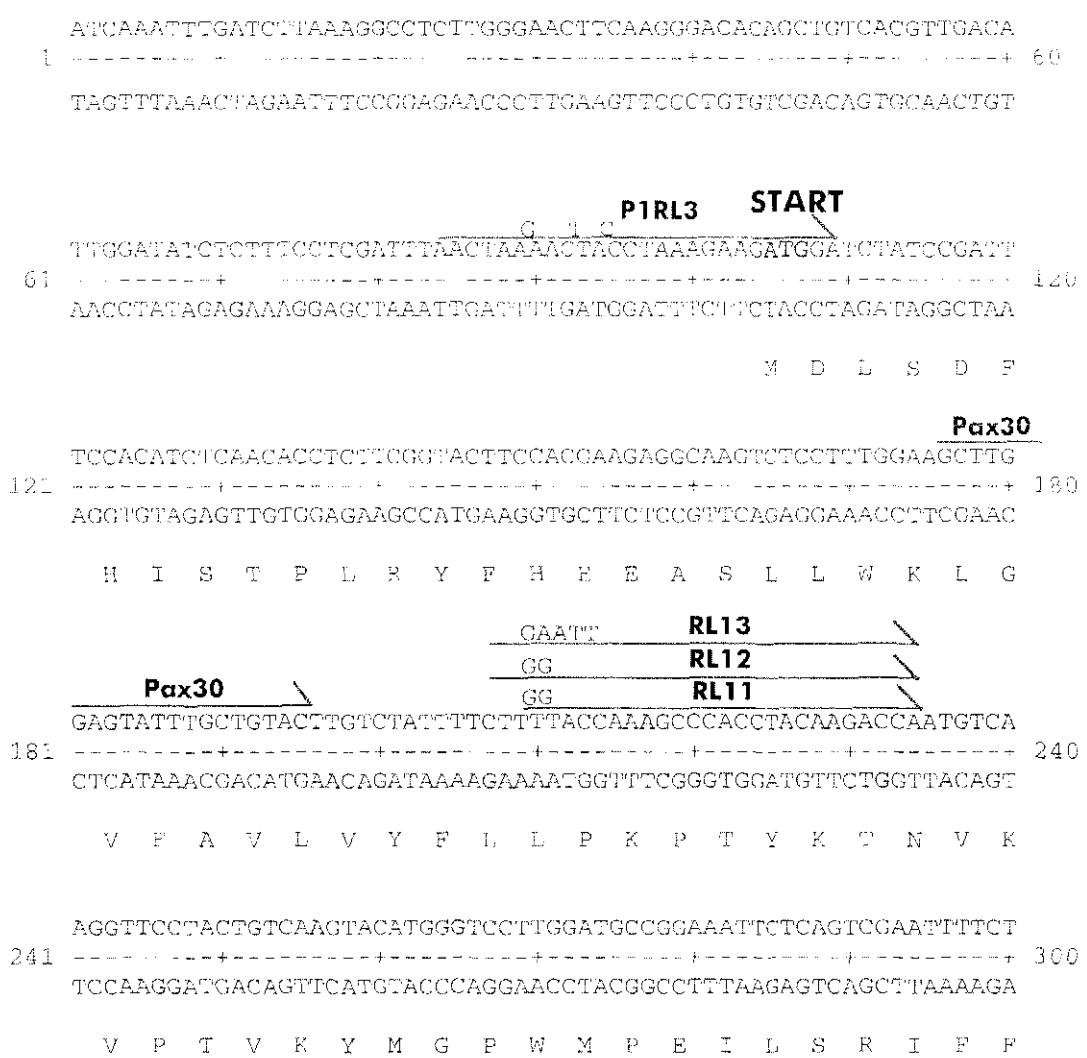
Template: pRL9

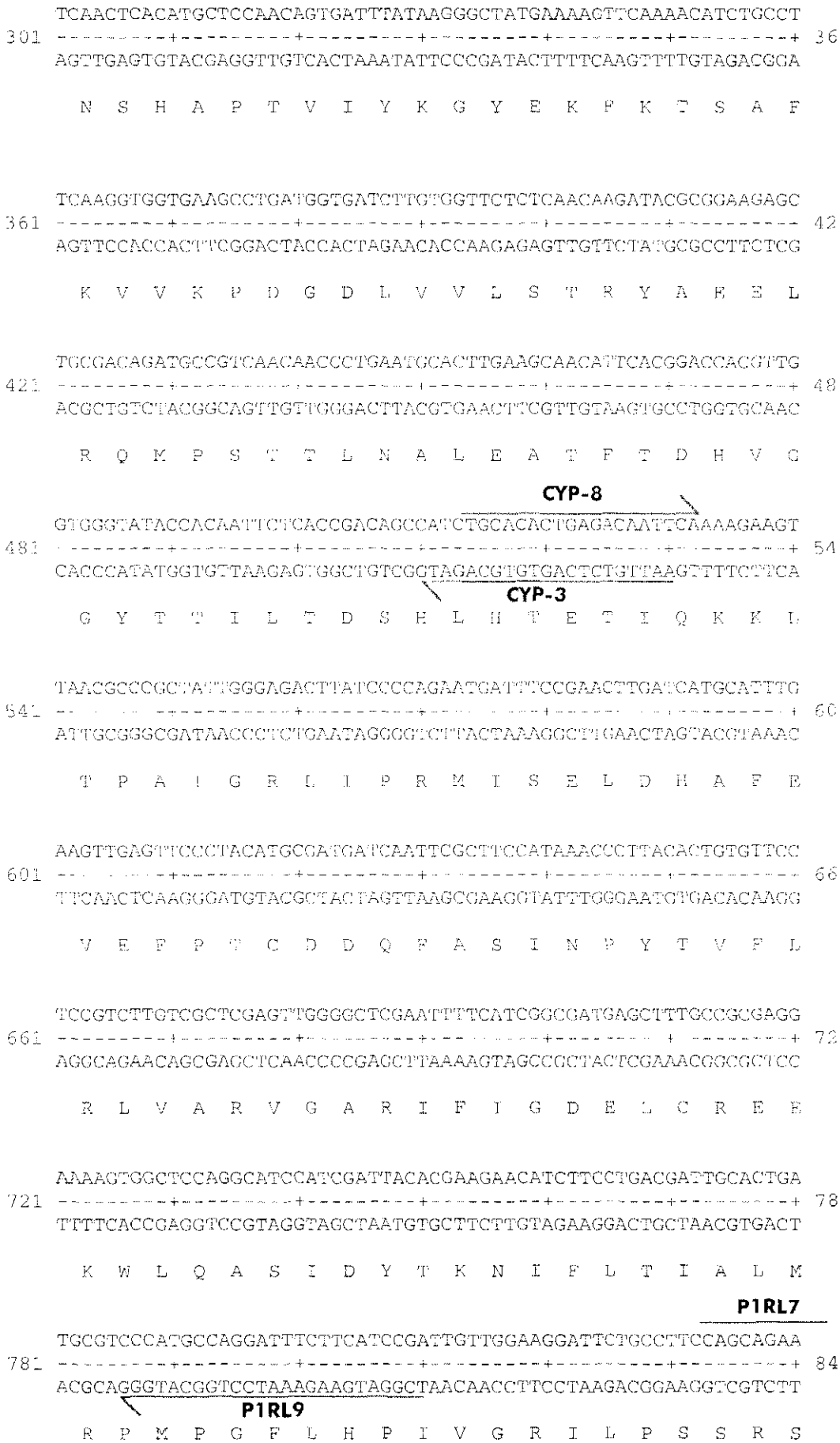
Primer: Pax29 (reverse direction)



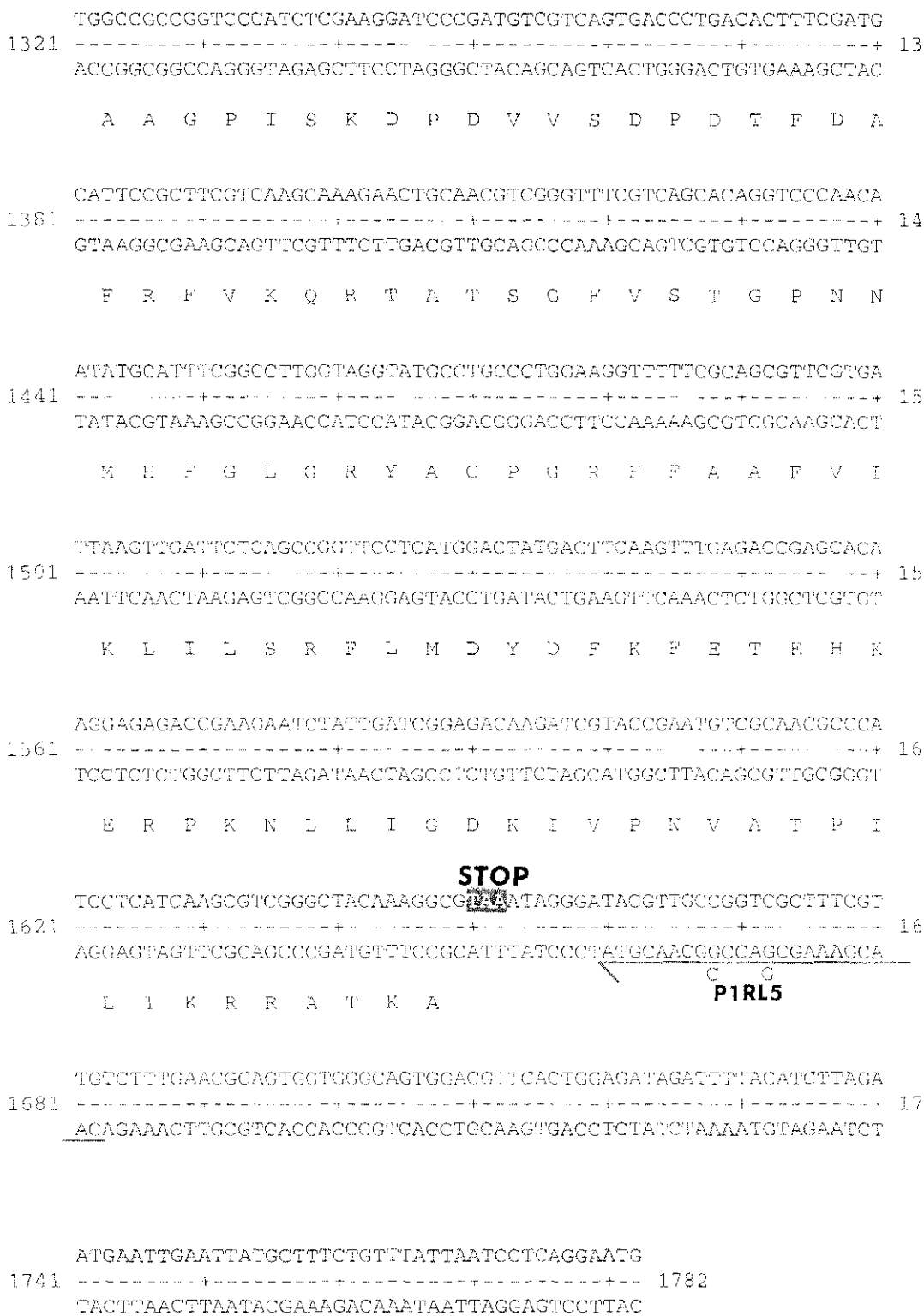
6.3 cDNA and deduced amino acid sequence of *P. paxilli paxP*

Both coding and template strands of *paxP* cDNA sequence are shown below. Nucleotide sequence is derived from accession number. AF279808, protein sequence is from accession number AAK11528. Nucleotides are aribtrarily numbered from the transcription start site. Deduced amino acid sequence is shown in frame with the coding sequence. Start and stop codons are highlighted. The annealing positions of primers used in this study (see Table 2.2) are highlighted with a line above the primer sequence. Primer base mismatches are listed above or below the sequence line.



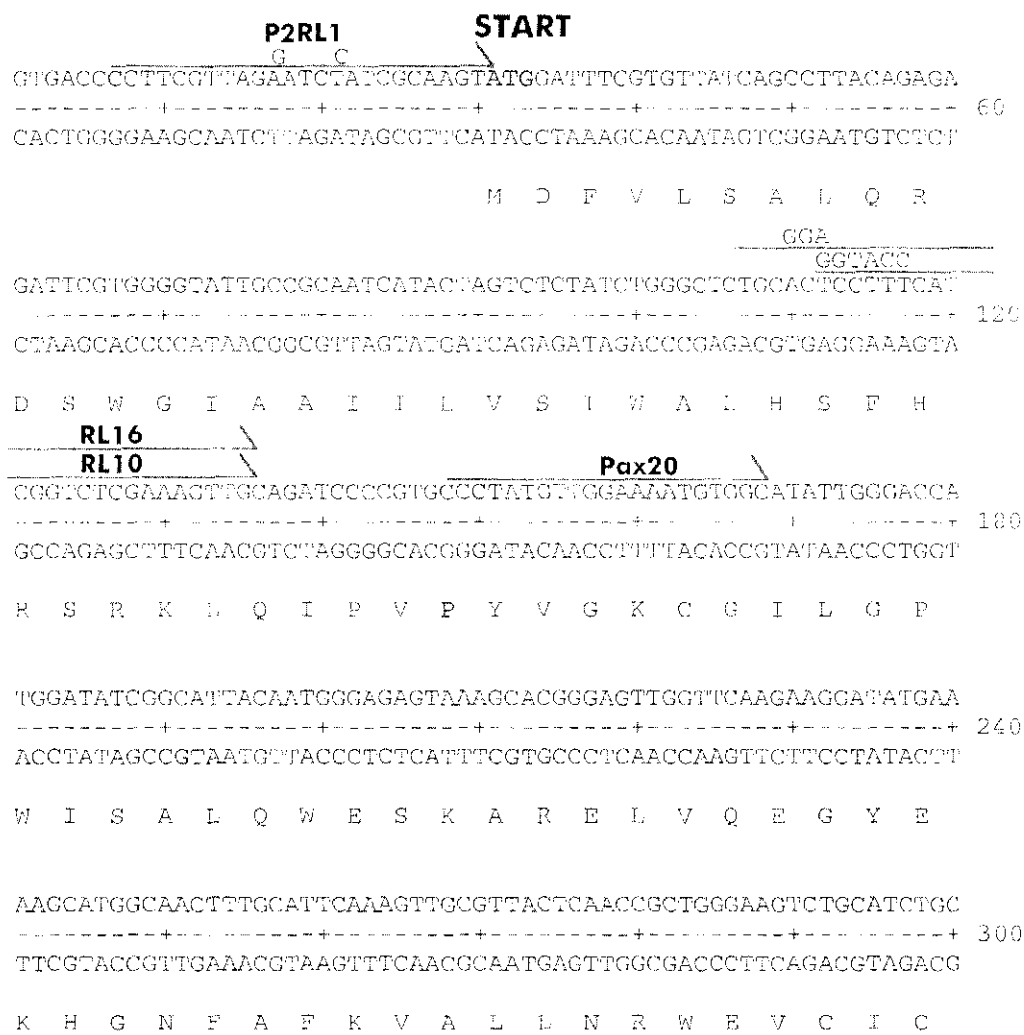

CYP-8
CYP-3
P1RL7
P1RL9

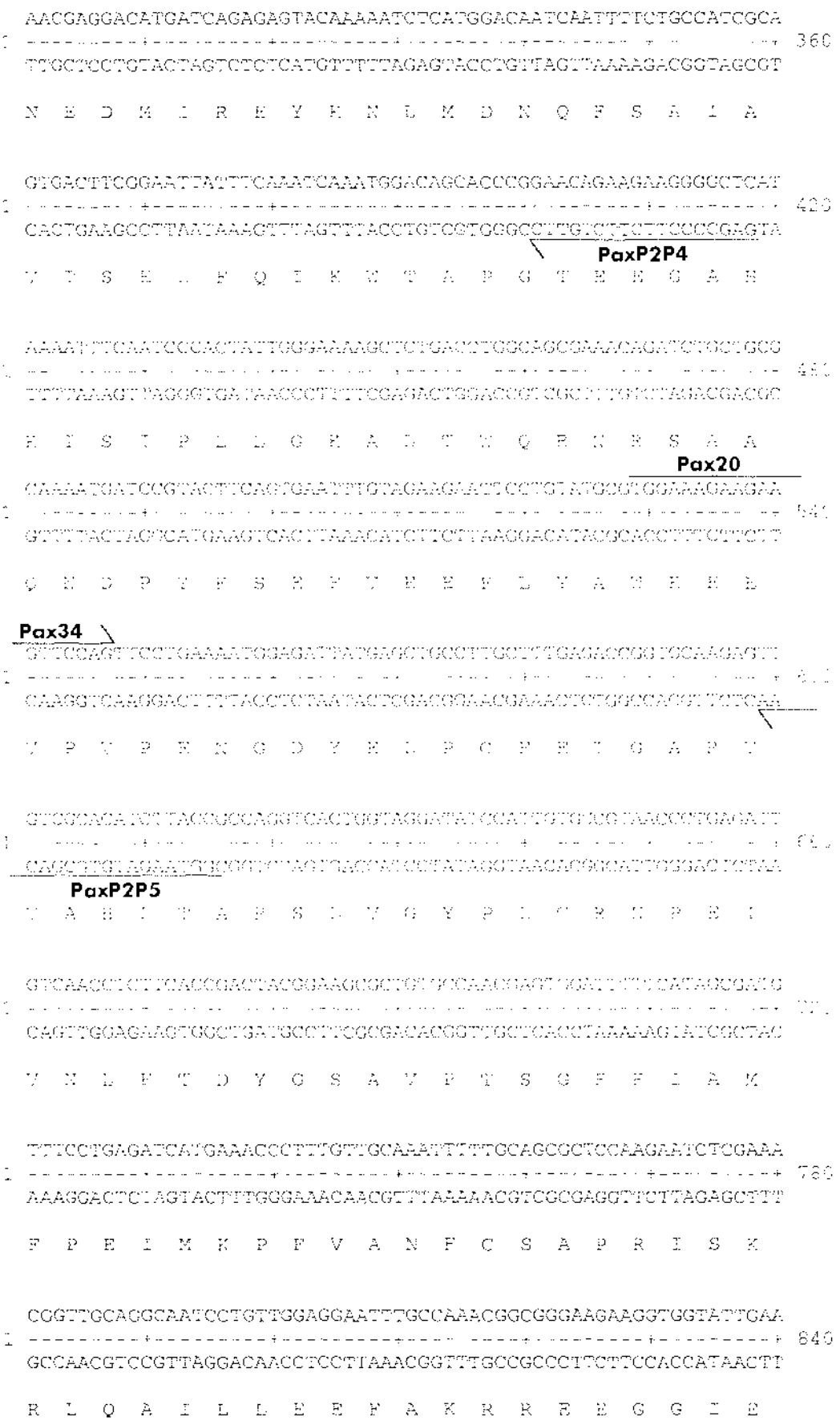

P1RL7
Pax29

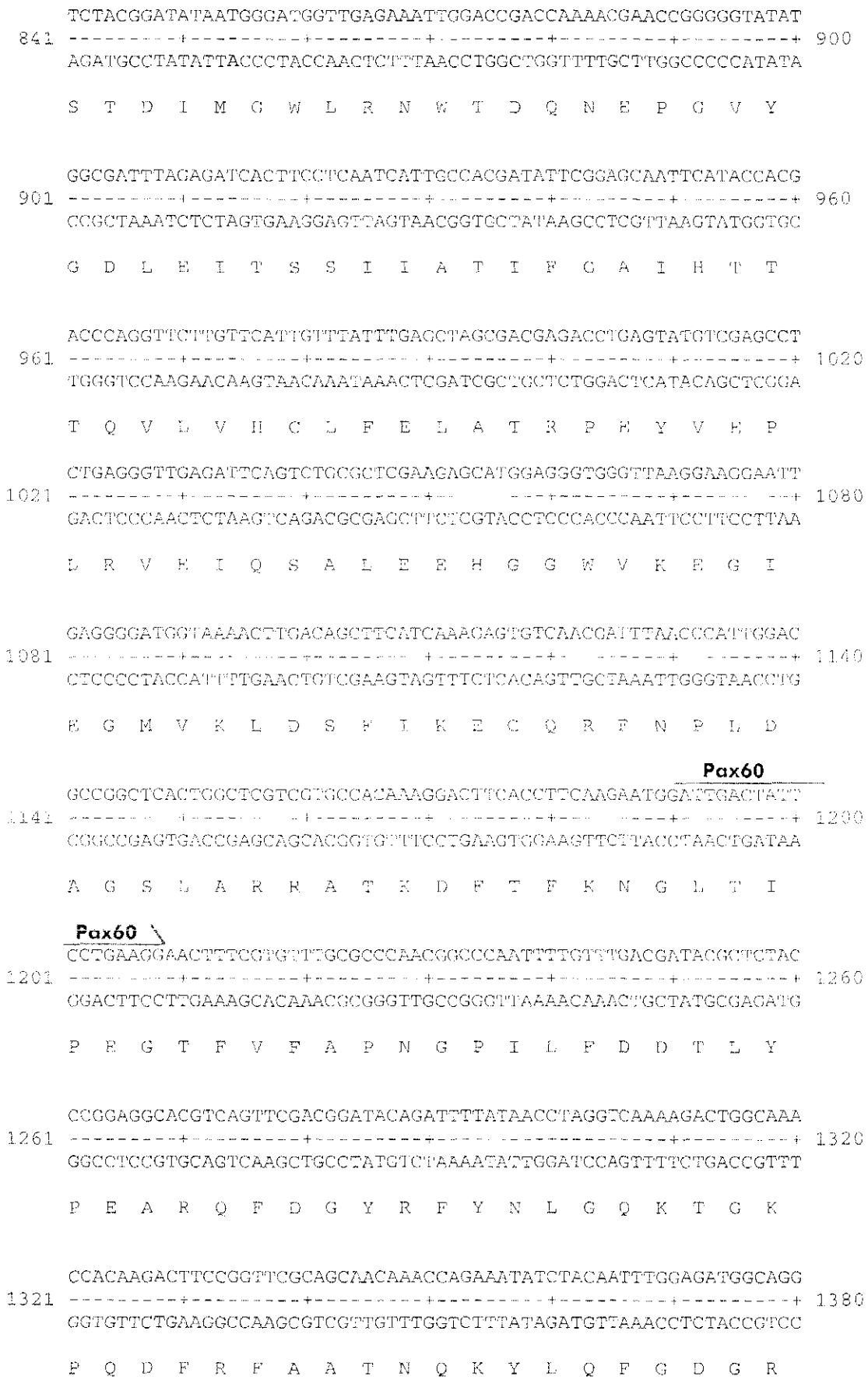

STOP
P1RL5

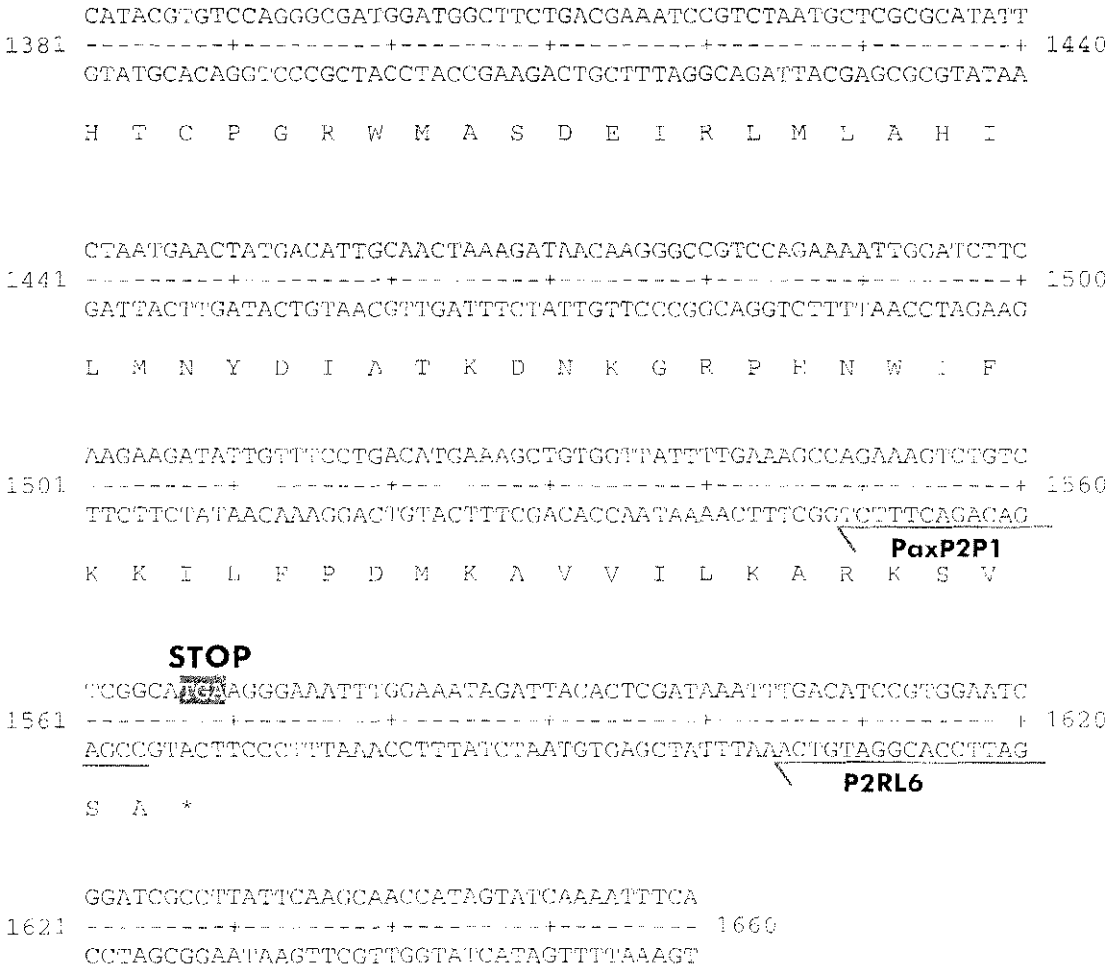
6.4 cDNA and deduced amino acid sequence of *P. paxilli paxQ*

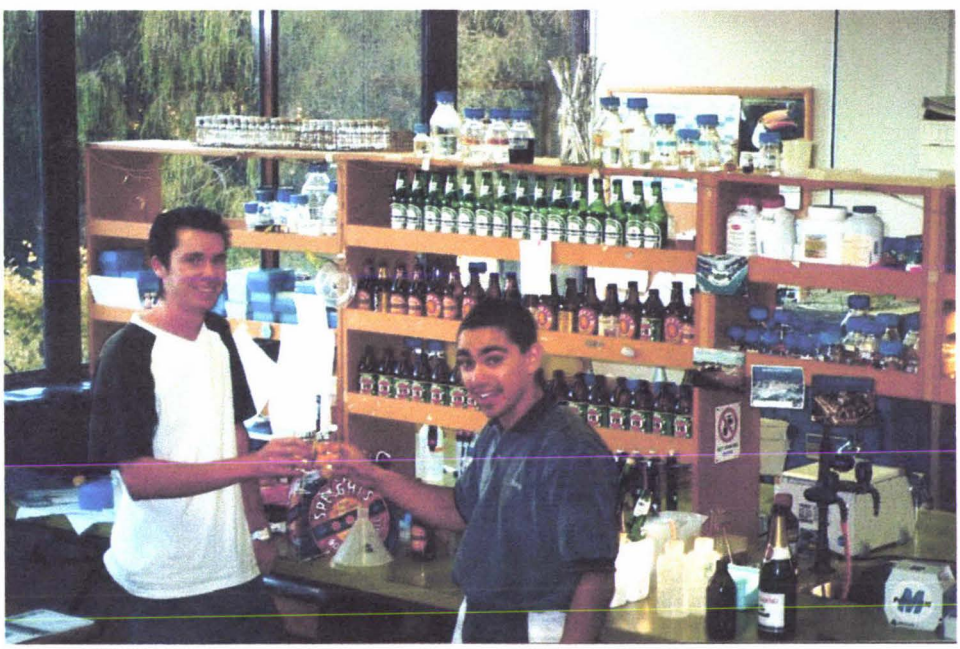
Both coding and template strands of *paxQ* cDNA sequence are shown below. Nucleotide sequence is derived from accession number AF279808, protein sequence is from accession number AAK11527. Nucleotides are aribtrarily numbered from the transcription start site. Deduced amino acid sequence is shown in frame with the coding sequence. Start and stop codons are highlighted. The annealing positions of primers used in this study (see Table 2.2) are highlighted with a line above the primer sequence. Primer base mismatches are listed above or below the sequence line




PaxP2P4
Pax20
Pax34
PaxP2P5


TCTACGGATATAATGGGATGGTTGAGAAATTGGACCGACCAAAACGAACCGGGGGTATAT
841
AGATGCCTATATTACCCTACCAACTCTTTAACCTGGCTGGTTTTGCTTGGCCCCCATATA
900
S T D I M G W L R N W T D Q N E P G V Y
GGCGATTTAGAGATCACTTCCTCAATCATTGCCACGATATTCGGAGCAATTCATACCACG
901
CCGCTAAATCTCTAGTGAAGGAGTTAGTAACGGTGCTATAAGCCTCGTTAAGTATGGTGC
960
G D L E I T S S I I A T I F G A I H T T
ACCCAGGTTCTTGTTCATTGTTTATTTGAGCTAGCGACGAGACCTGAGTATGTCGAGCCT
961
TGGGTCCAAGAACAAGTAACAAATAAACTCGATCGCTGCTCTGGACTCATACAGCTCGGA
1020
T Q V L V H C L F E L A T R P E Y V E P
CTGAGGGTTGAGATTCAGTCTGCGCTCGAAGAGCATGGAGGGTGGGTTAAGGAAGGAATT
1021
GACTCCCAACTCTAAGTCAGACGCGAGCTTCTCGTACCTCCCACCCAATTCCTTCCTTAA
1080
L R V E I Q S A L E E H G G W V K E G I
GAGGGGATGGTAAAACTTGACAGCTTCATCAAACAGTGTCAACGATTTAACCCATTGGAC
1081
CTCCCCTACCATTTTGAACTGTCGAAGTAGTTTCTCACAGTTGCTAAATTGGGTAACCTG
1140
E G M V K L D S F I K E C Q R F N P L D
Pax60
GCCGGCTCACTGGCTCGTCGTGCCACAAAGGACTTCACCTTCAAGAATGGATTGACTATT
1141
CGGCCGAGTGACCGAGCAGCACGGTGTTTCCTGAAGTGGAAGTTCTTACCTAACTGATAA
1200
A G S L A R R A T K D F T F K N G L T I
Pax60
CCTGAAGGAACTTTCGTGTTTGCGCCCAACGGCCCAATTTTGTTTGACGATACGCTCTAC
1201
GGACTTCCTTGAAAGCACAAACGCGGGTTGCCGGGTTAAAACAAACTGCTATGCGAGATG
1260
P E G T F V F A P N G P I L F D D T L Y
CCGGAGGCACGTCAGTTCGACGGATACAGATTTTATAACCTAGGTCAAAAGACTGGCAAA
1261
GGCCTCCGTGCAGTCAAGCTGCCTATGTCTAAAATATTGGATCCAGTTTTCTGACCGTTT
1320
P E A R Q F D G Y R F Y N L G Q K T G K
CCACAAGACTTCCGGTTCGCAGCAACAAACCAGAAATATCTACAATTTGGAGATGGCAGG
1321
GGTGTTCTGAAGGCCAAGCGTCGTTGTTTGGTCTTTATAGATGTTAAACCTCTACCGTCC
1380
P Q D F R F A A T N Q K Y L Q F G D G R


CATACGTGTCCAGGGCGATGGATGGCTTCTGACGAAATCCGTCTAATGCTCGCGCATATT
1381
1440
GTATGCACAGGTCCCGCTACCTACCGAAGACTGCTTTAGGCAGATTACGAGCGCGTATAA
H T C P G R W M A S D E I R L M L A H I
CTAATGAACTATGACATTGCAACTAAAGATAACAAGGGCCGTCCAGAAAATTGCATCTTC
1441
1500
GATTACTTGATACTGTAACGTTGATTTCTATTGTTCCCGGCAGGTCTTTTAACCTAGAAG
L M N Y D I A T K D N K G R P E N W I F
AAGAAGATATTGTTTCCTGACATGAAAGCTGTGGTTATTTTGAAAGCCAGAAAGTCTGTC
1501
1560
TTCTTCTATAACAAAGGACTGTACTTTCGACACCAATAAAACTTTCGGCCTTTCAGACAG
PaxP2P1
K K I L F P D M K A V V I L K A R K S V
STOP
TCGGCATGAAGGGAAATTTGGAAATAGATTACACTCGATAAATTTGACATCCGTGGAATC
1561
1620
AGCCGTACTTCCCTTTAAACCTTTATCTAATGTGAGCTATTTAAACTGTAGGCACCTTAG
P2RL6
S A *
GGATCGCCTTATTCAAGCAACCATAGTATCAAAATTTCA
1621
1660
CCTAGCGGAATAAGTTCGTTGGTATCATAGTTTTAAAGT



Cheers!