Physical and functional analysis of genes from the CAM catabolic plasmid encoding probable steps in the catabolism of camphor

By

Narongchai Tongyoo

A thesis submitted in partial fulfilment of the requirements for the degree of master of philosophy

Department of Biochemistry and Molecular Biology
University College London

November, 2002

Abstract

The ability to grow on either (+) or (-)-camphor, a bicyclic monoterpene ketone, of *Pseudomonas putida* NCIMB 10007 is conferred by the CAM plasmid, which has catabolic genes encoding enzymes that specify camphor degradation. Cloning of *camR* (cam repressor gene), *camD* (5-*exo*-hydroxycamphor dehydrogenase gene), *camC* (cytochrome P450cam gene), *camA* (putidaredoxin reductase gene) and *camB* (putidaredoxin gene) which are responsible for early steps of camphor degradation have been previously described. However, a study of the CAM plasmid involved in the further steps in camphor degradation has not been carried out. In this study, the nucleotide sequence of the 4485 bp on the left-hand side region of the *cam* operon on the CAM plasmid has been cloned. This nucleotide sequence consists of four possible open reading frames; *orf1*, *orf2*, *orf3* and *orf4*. The deduced amino acid of *orf1* shares identity with steroid monooxygenase (*Rhodococcus rhodochrous*), cyclohexanone monooxygenase (*Brevibacterium* sp. HCU) and cyclohexanone 1,2-monooxygenase (*Acinetobacter* sp. NCIMB 9871). Yet, the nucleotide sequence of *orf1* is incomplete and is probably missing its 5′end sequence. The deduced amino acid sequence of *orf2* is similar to putative limonene monooxygenase (*Rhodococcus erythropolis*), alkanal monooxygenase (*Xenorhabdus luminescens* HW) and luciferase related proteins. Moreover, the N-terminus of *orf2* deduced amino aicd is homologous (80% identity) with that of 2,5-diketocamphane 1,2-monooxygenase, an enzyme in the third step of camphor degradation. The Orf3 protein shares identity and similarity with a number of transcriptional regulator proteins in the tetR family. The deduced amino acid sequence of *orf4* is highly related to the methyl parathion hydrolase (*Plesiomonas* sp. M6), methyl parathion degrading protein (*Plesiomonas sp.* DLL-1), and other proteins in the β-metallo-lactamase family. A 40.7 kDa Orf2 monooxygenase was purified to homogeneity. However, the Orf2 monooxygenase has no activity towards (+)-and (-)-limonene, (+)-and (-)-pinene and cyclohexanone in the reaction with NADPH or NADH. The Orf4 protein is a membrane associated hydrolase possessing 29-leader signal peptide at its N-terminus. This signal peptide is; however, unprocessed by signal peptidease in *Escherichia coli*. Although whole cell activities using *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 showed that Orf4 hydrolase is capable of hydrolysing γ-butyrolactone and paraoxon with high initial hydrolysis rates of 84.2 and 13.0 μmol/h/mg-dry-cell weight respectively, the natural substrate for Orf4 hydrolase remains unknown.

Acknowledgements

I would like to thank

Dr. John Ward (my supervisor) for all advice, discussion and guidance through out the study

Prof. Peter piper (my mentor) for advice and discussion

Dr. Haitham Hussain for technical advice

Dr. Nicholas Harris for comment on the thesis

Royal Thai Government for the scholarship to study here

Everyone on the ground floor for friendly support

Table of contents

Page

List of tables

Table **Page**

List of figures

Figure **Page**

List of abbreviations

2,5-DKCMO	2,5-diketocamphane 1,2-monooxygenase
3,6-DKCMO	3,6-diketocamphane 1,6-monooxygenase
A	adenine
AHL	N-acylhomoserine lactonase
ATCC	American Type Culture Collection
AZT	3'-azido-3'-deoxythymidine
bp	basepair
C	cytosine
CHMO	cyclohexanone 1,2-monooxygenase
CoA	coenzymeA
DNA	deoxyribonucleic acid
dNTPs	dinucleotide triphosphates
EDTA	ethylene diamine tetra acetate
FAD	flavin adenine dinucleotide (oxidised form)
FADH	flavin adenine dinucleotide (reduced form)
FMN	flavin mononucleotide (oxidised form)
FMNH	flavin mononucleotide (reduced form)
G	guanine
HTH	helix-turn-helix
kb	kilobase
kDa	kilodalton
LMO	limonene monooxygenase
M	molarity
mRNA	massenger RNA
NAD	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
NADP	nicotinamide dinucleotide phosphate (oxidised form)
NADPH	nicotinamide dinucleotide phosphate (reduced form)
NCIMB	National Collections of Industrial and Marine Bacteria
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
RBS	ribosome binding site
RNA	ribonucleic acid
rpm	round per minute
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
T	thymine
tRNA	tranfer RNA
U	uracil
UV	ultraviolet

The Genetic code

	T	C	A	G
T	**TTT** Phe	**TCT** Ser	**TAT** Tyr	**TGT** Cys
	TTC Phe	**TCC** Ser	**TAC** Try	**TGC** Cys
	TTA Leu	**TCA** Ser	**TAA** Stop	**TGA** Stop
	TTG Lue	**TCG** Ser	**TAG** Stop	**TGG** Trp
C	**CTT** Leu	**CCT** Pro	**CAT** His	**CGT** Arg
	CTC Leu	**CCC** Pro	**CAC** His	**CGC** Arg
	CTA Leu	**CCA** Pro	**CAA** Gln	**CGA** Arg
	CTG Leu	**CCG** Pro	**CAG** Gln	**CGG** Arg
A	**ATT** Ile	**ACT** Thr	**AAT** Asn	**AGT** Ser
	ATC Ile	**ACC** Thr	**AAC** Asn	**AGC** Ser
	ATA Ile	**ACA** Thr	**AAA** Lys	**AGA** Arg
	ATG Met	**ACG** Thr	**AAG** Lys	**AGG** Arg
G	**GTT** Val	**GCT** Ala	**GAT** Asp	**GGT** Gly
	GTC Val	**GCC** Ala	**GAC** Asp	**GGC** Gly
	GTA Val	**GCA** Ala	**GAA** Glu	**GGA** Gly
	GTG Val	**GCG** Ala	**GAG** Glu	**GGG** Gly

Abbreviations for amino acids

Amino acid	Symbol	Three-letter symbol
Alanine	**A**	Ala
Arginine	**R**	Arg
Asparagine	**N**	Asn
Aspartic acid	**D**	Asp
Cysteine	**C**	Cys
Glutamine	**Q**	Gln
Glutamic acid	**E**	Glu
Glycine	**G**	Gly
Histidine	**H**	His
Isoleucine	**I**	Ile
Leucine	**L**	Leu
Lysine	**K**	Lys
Methionine	**M**	Met
Phenylalanine	**F**	Phe
Proline	**P**	Pro
Serine	**S**	Ser
Threonine	**T**	Thr
Tryptophan	**W**	Trp
Tyrosine	**Y**	Tyr
Valine	**V**	Val
Any amino acid	**X**	NNN

Chapter 1

Introduction

1.1 *Pseudomonas*

The genus *Pseudomonas* forms one of the largest genera of important bacteria. Members of the *Pseudomonas* species are, for example, *P. aeruginosa*, *P. fluorescens*, *P. syringae, P. diminuta*, *P. oleovorans*, *P. cepacia* and *P. putida*. Such Gram-negative bacteria contain an outer membrane which Gram-positive bacteria lack. The genus *Pseudomonas* is the most diverse group of bacteria found widely in environments; soil, water and marine. Bacteria in this genera is also found to be associated with plants and animals (Spiers *et al.*, 2000). In the environment, therefore, the roles of *Pseudomonas* are such as in animal pathogenicity (*P. aeruginosa*), plant-microbes interaction (*P. fluorecens* and *P. syringae*) and the biodegradation of natural and man-made chemicals (*P. diminuta, P. oleovorans, P. cepacia* and *P. putida*). *Pseudomonas* are also notable for their metabolic versatility and their ability to use various nutrients. In a single literature, Stanier and coworkers (1966) reported a collection of 297 *Pseudomonas* strains showed their ability to grow on more than a hundred different organic compounds. Frequently, the ability of *Pseudomonas* to metabolise chemical compounds is conferred by genes on plasmids.

1.2 The CAM plasmid

Plasmids can be defined as extra chromosomal DNAs found in most species of bacteria. The size of plasmid ranges from 2.6 to more than hundreds of kilobases and the copy number per cell ranges from 1 to 700 copies. During the 1940s to 1960s, early plasmid research has been mainly focused in *Eschericia coli*. Based on these early

studies, the plasmid phenotypes in *E. coli* are characterised into 3 different traits (Summers, 1996). First, fertility plasmids (F plasmids) allow bacteria to transfer genetic material and the F plasmid itself to other cells. Second, drug resistant plasmids (R plasmids) have genes that confer antibiotic resistance. Third, colicingenic plasmids (Col plasmids) synthesise colicins, proteins that have the ability to kill susceptible bacteria related to the host cell.

However, there is another kind of plasmid which is defined as a degradative plasmid. To the host cell, the degradative plasmid confers the ability to metabolise a number of organic or man-made compounds. Interestingly, degradative plasmids are especially found in saprophytic bacteria e.g. the *Pseudomonas* strains. Well-known degradative plasmids in *Pseudomonas* strains are: i) SAL (responsible for salicylate degradation, salicylate to catechol); ii) NAH (the plasmid in naphthalene degradation, naphthalene to pyruvate and acetaldehyde); iii) OCT (the plasmid that has a degradative function for *n*-alkanes; octane, hexane and decane); iv) TOL (the plasmid specifies for degradation of *m*-or *p*-toluates, and corresponding xylenes); and v) the CAM (the plasmid responsible for the early steps of camphor degradation, camphor to 5-*exo*-hydroxycamphor to 2,5-diketonecamphane or 3,6-diketonecamphane to campholide and (steps) to 3 acetates and isobutyrate) (Chakrabarty, 1976). The CAM plasmid is an interesting plasmid in *Pseudomonas* species. *P. putida* carrying the CAM plasmid is able to grow on camphor as a sole carbon and energy source.

P. putida grown on camphor was first reported in 1959 where the bacteria isolated from sewage sludge was studied (Bradshaw *et al.*, 1959). In 1971 Chakrabarty and Gunsalus found that the ability to use camphor as a sole carbon source is confered by the CAM plasmid.

The phenotypes of CAM plasmid are defined as Tra^+ Fi^+ (RP1) Fi^- (FP2) Phi (B3 B39 D3 E79 G101 M6 BP1) Cam and UV (Bukhari *et al.*, 1977). The CAM plasmid has the ability to promote self transfer, inhibit fertility of the RP1 plamids (Tra^+) but not the FP2 plasmid, inhibit many phages, use camphor as a sole carbon source and mobilise the

P. putida chromosome. The CAM plasmid is classified as an $IncP_2$ plasmid, sharing a common replication system, and in the same group as pMG1, pMG3 (an antibiotic resistant plasmid) and OCT (degradative plasmid). The IncP2 plasmids are transmissible between *Pseudomonas* strains but they are not transmissible to *E. coli*. The host-specific InP2 plasmids have replication systems which function in the *Pseudomonas* species, but will not function in enterobacteria (Bukhari *et al.*, 1977).

Predictions regarding the size of the CAM plasmid have ranged from 150 to 312 MDa or about 230 to 500 kb (Chakrabaty, 1976 and Hansen and Olsen, 1978).

1.3 Camphor

Camphor, whose systematic name is 1,7,7-trimethylbicyclo [2.2.1] heptan-2-one, is a compound which can be found in the camphor tree, *Cinnamomum camphora*, and common sage, *Salvia afficinalis* (Funk *et al.*, 1992). Camphor is a bicyclic monterpene ketone that has two mirror-image compounds or enantiomers which are designated (+) and (-) isomers (see Figure 1.1). This monoterpene ketone is commonly used in ointments and liniments, and in chemistry it is used as a starting substrate in organic synthesis.

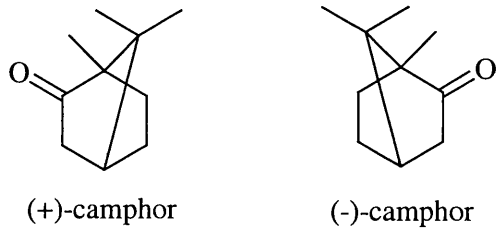


Figure 1.1 Chemical structure of (+) and (-)-camphor.

Almost 30% of (+)-camphor is a major part of the volatile oil of the leaves of common sage. However, the content of (+)-camphor declines when the plants enter the mature stage (Funk *et al.*, 1992). (+)-Camphor can be catabolised by sage itself or metabolised by microorganisms (Croteau *et al.*, 1984). Several microorganisms that can catabolise camphor have been reported such as *P. putida* and *Corynebacterium* T1.

1.4 Camphor degradation by *P. putida* NCIMB 10007

The degradation of (+) and (-)-camphor by *P. putida* C_1 (*P. putida* strain NCIMB 10007 or ATCC 17453) has been demonstrated by LeGall (1963), who showed that both (+) and (-)-camphor served equally as a growth substrate for the bacteria. This indicates that (+) and (-)-camphor are metabolised in parallel. Further studies also show that (+) and (-)-camphor enantiomers are catabolised in parallel through two routes (Gunsalus and Marshal, 1971). However, the study of (+)-camphor metabolism was carried out in much more detail than that of (-)-camphor.

In Figure 1.2, the main metabolic pathway of camphor degradation in *P. putida* is shown in detail. The known-camphor degradation pathway is from D-(+)-camphor to $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA (**IV**). First, the reaction starts with the hydroxylation of carbon 5 by cytochrome P450cam, NADH putidaredoxin reductase and putidaredoxin to form 5-*exo*-hydroxycamphor. In addition, NADH, FAD and O_2 are required in this step. Second, 5-*exo*-hydroxycamphor is dehydrogenated by 5-*exo*-hydroxycamphor dehydrgenase (2) which requires of NAD as a cofactor to form 2,5-diketocamphane. Third, the oxygenation of 2,5-diketocamphane proceeds by a Baeyer-Villiger reaction to form the unstable lactone, 5-oxo-1,2-campholide which is spontaneously rearranged to 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetate (**I**). In the third step, the oxygenating subunits of 2,5-diketocamphane monooxygenase (3) NADH dehydrogenase, NADH, FMN and O_2 are required. However, 2,5-diketocamphane 1,2-monooxygenase can utilise (+)-camphor to from stable campholide (**XI**) as well. Fourth, 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetate (**I**) is enzymetically converted into 2-oxo-Δ^3-4,5,5-

trimethylcyclopentenylacetyl-CoA (**II**) by 2-oxo-Δ^3-4,5,5- trimethylcyclopentenylacetyl-CoA synthetase (5). Then, 2-oxo-Δ^3-4,5,5- trimethylcyclopentenylacetyl-CoA is catalysed by another Baeyer-Villiger monooxygenase, 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA-1,2-monooxygenase (6), to form 5-hydroxy-3,4,4-trimethyl-Δ^2-pimelyl-CoA-δ-lactone (**III**). In this step, the enzyme requires NADPH (as the electron donor) and O_2 in this oxygenation reaction. Finally, $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA (**IV**) is formed by spontaneous ring opening. The catabolic reactions of $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA to isobutyrate and acetates are, however, not yet fully delineated.

Since there are several lactone intermediates of camphor oxidation in *P. putida*, the lactone hydrolase may be required in the cell. However, the lactone hydrolase has never been detected in hydrolysing of lactone intermediates of camphor catabolic pathway in *P. putida* NCIMB 10007 (Ougham *et al.*, 1983).

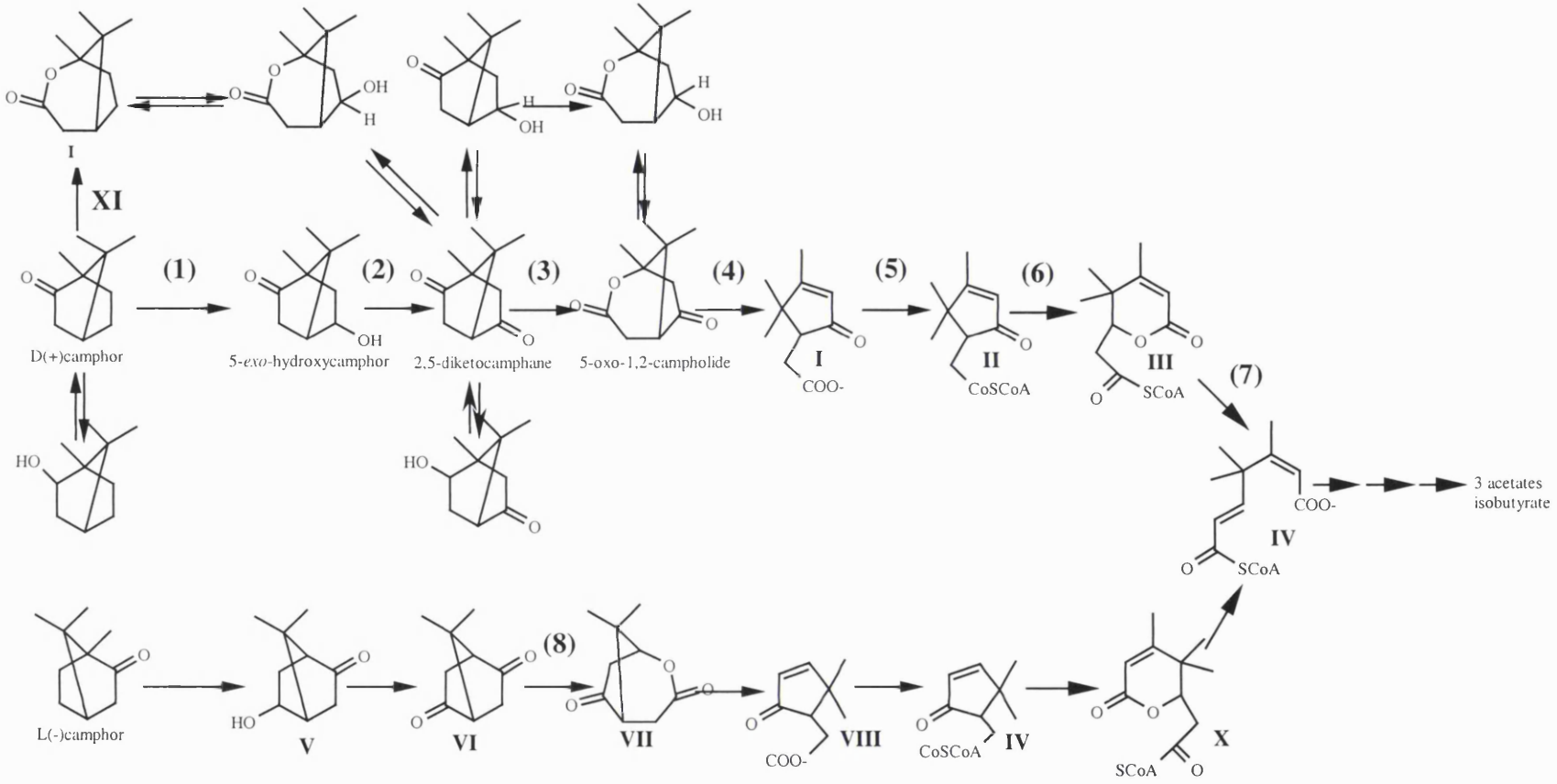


Figure 1.2 Early steps in the catabolism of camphor isomer by *P. putida* (strain ATCC 17453 or NCIMB 10007). (1) is cytochrome P450cam, putida redoxin reductase and putidaredoxin; (2) is 5-*exo*-hydroxycamphor dehydrogenase; (3) is 2,5-diketocamphane 1,2-monooxygenase; (4) is spontaneous reaction; (5) is 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA synthetase;

Figure 1.2 (continued) (6) is 2-oxo-Δ^3-45,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygense; (7) is a spontaneous reaction; and (8) is 3,6-diketocamphane 1,6-monooxygenase. **I**; 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetate, **II**; 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA, **III**; 5-hydroxy-3,4,4-trimethyl-Δ^2-pimelyl-CoA-δ-lactone, **IV**; $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA, **V**; 5-*endo*-hydroxycamphor, **VI**; 3,6-diketocamphane, **VII** ;the isomer of 5-exo-1,2-campholide and **VIII-X** are the isomers of **I**, **II** and **III** respectively.

Although the enzymatic pathway of camphor to $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA is well established, a further pathway of $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA to accetate and isobutyrate is still unclear. However, a possible pathway for this has been postulated as in Figure 1.3 (Sokatch, 1986).

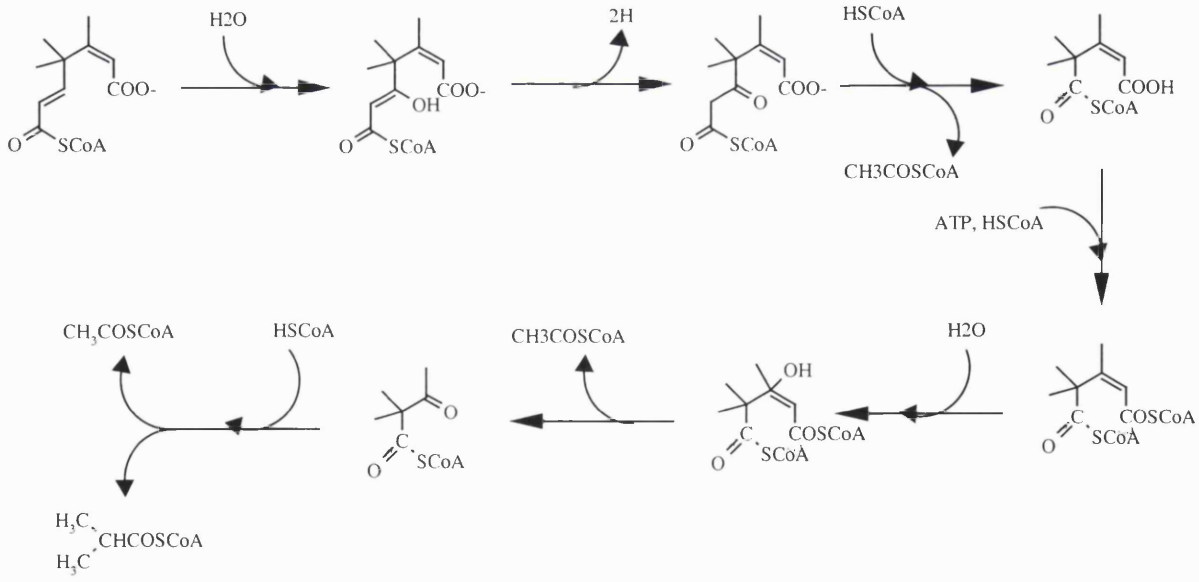
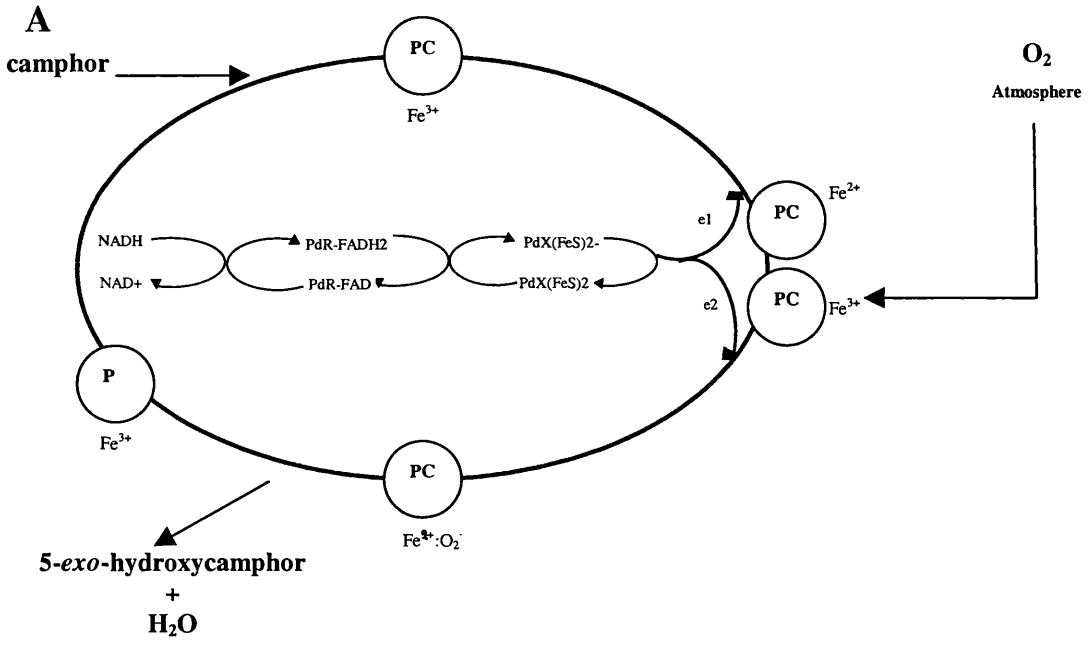


Figure 1.3 A postulated pathway of $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA to acetate and isobutyrate by *P. putida*.

1.5 The mechanism of (+)-camphor oxidation

The oxidation of bicyclic monoterpene (+)-camphor to 5-*exo*-hydroxycamphor was described in the 1960s (Katagiri *et al.*, 1968 and Hedegaard and Gunsalus, 1965). This oxidation requires a monooxygenase multicomponent system, cytochrome P450cam, putidaredoxin reductase and putidaredoxin. In addition, NADH, flavin adenine dinucleotide (FAD), O_2 and the substrate camphor are required in the reaction. First, an electron donor NADH is reduced by putidaredoxin reductase to form NAD^+. The electrons are transferred to FAD resulting in reduced flavin-adenine dinucleotide, $FADH_2$. Second, the putidaredoxin, iron-sulpher ($[FeS]_2$) protein, receives the electron from the reduced flavin ($FADH_2$). Then the electrons are transferred from putidaredoxin-$[FeS]_2^-$ to cytochrome P450cam which contains heme iron (Fe^{3+}), and cytochrome P450-Fe^2 is formed. Afterwards, the reduced cytrochrome P450-Fe^{2+} uses the electrons to cleave molecular oxygen (O_2) and incorporates an oxygen atom into the camphor molecule. The end product of camphor hydroxylation, 5-*exo*-hydroxycamphor, is then dehydrogenated by 5-*exo*-hydroxycamphor dehydrogenase to form 2,5-diketocamphane, see Figure 1.4.



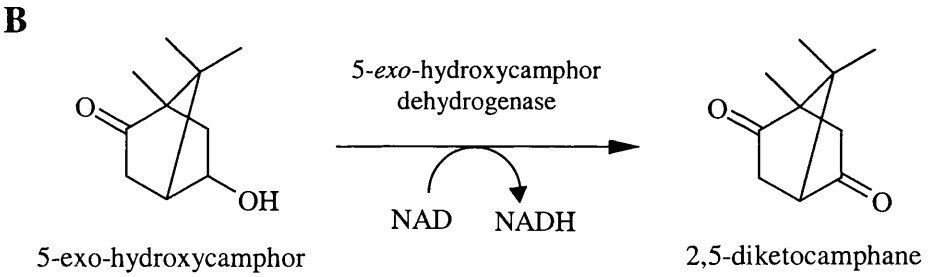


Figure 1.4 The electron cascade of camphor hydroxylation (**A**) (P: cytochrome P450cam; PC : cytochrome P450cam-camphor; PdR: putidaredoxin reductase; and PdX: putidaredoxin) and dehydrogenation of 5-*exo*-hydroxycamphor to form 2,5-diketocamphane by 5-*exo*-hydroxycamphor dehydrogenase (**B**).

1.6 Mechanism of the 2,5-diketonecamphane lactonising reaction

There are two enzymes involved in the lactonisation of 2,5-diketonecamphane, namely 2,5-diketocamphane 1,2-monooxygenase and NADH dehydrogenase. In addition, this lactonizing system requires flavin mononucleotide (FMN), NADH and O_2. The overall electron transfer in the oxygenation of diketocamphane to campholide can be formulated as in Figure 1.5. As can be seen in the Figure, the NADH dehydrogenase reduces NADH to NAD^+ and transports electrons to 2,5-diketonecamphane 1,2-monooxygenase, which carries a flavin mononucleotide (FMN). The 2,5-diketonecamphane 1,2-monooxygenase with reduced FMN then reacts with molecular oxygen (O_2), which forms an enzyme-hydrogen peroxide (enzyme-FMN-OOH), and incorporates one oxygen atom into the 2,5-diketocamphane to form 5-oxo-1,2-campholide lactone. This mechanism has been proposed by Taylor and Trugill (1985).

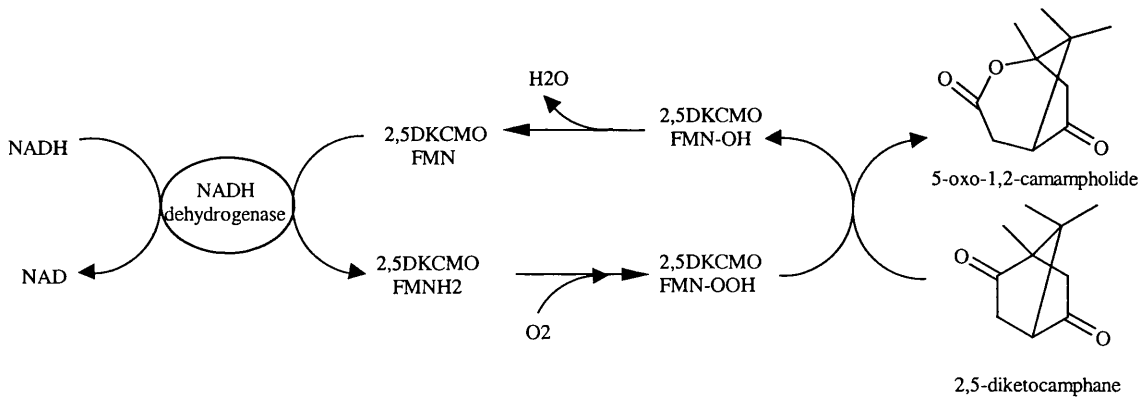


Figure 1.5 Lactonisation of 2,5-diketocamphane to form 5-oxo-1,2-campholide lactone by NADH dehydrogenase and 2,5-diketocamphane 1,2-monooxygenase (2,5-DKCMO).

1.7 Camphor catabolic genes and their enzymes

In 1971, Chakrabarty and Gunsalus studied the physical characteristics of the CAM plasmid. A transduction experiment, which provided information on gene order and genetic distance in the CAM plasmid, was conducted. The experiment involves transferring CAM plasmid from Cam^+ to Cam^- *P. putida* host cells and other Cam^- Ibu^+ (isobutyrate utilisation) *Pseudomonas* species *i.e. P.aeruginosa*, *P. fluorescens* and *P. olevorans*.

The result, the kinetic transfer and physical arrangement of the genes in the CAM plasmid, is shown in Figure 1.6. When the hydroxylase gene (cam100 and cam101) was marked as the first marker and dehydrogenase gene (cam120 and cam121) was marked as the last marker in $cam^+ \times cam^-$ cross, the hydroxylase gene entered at 8 minutes and the dehydrogenase gene at 14 minutes. From the transduction study, it can be established that the *cam* genes specified enzymes for the catabolism of camphor to isobutyrate in the plasmid. However, the genes that code enzymes for anaplerotic and amphibolic metabolism of isobutyrate are thought to be in the chromosome (Rheinwarld *et al.*, 1973).

The investigation of genes and enzymes which are responsible for the early steps of camphor degradation on the CAM plasmid have been previously described (Koga *et al.*, 1989 and Aramaki *et al.*, 1993). The cytochrome P450cam monooxygenase, putidaredoxin reductase, putidaredoxin and 5-*exo*-hydroxycamphor dehydrogenase, the enzymes in the first two steps of camphor degradation, are encoded by *camC*, *camA*, *camB* and *camD* respectively. The genetic arrangement of these genes, in array, is *camDCAB*. The structural genes of *camDCAB* are regulated by CamR repressor protein encoded by *camR*.

According to the sequential camphor degradative pathway, the next review will be on *camC*, *camA*, *camB*, *camD* and *camR* which encode for cytochrome P405cam, putidaredoxin reductase, putidaredoxin, 5-*exo*-hydroxycamphor dehydrogenase and CamR repressor protein respectively.

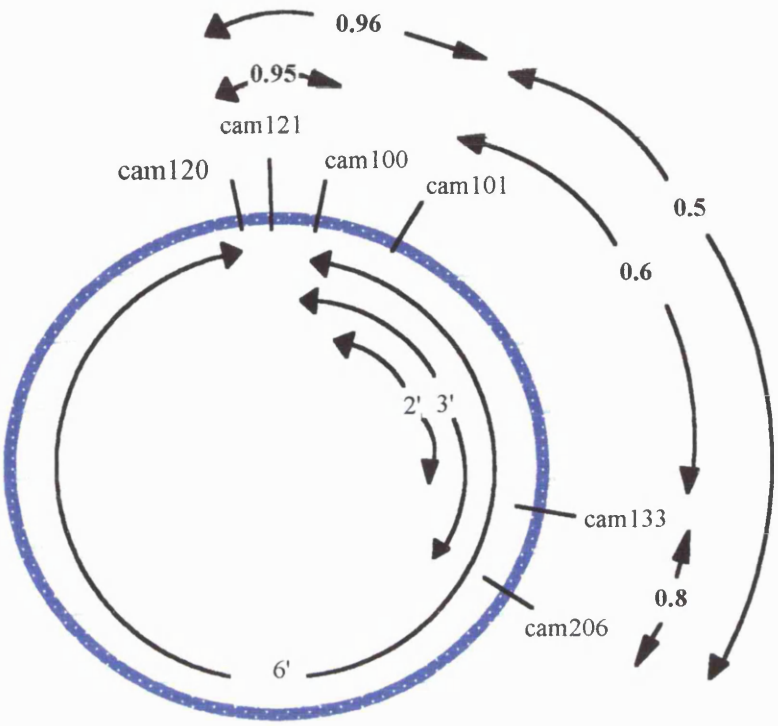


Figure 1.6 The kinetic transfer and physical arrangement of the CAM plasmid (cam100 and cam101: mutations in the hydroxylase genes; cam120 and cam121: mutations in the dehydrogenase genes; cam133: mutations in the ketolactonase genes and cam206: mutations in the genes between ketolactonase genes and isobutyrate genes), where the numbers in bold represent the transductional cotransfer frequencies, indicating the relation between the distance between two selected markers and the size of the transducing fragment in minutes.

1.7.1 Cytochrome P450cam monooxygenase

The first step of camphor degradation by *P. putida* employs the monooxygenase system; putidaredoxin, NADH-putidaredoxin reductase and cytochrome P450cam. Cytochrome P450 is the enzyme in an iron-heam compound, Fe^{2+}-carbonmonoxide complex, which shows an absorbance sprectrum at 450 nm. The interest of the CAM plasmid has been mainly on the P450cam gene. The P450 gene is an important genetic system in microorganisms, animals and human. The cytochrome P450 gene encodes cytochrome P450 monooxygenase which is significant in biosynthesis and in the degradation of a variety of endogenous and xenobiotic compounds. Moreover, the cytochrome P450cam has been studied in considerable detail since this enzyme in bacteria is a soluble protein, compared to mammalian cytochrome P450s, which are membrane bound proteins (Haniu *et al.*, 1982).

The complete 412 amino acid sequence of cytochrome P450cam has been determined by protein sequencing. The molecular mass of the cytochrome P450cam enzyme is reported as 46.8 kDa (Mitsuru *et al.*, 1982).

The first cloning of the P450cam gene was reported by Koga and coworkers in 1985. The gene of cytochrome P450cam was cloned from *P. putida* strain PpG1/pRG1 (*P. putida* ATCC 17453). Briefly, the total DNA of *P. putida* ATCC 17453 was used because they were unable to isolate the CAM plasmid. This genomic DNA was digested with restriction endonuclease *Pst*I. The 2.3 kb *Pst*I fragment of CAM plasmid DNA was ligated into a pKT240 vector and transformed into the mutant strain PpG543, which has a mutation on the pRG1 CAM plasmid and an inability to grow on a camphor plate. A transformant, which grew on a camphor plate supplemented with kanamycin, was selected, and the recombinant plasmid was designated pKG201. The plasmid pKG201 was transformed into the *P. putida* JPS3 to express the P450cam protein. The expression of the *camC* was independent of the presence of camphor as an inducer, which indicated that the 2.3 kb fragment of CAM plasmid DNA-*Pst*I lacks a regulatory gene.

The determination of the nucleotide sequences of the *camC* was documented in detail in 1986 (Unger *et al.*, 1986). The complete nucleotide sequence of the cytochrome P450cam gene contains the G+C content of 59.0% with preferred usage of G and C at the wobble position (72.8%). In contrast, the G+C content of the first and second position is 63.5% and 40.6% respectively. The *camC* has 1245 nucleotides encoding for 414 amino acid residues with a predicted molecular mass of 49.0 kDa. The start and stop codons of *camC* are ATG and TAA respectively. The amino acid sequence from the cloning of this *camC* was also compared to the protein sequences obtained from the purified protein (Haniu *et al.*, 1982). The comparison showed the additional amino acids of Trp and Thr were between Val (54) and Arg (55) (Unger *et al.*, 1986).

A sequence homology search using amino acid sequence of *camC* showed that cytochrome P450cam protein is related to many known and putative cytochrome P450 proteins in the GenBank protein database. Three known cytochrome P450 proteins that share the highest level of homology to cytochrome P450cam in the current database are cytochrome P450 [*Mesorhizobium loti*] (GenBank accession number AP003007), cytochrome *P450cin* [*Citribacter braakii*] (GenBank accession number AF456128) and cytochrome P450meg (steroid 15-beta-monooxygenase) [*Bacillus megaterium*] (GenBank accession number Z21972) with 31%, 27% and 30% identity to cytochrome P450cam respectively.

1.7.2 Putidaredoxin reductase

Two enzymes of a flavoprotein, putidaredoxin reductase and iron-sulfur protein, putidaredoxin, another two component monooxygenase in initial 5-*exo*-hydroxylation of (+)-camphor have previously been studied (Peterson *et al.*, 1990).

The putidaredoxin reductase has a nucleotide sequence of 1215 base pairs. The start and stop codons for the putidaredoxin reductase gene (*camA*) are GTG and TGA respectively. The initiation codon GTG for *camA* is a rare start codon. This rare start

codon is thought to be important in the control of putidaredoxin reductase abundance. A study of site-directed mutagenesis of GTG to ATG showed the expression of putidaredoxin reductase in *E. coli* cells increased 18 fold. The *camA* gene is adjacent to the 5′-end of *camB* (the putidaredoxin gene) and translated to 422 amino acid residues. The molecular mass of putidaredoxin reductase has been reported as approximately 26.0 kDa. The potential FAD and/or NAD-binding site on putidaredoxin reductase has been identified as GGGYIG (GXGXXG) (Peterson *et al.*, 1990). The binding of NADH, an electron donor, to putidaredoxin reductase is required in the hydroxylation reaction of camphor to 5-*exo*-hydroxycamphor.

In the recent BLAST search, the NADH dependent putidaredoxin reductase shares homology with many proteins in the ferredoxin reductase family. The proteins in this family resembling the putidaredoxin reductase are, for example, ferredoxin reductase [*Acinetobacter* sp.] (GenBank accession number AJ311718), ferredoxin reductase [*Caulobaacter crescentus*] (GenBank accession number AE006011) and rhodocoxin reductase, ThcD [*Rhodococcus erythropolis*] (GenBank accession number U17130). These three reductases are highly related to the putidaredoxin reductase, with 44%, 43% and 41% identity respectively.

17.3 Putidaredoxin

Tanaka (1974) has determined the amino acid sequence of putidaredoxin from the isolated protein which contains 106 amino acid residues and, interestingly, it appears to be 37.2% homologous to the sequence of adrenodoxin, i.e. a steroid hydrolase in the adrenal cortex.

The putidaredoxin gene obtained by DNA cloning was published many years later (Koga and *et al.*, 1989 and Peterson *et al.*, 1990). The location of *camB* is at the 3′end of the *camA* gene. The putidaredoxin gene (*camB*) is 321 base pair long and encodes for 106 amino acids, it is the smallest polypeptide of the *cam* operon proteins. The start and

stop codons of *camB* are ATG and TAA respectively. The molecular mass of the putidaredoxin is 11.7 kDa. DNA sequence analysis showed a transcriptional termination sequence of CCGGGCTCCAAGCAAGGAGCCCGGAATCTCTC. The sequence of GCCCG followed by a short AT string and then a TCTC, which would form a stem-loop, rho factor independent transcription termination, presented at the end of *camB*. These sequences are similar to the transcriptional termination sequence in *E. coli* genes which has a CGGGC preceded by a short AT string and TCTG. A comparison of both polypeptide sequences of *camB* determined by isolated protein and deduced amino acid sequences of cloned *camB* showed complete agreement.

The putidaredoxin is closely related to the proteins in the feredoxin family. Three of the most highly related ferredoxins to putidaredoxin from the current GenBank database are ferredoxin [*Rhodobacter capsulatus*] (GenBank accession number Y11304), ferredoxin, 2Fe-2S [*Caulobacter crescentus*] (GenBank accession number AE006011) and ferrdoxin VI, FdVI [*Rhodobacter capsulatus*] (GenBank accession number S45612), with 42%, 45% and 41% identity respectively.

1.7.4 5-exo-hydroxycamphor dehydrogenase

The first three enzymes described previously are the enzymes specify for the first step of camphor hydroxylation. For the second step of camphor degradation, 5-*exo*-hydroxycamphor dehydrogenase is the enzyme catalysing 5-*exo*-hydroxycamphor to form 2,5-diketocamphane. This terpene alcohol dehydrogenase is responsible for the dehydrogenation of the hydroxyl group at the carbon position 5 of 5-*exo*-hydroxycamphor, requiring NAD, yielding 2,5-diketocamphane and NADH.

The gene *camD,* which encodes the 5-*exo*-hydroxycamphor dehydrogenase enzyme, is also carried on the CAM plasmid. It encodes 1083 nucleotides located upstream, adjacent to the 5′ end of *camC*, of the gene cluster of *camCAB* (see Figure

1.7). The transcription direction of the *camD* is from left to right and the transcript extends through *camC*, *camA* and *camB*.

The nucleotide sequence of 1086 base pairs of *camD* encodes a 361 amino acid protein. The 5-*exo*-hydroxycamphor dehydrogenase has a molecular mass of 38.4 kDa. The start and stop codons of the *camD* are ATG and TAG respectively. The promoter sequences of TTGACC (-35) and TATGCT and ribosome binding site (ACGAG) were also identified in the upstream region of the *camD*. The G+C content of *camD* has been determined as 62%, with preferential usage of G and C at the wobble position (Aramaki *et al.*, 1993).

The active-site zinc ligands of Cys (140), His (62) and Cys (158) and second zinc ligands of Cys (98), Cys (101), Cys (104) and Cys (112) on 5-*exo*-hydroxycamphor dehydrogenase have also been identified. These amino acid residues of the active-site zinc ligands and second zinc ligands are highly conserved in the 5-*exo*-hydroxycamphor dehydrogenase when compared with other alcohol dehydrogenases. Aramaki (1993) also identified a NAD-binding site of GAGPVG which interacts with the coenzyme NAD required in the dehydrogenation of 5-*exo*-hydroxycamphor to 2,5-diketocamphane by 5-*exo*-hydroxycamphor dehydrogenase.

The 5-*exo*-camphordehydrogenase of *camD* is closely related to other proteins in the dehydrogenase family. An alcohol dehydrogenase [*Thermotoga maritima*] (GenBank accession number AE001722), alcohol dehydrogenase class III [*Pseudomonas aeroginosa*] (GenBank accession number AE004784) and sorbitol dehydrogenase [*Homo sapiens*] (GenBank accession number BC025295) are the most related dehydrogenases to 5-*exo*-hydroxycamphor dehydrogenase, in a recent database search, sharing 31%, 29% and 30% identity to 5-*exo*-hydroxycamphor dehydrogenase respectively.

1.7.5 CamR repressor protein

Gene regulation is the crucial manoeuvring of every cell to control whether or not genes are expressed. Gene regulations correspond to changes in the environment and the demands of the cell in order to control and balance the mechanisms in the cells.

Mostly, a substrate or final product is the key to determining the regulation mechanism of genes. In a degradative pathway, the regulation of a gene is usually achieved by a substrate; however, the regulation can be either under negative or positive control. In contrast, in a biosynthetic pathway, the genes are regulated by a final product, and the regulation is under negative control.

In *Pseudomonas*, the regulatory mechanisms in various degradative pathways have been studied and characterised in detail (Nakazawa *et al.*, 1996); for example the regulatory mechanism in *P. putida*, the regulator *xylR*, which activates the expression of the TOL plasmid upper pathway; the regulator *xylS*, which activate the expression of the TOL lower pathway; and the regulator *camR*, which represses the expression of the *camDCAB* operon of the CAM plasmid.

The gene that encodes the CamR repressor is *camR*, which is the first gene upstream of the *cam* operon (see also Figure 1.7). The *camR* consists of 1134 nucleotides, which encodes a 180 amino acid protein with a molecular mass of 20.4 kDa. This repressor protein is transcribed in the opposite direction to the *camDCAB* operon. The expression of *camR* is subjected to autoregulation, however, the expression and regulation of the *camR* also depend on camphor (Koga *et al.*, 1986; Aramaki *et al.*, 1993; Fujita *et al.*, 1993 and Aramaki *et al.*, 1995). In the presence of camphor, it acts as an inactivator of the CamR repressor; it binds and inactivates the CamR repressor, which then allows cytochrome P450cam hydroxylase operon (*camDCAB*) to be transcribed (see Figure 1.7).

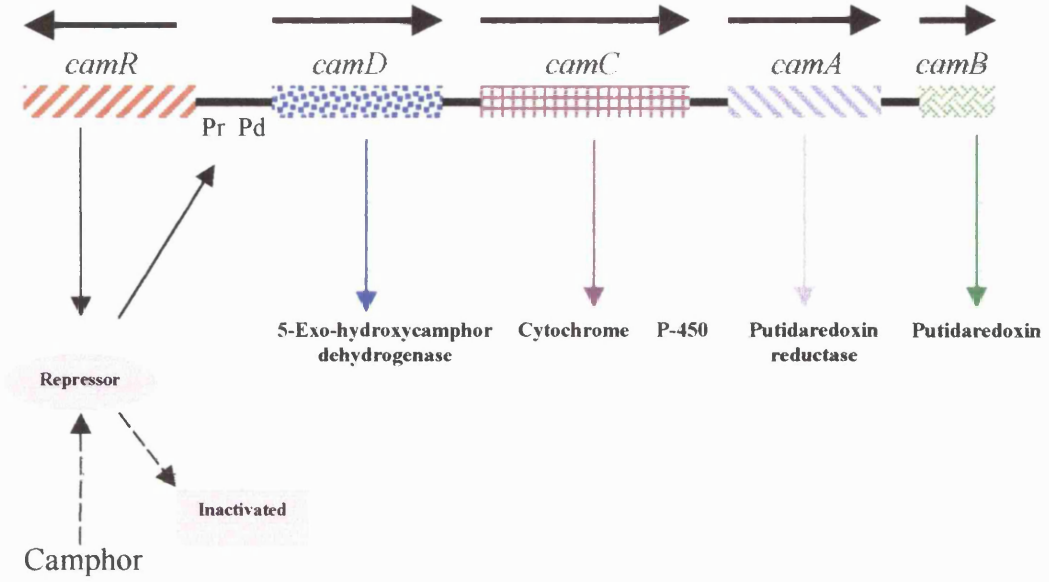


Figure 1.7 Location of *camRDCAB* and their translated proteins (Pd: Promoter for *camDCAB* operon and Pr: Promoter for *cam R* gene). *camR* encodes the repressor binding to the promoter region which regulates *camR* and the *camDCAB* promoter. As a result of the CamR repressor binding, it inhibits *camR* and *camDCAB* expression. In the presence of camphor, it acts as an inactivator to inactivate the camR repressor, which consequently can not bind to the CamR promoter, and allows *camDCAB* to be transcribed.

In detailed DNA sequence analysis of the region between *camR* and *camD*, it has been found that both the promoter of *camR* (Pr) and the promoter of *camDCAB* (Pd) are located on this region. Fujita (1993) found that the *camR* promoter has putative hexanucleotides of the –35 and –10 promoter regions as TTGACC and TATGCT. For *camDCAB*, the RNApolymerase binding site has hexanucleotides as TTGTTC (–35 region) and TCATAT (-10 region). Moreover, the study showed the palindromic sequence (ATATAGCGGCTATAT) of the common operator overlapping both the Pd and Pr promoters. Using DNA footprinting technique showed that the region which is bound by CamR protein is protected from DNaseI attack, and the CamR covered almost all the Pd and Pr regions (Fujita *et al.*, 1993) (also see Figure 1.8).

The repressor protein of CamR has a conserved domain of the bacterial TetR repressor family and shares an identity with many putative transcription regulators in the TetR family. A BLAST search showed that three highly identical regulatory proteins related to CamR repressor are the transcriptional regulator, the TetR family [*Caulobacter crescentus*] (GenBank accession number AE005796), the transcriptional regulator, the TetR family [*Deinococcus radiodurans*] (GenBank accession number AE002049) and the TetR-family transcriptional regulator protein [*Streptomyces avermitilis*] (GenBank accession number AB070948), which have 25%, 30% and 29% identity to the CamR repressor respectively.

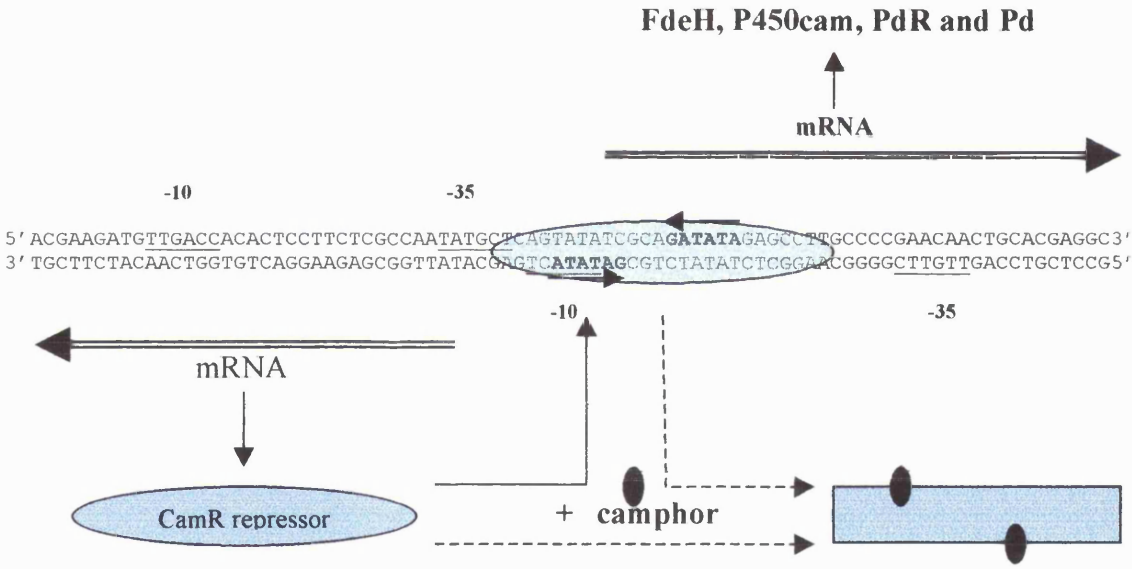


Figure 1.8. Nucleotide sequences in the cam operon controlling region. The arrows on the DNA sequences represent the paralindromic DNA sequences (-35 and –10 promoter regions of *camR* and *camD* are underlined; FdeH: 5-*exo*-hydroxycamphor dehydrogenase; P450cam : cytochrome P450cam; Pdr: putidaredoxin reductase; Pd: putidaredoxin). In the absence of camphor, the CamR repressor binds to the palindromic region preventing the expression of *camR* and *camDCAB*.

1.8 2,5- and 3,6-Diketocamphane monooxygenase

Two parallel enantiomeric specific enzymes are responsible for the third step of (+) and (-)-camphor degradation, the metabolism of diketocamphanes to campholides. These two enzymes are 2,5-diketocamphane 1,2-monooxygenase and 3,6-diketocamphane 1,6-monooxygenase, which are responsible for 2,5-diketocamphane and 3,6-diketocamphane oxygenation respectively (Figure 1.9). Both 2,5 and 3,6-diketonecampane monooxygenases are classified as Baeyer-Villiger monooxygenases; the enzymes catalyse the insertion of oxygen into ketones to give esters or lactones. The biochemical and immunological characteristics of 2,5-diketocamphane 1,2-monooxygenase and 3,6-diketocamphane 1,6-monooxygenase have been studied. The results of the studies show these two enzymes have very similar properties.

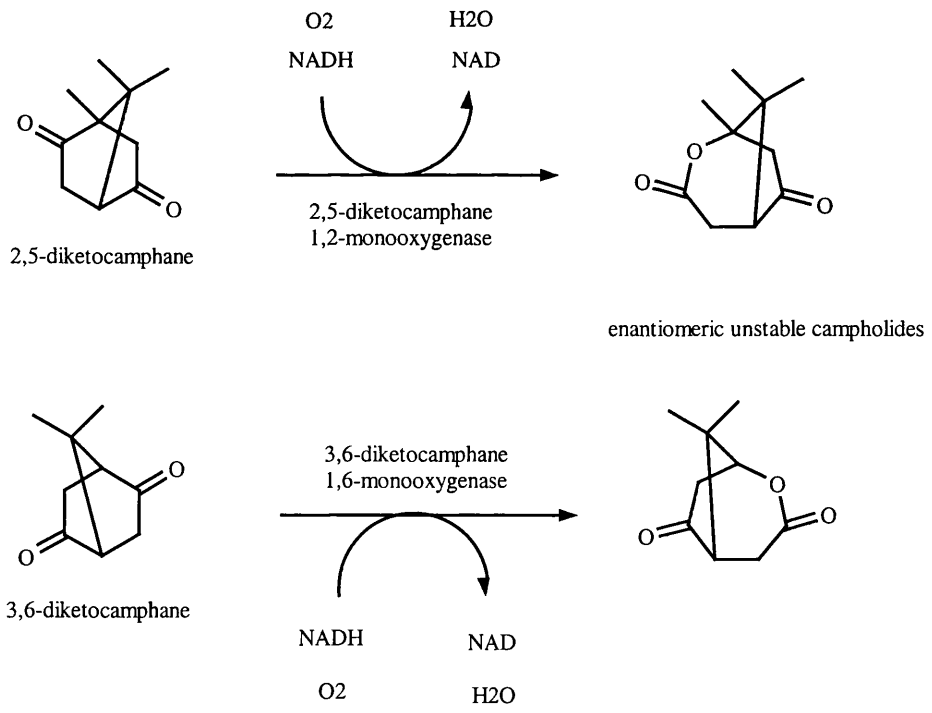


Figure 1.9 The oxygenating reaction of 2,5 and 3,6-diketocamphane by 2,5 and 3,6-diketocamphane monooxygenase in *P. putida* NCIMB 10007.

1.8.1 2,5-diketocamphane 1,2-monooxygenase

The early study of the 2,5-diketocamphane 1,2-monooxygenase suggested that the enzyme consists of three protein components: two oxygenating components of NADH oxygenase and a bipyridyl-sensitive oxygenating flavoprotein (Yu and Gunsalus, 1969). The more precisely defined study of the 2,5-diketocamphane 1,2-monooxygenase complex by Taylor and Trudgill (1986) shows that the enzyme is 78.0 kDa, consisting of two subunits of almost equal molecular weight. This active complex is composed of 2,5-diketocamphane 1,2-monooxygenase, FMN-binding enzyme, and NADH dehydrogenase. The oxygenating component is 2,5-diketocamphane 1,2-monooxygenase, which has a molecular mass of approximately 39.0-40.0 kDa. The NADH dehydrogenase has an estimated molecular mass of 36.0 kDa. This has confirmed the original report of the 2,5-diketocamphane 1,2-monooxygenase, which consists of two component (Conrad *et al.*, 1965a,b). The absorbance spectrum of 2,5-diketonecamphane 1,2-monooxygenase and the protein FMN-bound complex displays the spectrum maximum at 274, 372 and 445 nm, see detail Figure 1.10. Conrad (1965) demonstrated that when a 2,2′-bipyridyl, chelating agent of Fe, was added to the solution containing FMN-bound 2,5-diketonecamphane 1,2-monooxygenase, it resulted in dissociation of the bound flavin. As a result, Fe was thought to be involved in the catalytic reaction of the oxygenation of 2,5-diketocamphane. However, this suggestion was challenged by Taylor and Trudgill 's study (1986), which did not shown that Fe is involved in the catalytic reaction of oxygen insertion in 2,5-diketocamphane. Taylor and Trudgill also proposed the mechanism of oxygen insertion of 2,5-diketocamphane (as described previously Figure 1.5).

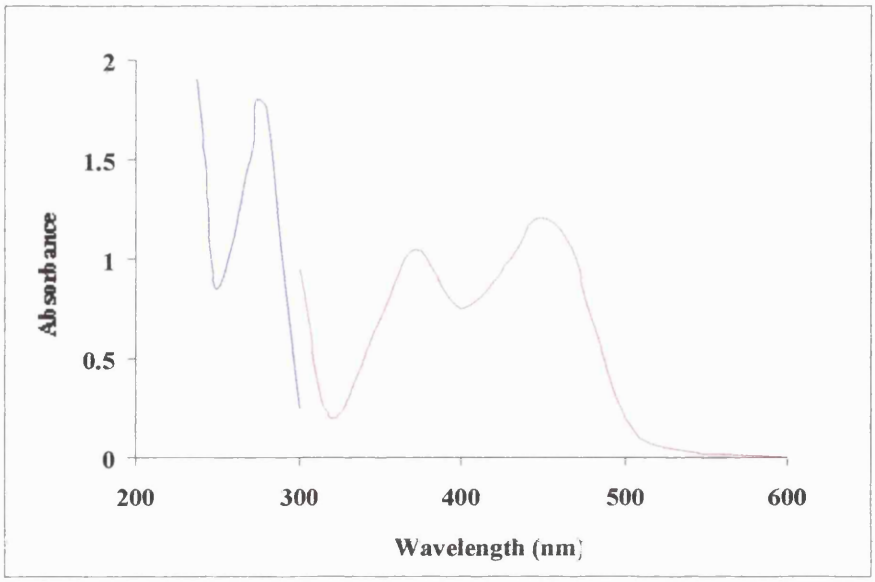


Figure 1.10 Absorbance spectrum of 2,5-diketocamphane 1,2-monooxygenase apoenzyme and its FMN-bound complex. A cuvette with a 1-cm light path contained 1.01 mg of protein in 1 ml of 42 mM Na^+, K^+ phosphate buffer (adapted from Taylor and Trudgill, 1986).

1.8.2 3,6-diketocamphane 1,6-monooxygenase

Only one literature that studied biochemical characteristics of 3,6-diketocamphane monooxygenase in detail, by Jones and coworkers (1993). The molecular mass of the apoenzyme of the 3,6-diketocamphane 1,6-monooxygenase is 76.0 kDa, which consists of an oxygenating subunit and NADH dehydrogenasae. The oxygenating subunit of 3,6-diketocamphane monooxygenase has a molecular mass of 39.5-41.1 kDa, forming a very lose complex to NADH dehydrogenase (molecular mass of about 36,000 daltons) and a prosthetic cofactor of FMN (Jones *et al.*, 1993). The pI value for 3,6-diketocamphane 1,6-monooxygenase is 5.5. The purified enzyme of 3,6-diketocamphane 1,6-monooxygenase exhibited a spectrometric spectrum of λ_{max} at 282 nm, distinct from 2,5-diketonecamphane 1,2-monooxygenase, which exhibited a spectrometric spectrum of a λ_{max} at 274 nm. Like 2,5-diketocamphane 1,2-monooxygenase complex, NADH

dehydrogenase forms a complex with 3,6-diketocamphane 1,6-monooxygenase and function as an electron donor in the catalytic reaction of oxygen insertion of 3,6-diketocamphane.

Although the 3,6-diketocamphane 1,6-monooxygenase is closely related to 2,5-diketocamphane 1,2-monooxygenase in the size, it is translated by different genes (Jones *et al.*, 1993). The 2,5-diketocamphane 1,2-monooxygenase has been shown to be inactive on (-)-camphor, and genetic studies using mutants show that an isofuntional enzyme, 3,6-diketocamphane 1,6-monooxygenases is specific for (-)-camphor degradation (Jones *et al.*, 1993). Both 2,5-and 3,6-diketocamphane monooxygenase, specific biological Baeyer-Villiger monooxygenases in the third step of camphor degradation, have very similar aspects; each uses FMN as a coenzyme and both are able to make use of the same NADH dehydrogenase.

Recently, the N-terminal sequencesof 2,5 and 3,6-diketocamphane monooxygenase from *P. putida* NCIMB 10007 have been determined (Kelly *et al.*, 1996). By using the N-terminal sequences of 2,5- and 3,6-diketocamphane monooxygenase, it is possible to show that these two monooxygenases are related to luciferase proteins, LuxA and LuxB, of *Vibrio harveyi* and *Vibrio fischeri*, which are FMN-binding proteins (Willetts, 1997) (Figure 1.11).

```
LUXB_Har_      -MKFGLFFLNFMFSKR--SS---------------- 17
LUXB_Fis_      -MKFGLFFLN--FQKD--GI---------------- 15
LUXA_Har_      -MKFGNFLLT--YQPPELSQTEVMKRLVNLG----- 28
LUXA_Fis_      -MKFGNISFS--YQP---SG-E-------------- 15
2,5-DKCMO      -MQAGFFGTP--YDLPTRTAR-----------QM-- 20
3,6-DKCMO      AMETGLIFHP--YMRPGRSARGTFDWGIKSAVQADR 34
                *: * :     :
```

Figure 1.11 Multiple alignment of the N-terminal of 2,5-DKCM, 3,6-DKCMO, *LuxA* and *LuxB* protein (LUXB_Har: LuxB from *V. harveyi*; LUXB_Fis: LuxB from *V. Fischeri*; LuxA_Har: LuxA from *V. harveyi*; LuxA_Fis: LuxA from *V. Fischeri*; 2,5-DKCMO: the N-terminal of 2,5-diketocamphane 1,2-monooxygenase and 3,6-DKCMO: the N-terminal of 3,6-diketocamphane 1,6-monooxygenase) aligned by CLUSTALW program. Identical amino acids are shaded; and dashes are gaps .

1.9 2-oxo-Δ^3-4,5,5-trimethylcyclopentanylacetyl-CoA synthetase and 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygenase

2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA synthetase, ATP dependent, is the enzyme that catalyses the formation of a 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA, a combined molecule between a 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetic acid (VII) and coenzyme A (3-phosphoadenosine-5-diphospho-4-pantetheine) in the fourth enzymatic step of camphor catabolism (see Figure 1.12). This reaction as well as the further oxygenation of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA by a 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygenase has been demonstrated by Ougham (1983).

The 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygenase is another monooxygenase involved in the oxidation of 2-oxo-Δ^3-4,5,5-trimethylcyclopetanylacetyl-CoA to form 5-hydroxy-3,4,4-trimethyl-Δ^2-pimelyl-CoA-δ-lactone. This monooxygenase inserts an oxygen atom between carbon position 3 and 4 of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA (Ougham *et al.*, 1983) (see also Figure 1.12). This acetyl CoA monooxygenase is distinct from the 2,5 and 3,6-diketocamphane monooxygenase enzymes in both substrate specificity and coenzyme requirement. The acetyl CoA monooxygenase has specificity for 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA as a substrate, and it uses a NADPH coenzyme as an electron donor.

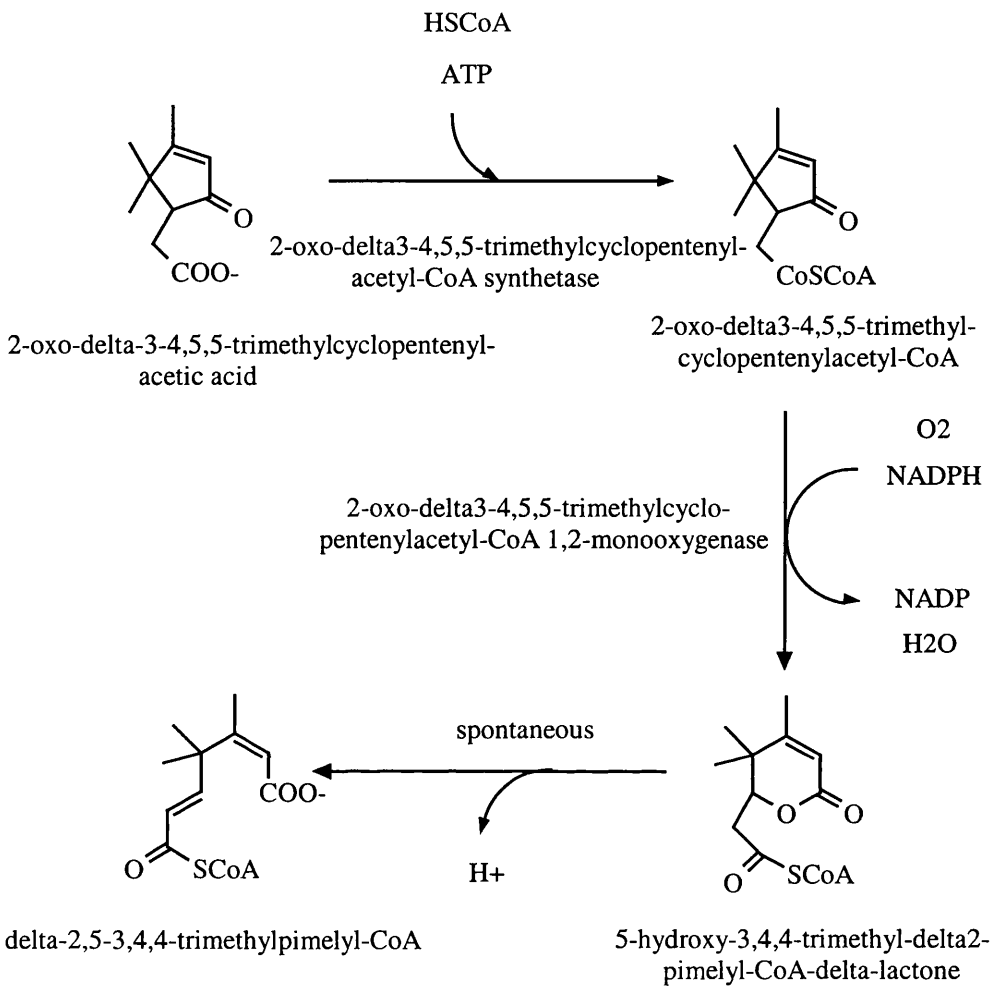


Figure 1.12 The metabolism of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetic acid by CoA ester synthetase and acetyl-CoA monooxygenase in *P. putida* NCIMB 10007.

The native enzyme of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygenase has a molecular mass of 106.0 kDa and is composed of two identical subunits with a molecular mass of 56.0 kDa each (Ougham *et al.*, 1983). The absorption spectrum of the purified enzyme showed a maximum peak at 274, 378 and 438 nm, indicating this monooxygenase is a flavin associated enzyme and is another Baeyer-Villiger monooxygenase. The addition of NADPH as an electron donor to the purified

enzyme in the oxygenation of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA gave activity of 100% (Ougham *et al.*, 1983). After the ring-oxygen insertion reaction of 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA to form 5-hydroxy-3,4,4-trimethyl-Δ^2-pimelyl-CoA-δ-lactone, the hydrolysis of this lactone took place to form $\Delta^{2,5}$-3,4,4-trimethylpimelyl-CoA spontaneously. However, studies of the genes encoding the CoA synthtase and CoA monooxygenase have not been reported. Up-to-date, the investigation of the camphor degradation pathway has been confined to these steps.

For a summary of the known genes and enzymes responsible for the steps in the camphor degradation pathway by *P. putida* NCIMB10007, see Table1.1.

Table 1.1 Summary of the known genes and enzymes responsible for the steps in camphor degradation in *P. putida* NCIMB10007.

Gene	Enzyme	Basepairs	Amino acids	Mw (kDa)	Regulation	Function	Cofactor
camD	5-*Exo*-camphane dehydrogenase	1083	361	38.4	By camR repressor	3-*Exo*-camphor to 2,5 or 3,6-diketocamphane	NAD
camC	Cytochrome P450cam	1245	414	46.8	By camR repressor	Camphor to 3-*exo*-hydroxycamphane	Haem
cam A	NADH-putidaredoxin reductase	1266	422	45.57	By camR repressor	Camphor to 3-*exo*-hydroxycamphane	NADH
cam B	Putidaredoxin	318	106	11.49	By camR repressor	Camphor to 3-*exo*-hydroxycamphane	-
cam R	CamR repressor	558	186	20.4	Auto regulated (camphor is inhibitor)	*CamDCAB* regulation	-
cam?	2-Oxo-Δ^3-4,5,5-trimethylcyclopentanyl acetyl-CoA synthetase	Unknown	Unknown	Unknown	Unknown	Cyclopentenylacetate to cyclopentenylacetyl-CoA	HSCoA and ATP
cam?	2-Oxo-Δ^3-4,5,5-trimethylcyclopetanyl-acetyl-CoA 1,2-monooxygenase	Unknown	Unknown	56.0	Unknown	Cyclopentenylacetyl-coA to pimelyl-CoA-δ-lactone	NADPH
cam?	2,5-Diketocamphane 1,2-monooxygenase	Unknown	Unknown	38.0	Unknown	2,5-Diketocamphane to 5-oxo-1,2-campholide	NADH
cam?	3,6-Diketocamphane 1,6-monooxygenase	Unknown	Unknown	36.0	Unknown	3,6-Diketocamphane to unstable lactone	NADH

cam? ; the gene not known

1.10 Baeyer-Villiger monooxygenases and biotransformation

Monooxygenases are enzymes that incorporate one oxygen atom from dioxygen into a substrate, with the remaining oxygen atom converted into H_2O. The Baeyer-Villiger monooxygenases are monooxygenase enzymes which catalyse the oxygen insertion of C-C bonds beside ketone groups. The oxygenation by monooxygenases usually requires NADH or NADPH as a coenzyme during the reaction. The oxygenation reaction is very interesting especially within non-activated compounds whose oxygenated forms may be potentially very useful.

Biotransformation is defined as a process that is exploited to convert a chemical compound by using biological systems. Benefits of biotransformation are, for instance, in the pharmaceutical industry and in organic synthesis.

In the pharmaceutical industry, the production of a single enantiomer is crucial. Adverse side effects and toxicity of a drug may occur if another enantiomer is included. In any application of organic synthesis in which regio- or enantiospecificity is desired, biotransformation techniques play an important role in producing a variety of organic compounds. The advantages of biotransformation are not only that regio and enantiomeric specific molecules are obtained; it may also lowers the cost of manufacture.

The organic synthesis industry has shown a growing trend in using biotransformation, especially in amino acid, steroid and antibiotic manufacture (Grogan *et al.*, 1992; Grogan *et al.*, 1993; Gagnon *et al.*, 1995; Furstoss and Petit, 1995 and Kelly *et al.*, 1998). The biotransformation are performed in mild conditions and considered to be environmentally friendly (Hanson, 1995).

In extending the range of useful biotransformations, oxygenation by Baeyer-Villiger monooxygenases has been employed in organic synthesis. These monooxygenases can be used to perform chemeoenzymatic synthesis on various chemicals. For example, cyclohexanone monooxygenase (CHMO) from *Acinetobacter*

calcoaceticus is a NADPH dependent Baeyer-Villiger monooxygenase. This CHMO has been used in the organic synthesis of many useful intermediate compounds. By CHMO, the biotransformation of ketone, bicyclo (3.2.0) hept-2-en-6-one yeilds the chiral syntons, which can subsequently enter a variety of routes to chemoenzymatic synthesis of valuable compounds such as pheromones, antibiotic sarkomycin A, viridine, ionomycin and derivatives with interesting antileukaemic activities (Willetts, 1997). Also, employing the oxygen insertion reaction to the ketone of bicyclo (3.2.0) hept-2-en-6-one can be a potential route to the synthesis of prostaglandins and nucleosides (Newton and Roberts, 1980). Moreover, using either cyclopentanone monooxygenase or 2-oxo-4,5,5-trimethylcyclopentenylacetyl-CoA monooxygenase in the chemoenzymatic synthesis of the intermeadiate (S)-methyl 6,8-dihydroxyoctanoate can lead to the production of (+)-(*R*)-lipoic acid (Adger *et al.*, 1997). In ketone synthesis, *Acenetobacter calcoaceticus* has been used to oxidise dihalogenoketone, yielding optically pure ketone (40%) that has been oxidised to lactone by chemical process and further steps to an AZT analogue (Roberts and Willetts, 1993) (see Figure 1.14). However, there have to be clear benefits to employing the CHMO and cyclopentanone monooxygenase, because of the cost and cofactor involvements since the CHMO and cyclopentanone monooxygenase utilise NADPH which is expensive and difficult to recycle (Roberts and Willetts, 1993). This introduces the thought of employing an alternative enzyme which might lower the cost and be put to practical use.

The diketocamphane monooxygenases from the camphor degradation pathway utilising NADH as a cofactor can be useful in organic synthesis compared to the monooxygenases CHMO (Roberts and Willetes, 1993). There are at least three Baeyer-Villiger monooxygenases in the camphor degradation pathway: 2,5-diketocamphane 1,2-monooxygenase, 3,6-diketonecamphane 1,6-monooxygenase and 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA monooxygenase. Pilot studies using these Baeyer-Villiger monooxygenases in the biotransformation of chemical compounds have been carried out. Preliminary studies of the Baeyer-Villiger monooxygenases in microbial transformation using whole cells of *P. putida* NCIMB10007 which carried the CAM plasmid, were used to transform a ketone into two isomeric optically active lactones

(Roberts and Willetts, 1993) (Figure 1.13). The products of this enzymatic reaction represent a mirror image of cyclohexanone monooxygenase (CHMO) biotransformation of the same substrate.

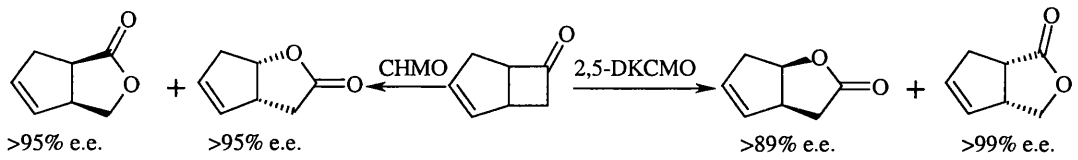


Figure 1.13 Biotransformation of bicyclo (3.2.0) hept-2-en-6-one by cyclohexanone monooxygenase (CHMO) and 2,5-diketocamphane 1,2-monooxygenase (2,5-DKCMO) (Willetts, 1997)

Therefore, diketocamphane monooxygenases may hold great promise and can be considered as alternatives to existing Baeyer-Villiger monooxygenases in biotransformation to produce useful compounds.

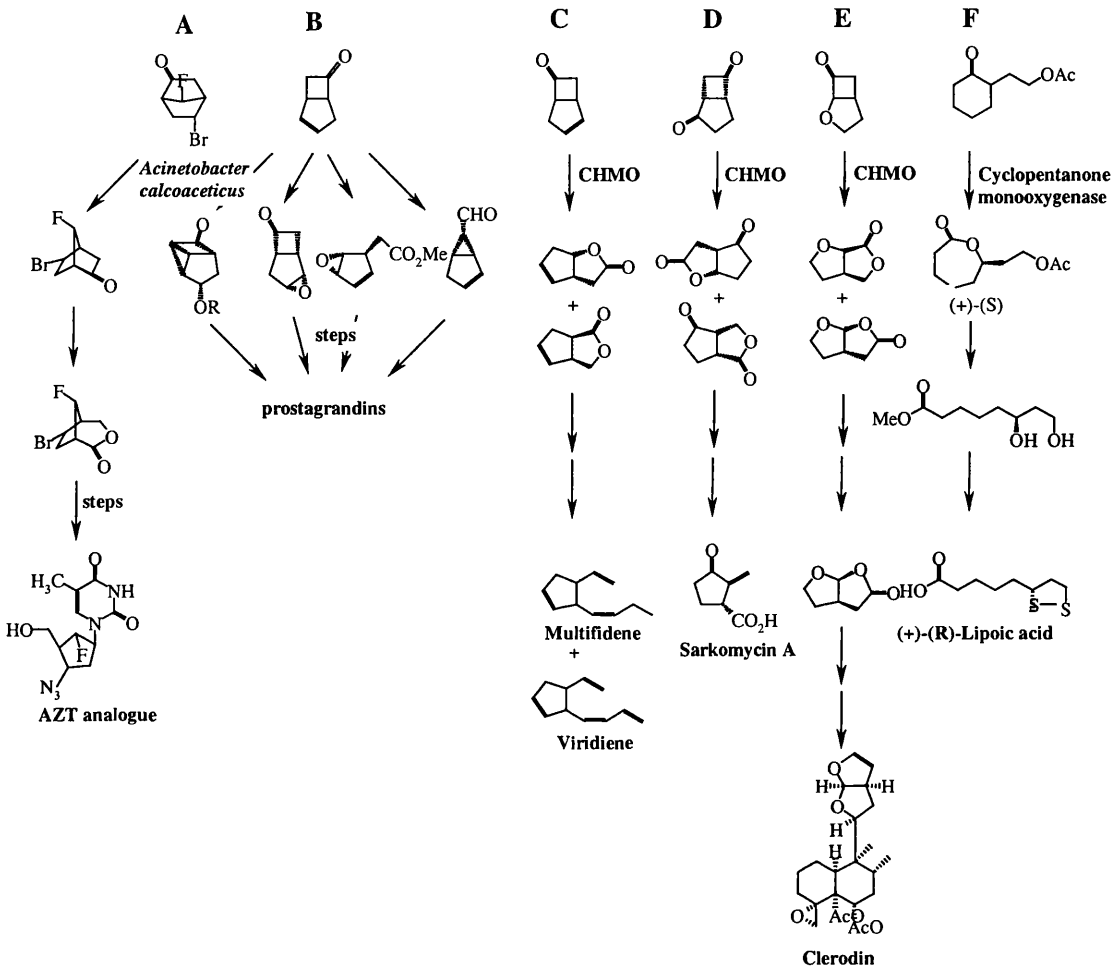


Figure 1.14 Chemoenzymatic approaches that employ Baeyer-Villiger monooxygenases to yield target molecules in order to use these intermediates in the synthesis of useful compounds (Newton and Roberts, 1980; Willetts, 1997; and Adger *et al.*, 1997).

A: Biotransformation of dihalogenoketone by *A. calcoaceticus* strain NCIMB 9871, which gave lactone that can be used in the synthesis of an AZT analogue. **B**: Oxygenation of bicyclo[3.2.0]hept-2-en-6-one to compounds that can be used in preparing prostraglandins. **C**, **D** and **E**: Biotransformation of bicyclo[3.2.0]hept-2en-6-one by cyclohexanonone monooxygenase in *A. calcoaceticus* which yields intermediate molecules that can be useful in the organic synthesis of pheromones (multifidene and viridine), antibiotic (sarkomycin A) and insect antifeedant (clerodin).

Figure 1.14 (continued). **F**: The use of cyclopentanone monooxygense in the chemoenzymetic synthesis of the intermediate of (S)-methyl 6,8-dihydroxyoctanoate to prepare (+)-(*R*)-lipoic acid.

1.11 Phosphotriesterase (parathion hydrolase)

Hydrolases are enzymes which catalyse hydrolysis, the rupture of chemical bonds involving the addition of atoms from water, in several compounds such as ester, amides, nitriles, epoxides and anhydrides. One of the hydrolases is phosphotriesterase, also known by its substrate names parathion hydrolase, paraoxonase and parathion aryl esterase. It catalyses the hydrolysis of organophosphorous compounds, and is an interesting enzyme in the detoxification of organophosphorous compounds. These compounds are chemical substances containing the phosphoric group, usually used as agricultural insecticides, pesticides and include organophosphorous compounds developed as nerve gases. The members of the organophosphorous compounds include parathion (*O*,*O*-diethyl-*O*-p-nitrophenyl phosphorothioate), methyl parathion (*O*,*O*-dimethyl-*O*-p-nitrophenyl phosphorothioate), coumaphos (*O*,*O*-diethyl-*O*-(3-chloro-4methyl-2-oxo-2*H*-1-benzophos) and sarin (*O*-isopropyl methylphosphonofluoridate). These compounds inhibit the activity of acetylcholinesterase, a crucial enzyme for nerve transmission in insects, animals, and humans. The increasing use of these organophosphorous compounds in agriculture causes concern about environmental contamination. The extensive use of such compounds has left them in the environment and they can contaminate agricultural products, soil, and water resources. In the United States alone more than 40 million kilograms of organophosphorous pesticides have been used (Food and Agricultural Organisation (FAO), 1989). One of the organophosphorous compounds, which is widely used as an insecticide in the U.S., is parathion.

Parathion is believed to be degraded by sunlight, however, the result of this biodegradation is paraoxon, which is more toxic than the parent compound. Degradation of parathion by soil bacteria has also been reported. The first bacteria phosphotriesterase

to catalyse the hydrolysis of parathion in *Flavobacterium* sp. strain ATCC 27551 was reported by Sethunathan and Yoshida in 1973. The second strain of parathion degrading bacteria, *P. diminuta* GM, was reported by Mulbry in 1986. The first strain was isolated from the Philippines and the second strain from the United States. The degradation mechanisms of parathion, methyl parathion and paraoxon are shown in Figure 1.15.

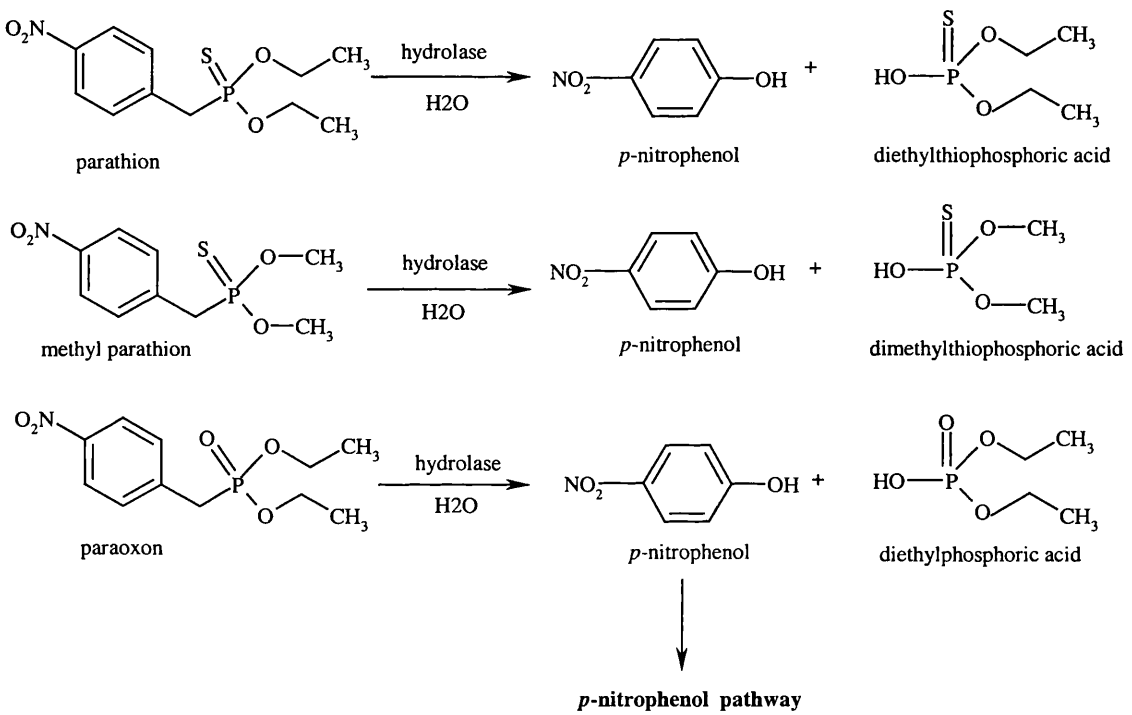


Figure 1.15 The hydrolysis of parathion methyl parathion and paraoxon.

It has been proposed that the parathion and paraoxon compounds are cleaved by phosphotriesterase to form *p*-nitrophenol and phosphoric acid compounds. The *p*-nitrophenol can be degraded by other microorganisms via *p*-nitrophenol pathway.

1.12 Phosphotriesterase encoded by the plasmid-borne gene (*opd*) of *Pseudomonas diminuta* MG

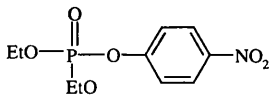
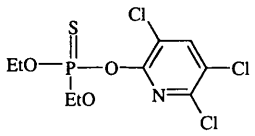
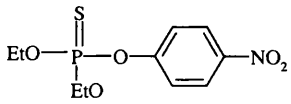
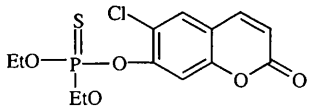
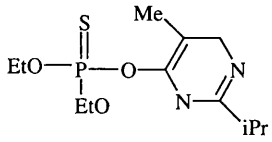
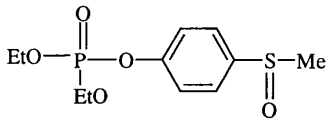
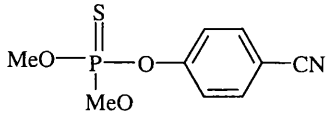
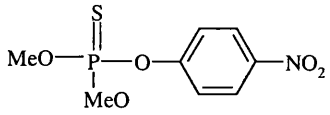
The phosphotriesterase gene in *P. diminuta* MG encods an enzyme that catalyses the hydrolysis of a broad range of organophosphorous compounds and was isolated in 1982. This plasmid-borne gene is on a 66-kb plasmid and has been defined as *opd*, which strands for **o**rgano**p**hosphhate **d**egradation. This opd gene is 978 nucleotides in lenght and codes for a protein of 325 amino acid residues with a molecular mass of 35.4 kDa. Though the phosphotriesterase has been predicted to have a molecular weight of 35.4 kDa, the active enzyme suggests that the enzyme is a dimer of two individual monomers to form the apoenzyme (McDaniel *et al.*, 1988). The hydrolysis of substrate by the membrane fraction of *P. diminuta* MG has demonstrated that this phosphotriesterase is membrane-associated protein. By isolating the membrane fractions of *P. diminuta* MG, the study showed that 80-90% of the enzyme activity was associated with the membrane fractions.

The phosphotriesterase from *P. diminuta* MG was purified by Dumas in 1989. However, in this study Dumas showed this hydrolase enzyme has an approximate molecular mass of 39.0 kDa on SDS-polyacrylamide gel, slightly different from the previous study. The enzymatic activity of the phosphotriesteras is relatively high toward paraoxon with the kinetic rate constant, K_{cat} and K_{cat}/K_m of 2,100 S^{-1} and 4×10^7 M^{-1} S^{-1} respectively. This indicated that the enzyme has substantial substrate specificity to paraoxon. In addition, the phosphotriesterase was found to hydrolyse the other organophosphorous insecticides; diazinon, dursban, coumaphos, cyanophos, fensulfothion, methyl parathion and parathion, see Table 1.2 (Dumas *et al.*, 1989). This study also indicated the preference of phosphotriesterase from *P. diminuta* MG in hydrolysing the organophosphorous P-O bond.

It has been found that the phosphotriesterase of *P. diminuta* MG has 1 mole of zinc bound to 1 mole of enzyme. The addition of zinc chelator, *o*-phenanthroline, to the enzyme-substrate reaction showed that the ability to hydrolyse the phosphorous substrate

was lost. The loss was due to the chelator, *o*-phenanthroline, binding to the zinc that is essential for the enzyme in the hydrolytic reaction. However, the hydrolase activity can be restored by dialysis of the apoenzyme with $ZnCl_2$ (Dumas *et al.*, 1989).

Table 1.2 Kinetic constants for the hydrolysis of organophosphorous compounds by phosphotriesterase from *P.diminuta* MG (Dumas *et al.*, 1989).

Structure	Common name	K_m	V	$(V/K_m)_{rel}$
	Paraoxon	0.09	100	100
	Dursban	0.11	0.08	0.08
	Parathion	0.24	30	11.25
	Coumaphos	0.39	29	6.70
	Diazinon	0.45	8.4	1.68
	Fensulfothion	0.46	3.2	0.63
	Cyanophos	2.1	7.5	0.32
	Methyl parathion	0.84	2.4	0.26

The mechanism for the hydrolysis of paraoxon by phosphotriesterase *of P. diminuta* MG has been proposed as in Figure 1.16. The activated water attacks the phosphorous centre of paraoxon whose phosphoryl oxygen is polarised by the active site zinc of phosphotriesterase. The rearrangement of chemical moieties around the phosphorous centre and hydroxyl of water results in the leaving of the *p*-nitrophenol molecule.

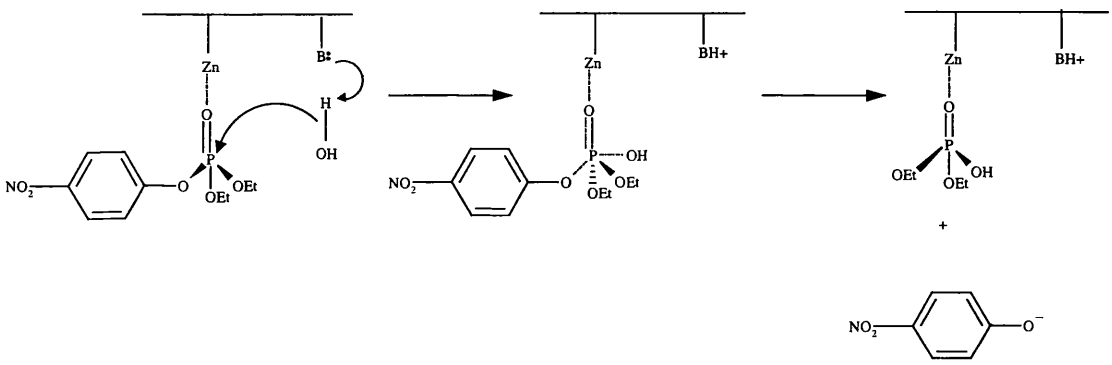


Figure 1.16 The model mechanism for the enzymatic hydrolysis of paraoxon by the phosphotriesterase from *P. diminuta* MG (proposed by Dumas *et al.*, 1989).

Determination of the deduced amino acid of *opd* of *P. diminuta* MG showed the N-terminus of this phosphotriesterase has the signal peptidase sequence of bacteria membrane proteins. This signal sequence is thought to be recognised by prokaryotic signal peptidase I and II in order to export the catalytic domains of enzyme to the membrane. The amino acid terminal sequence of phosphotriesterase of *P. diminuta* MG is shown in Figure 1.17.

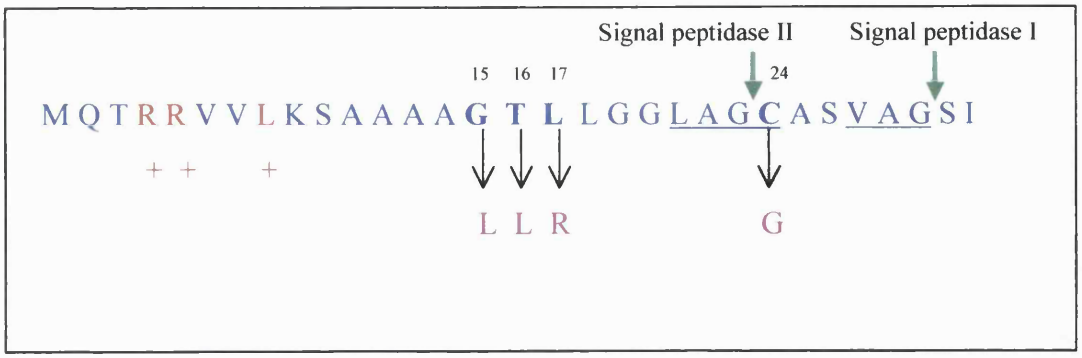


Figure 1.17 The N-terminal sequence of phosphotriesterase of *P. diminuta* MG. The potential signal peptidase cleavage sites are indicated as underlined and the green arrows show where the polypeptide is cleaved by signal peptidase I and II (+ shows the hydrophobic amino acid and the black arrows show where the mutation has been introduced in amino acid position 15, 16 and 24 (see text below)).

1.13 Heterologous expression of *opd* gene from *P. diminuta* MG in *Streptomyces lividans* and *E. coli*

The heterologous expression of the *opd* from *P. diminuta* MG in *S. lividans* and *E. coli* has been carried out (Steiert *et al.*, 1989). When the *opd* was cloned into *S. lividans*, the protein was found to be secreted out of the cell in a processed form, cleaved by the signal peptidase I (Black *et al.*, 1989). However, the expression of phosphotriesterase in *E. coli* gave very low levels of expression.

There have been mutational studies in the N-terminal sequence of phosphotriesterase from *P. diminuta* MG in *E. coli* in order to investigate the effect of the change of the N-terminal sequence on the stability and localisation of the protein (also see Figure 1.17.). The hydrophobic disruption by the change of a leucine 17 to arginine

prohibited the processing of the protein modification and the level of enzyme was found to be increased 2 folds (Dave *et al.*, 1993). This suggested the increasing stability of the phosphotriesterase. On the other hand, when the hydrophobic has been enhanced by altering the threonine (16) and glycine (15) to leucine, the enzyme has not been processed and the enzymatic activity has lost more than 98% (Dave *et al.*, 1993). In addition, the change of cysteine (24) on the potential signal peptidase II cleavage site to glycine had no effect on the efficiency of modification and translocation nor on stability of the modified protein (Dave *et al.*, 1993). The improvement of hydrophobic strength in the amino acid of the signal sequence of this membrane bound protein decreased its stability of the heterologous phosphotriesterase in *E. coli.*

Moreover, the mutation of histidine 55 and 57 to asparagine has been conducted to study the effect on the activity of the residues involved in the binding of zinc, substrate and enzyme. The mutational changes of H55 and H57 to N resulted in the enzymatic activity of the phosphotriesterase being almost completely lost.

1.14 Phosphotriesterase from *Flavobacterium* strain ATCC 27551

Mulbly and Karns (1989) have characterised three parathion hydrolases from *Flavobacterium* strain ATCC 27551, B-1 and Sc. These three phosphotriesterases are found in different part of the cells and they have different molecular weights. The hydrolase of *Flavobacterium* strain ATCC 27551 and SC are membrane bound proteins; however, the hydrolase from strain B-1 is cytoplasmic protein. The parathion hydrolase of *Flavobacterium* strain ATCC 27551 is composed of a single subunit with a molecular weight of 35.0 kDa. The parathion hydrolase of B-1 strain is also composed of a single subunit but it has a protein molecular weight of 43.0 kDa. The active hydrolase of the Sc strain is composed of four identical subunits with a molecular weight of 67. kDA.

The substrate specificities in these three parathion hydrolases are different. In that study, *O*-ethyl-*O*-4-nitrophenyl phenylphosphonothioate (EPN) and parathion are used as

a substrate. Both *Flavobacterium* strain ATCC 27551 and B-1 hydrolase can hydrolyse parathion and EPN. The SC-hydrolase is able to hydrolyse parathion in relatively high amounts compared to the ATCC 27551 and B-1 parathion hydrolase; however, it is unable to hydrolyse the EPN (Mulbry and Karns, 1989).

1.15 Heterologous expression of phosphotriesterase from *Flavobacterium* sp. ATCC 27551 in *E. coli* and *S. lividans*

Mulbry and Karns (1989) studied the heterologous expression of the parathion hydrolase of *Flavobacterium* sp. strain ATCC 27551 in *E. coli* cells. A *lacZ-opd* gene fusion was constructed by the deletion of the first 33 amino acid residues of *opd* hydrolase and replacing them by the first five amino acid residues of *lacZ*. This parathion hydrolase was constructed by deletion of its N-terminus as it would have been cleaved by signal peptidase in *Flavobacterium*. The *lacZ-opd* hydrolase fusion protein had the same molecular weight as the native hydrolase enzyme from *Flavobacterium*. Moreover, the hydrolase activity of the *lacZ-opd* fusion protein for parathion hydrolase showed higher activity than the cells harbouring the plasmid carrying *opd* gene (Mulbry and Karns, 1989). This study also demonstrated that nucleotide sequences of parathion hydrolase gene in both *P. diminuta* MG and *Flavobacterium* sp. ATCC 27551 are identical.

Heterologous expression of phosphotriesterase of *Flavobacterium* sp. in *S. lividans* has been studied (Rowland *et al.*, 1991). This heterologous parathion hydrolase in *S. lividans* is an extracellular protein secreted from the cells compared to the protein in its native host which is a membrane bound protein. However, the N-terminus of parathion hydrolase from *S. lividans* has an identical sequence to the native memebrane-bound parathion hydrolase of *Flavobacterium*. The lack of a signal peptidase site at the N-terminal of protein indicated that the heterologous parathion hydrolase has been modified in the same process as the native protein in *Flavobacterium* (Rowland *et al.*,

1991). The methyl parathion is a preferred substrate for methyl parathion hydrolase (Rowland *et al.*, 1991). Inactivation of the parathion hydrolase by *o*-phenanthroline can be reversed by adding Zn^{2+}. The study of Rowland, in its conclusion, showed that both herologous and native parathion hydrolase, in *S. lividans* and *Flavobacterium* respectively, have the same characteristics except that the heterologous parathion hydrolase in *S. lividans* is an extracellular protein.

1.16 Phosphotriesterases in other bacteria

1.16.1 Phosphotriesterase in *Plesiomonas* sp. M6

Methyl parathion hydrolase in *Plesiomonas* sp. M6, which has ability to hydrolase methyl parathion to *p*-nitrophenol, has recently been studied (Zhongli *et al.*, 2001). Parathion hydrolase (or its different names of phosphotriesterase, paraoxonase and parathion aryl esterase) is classified in the class of enzymes E C 3.1.3, hydrolytic enzymes hydrolysing an ester bond.

Cloning the *mpd* gene of the methyl parathion hydrolase has shown an ORF of 1062 nucleotide sequences. Protein encoded by the methyl parathion hydrolase has a molecular weight of approximately 35.0 kDa on SDS-PAGE. A BLAST search of the methyl parathion hydrolase protein sequence to the proteins in the GenBank database found this protein had the highest identity (31%) with hypothetical beta-lactamase (*S. coelicolor*) and 27-29% identity with other beta-lactamases (Zhongli *et al.*, 2001). This suggested that methyl parathion hydrolase is likely to share a common ancestor with the beta-lactamase superfamily proteins. However, the *mpd* has no homology to the parathion hydrolase genes of either *P.diminuta* MG or *Flavobacterium* sp. ATCC 27551. A study by Zhongli in 2001 also suggested that the parathion hydrolase is an endoenzyme (Zongli *et al.*, 2001). The activity of methyl parathion hydrolase for methyl parathion is 1.5×10^{-4} μmol/μg-protein/min of the crude cell-free extract of *E. coli* BL21 harbouring the *mpd* gene.

1.16.2 Organophosphate anhydrolase in *Alteromonas* sp. strain JD 6.5

Organophosphophate anhydrase is the enzyme catalysing the hydrolysis of a variety of organophosphorous compounds, including paraoxon, sarin (*O*-cyclohexyl methylphosphorofluoridate) , soman (*O*-isopropyl methyl phosphonofluoridate) and GF (*O*-pinacolyl methylphosphonofluorididate) (DeFrank *et al.*, 1991 and Hill *et al.*, 2000). This enzyme is encoded by the *opaA* gene on a 1.74 kb *Pst*I-*Hind*III DNA fragment from the chromosome of *Alteromonas* sp. JD 6.5 (Cheng et al., 1996). The *opaA* is composed of 1716 nucleotides and encodes for a 571 amino acid which has a molecular mass of about 56.0 kDa. The anhydrase shares a high similarity with prolidase, PepQ (47% identities) from *E. coli*, and a human prolidase (28% identities) (Cheng *et al.*, 1996).

The heterologous anhydrolase in *E. coli* has a specificity for DFP (diisopropyl fluorophosphate) of about 0.7 mmol/min/mg of protein. The OPA anhydrase is also capable of the hydrolysis of P-F and P-O bonds of various organophosphate compounds, GB (O-isopropyl methylphosphonofluoridate), GD (O-pinacolyl methylphosphonofluoridate) and NP-GD (a chromogenic soman analog) (Cheng *et al.*, 1996).

1.16.3 A novel phosphotriesterase in *Pseudomonas monteilli* strain C11

An attempt to isolate a *P. monteilli* C11 with a novel phosphotriesterase has been made by Horne (2002). The phosphotriesterase from the soluble fraction of the *P. monteilli* C11 can hydrolyse coumaphos and coroxon. Horne revealed that this novel phosphotriesterase has a specific activity for coumaphos and coroxon of 5.14 and 7.86 nmol/h/mg of protein respectively. By a southern hybridisation experiment, the C11 phosphotriesterase is not related to phosphotriesterase from *Flavobacterium* sp. ATCC 27551 (Horne *et al.*, 2002).

1.17 The structure of phosphotriesterase

The first three-dimensional structure of zinc bound phosphotriesterase from *P. diminuta* was solved by Vanhooke and coworkers in 1996. In later studies, a 3D structure of the phosphotriesterase was carried out by Benning and coworkers in 2000 and 2001. The structure of phophotriesterase apoenzyme is a α/β barrel, which consists of eight β-pleated sheets in parallel orientation (see Figure 1.18 and 1.19).

```
1   GDRINTVRGP ITISEAGFTL THEHICGSSA GFLRAWPEFF GSRKALAEKA
     EEEECCEE EEHHHHCCCEECCCCCEECCCCCHHHHCCCCC CCHHHHHHHH

51  VRGLRRARAA GVRTIVDVST FDIGRDVSLL AEVSRAADVH IVAATGLWFD
    HHHHHHHHHC CCCEEEECCC CCCCCCHHHH HHHHHHHCCC EEEEEEECCC

101 PPLSMRLRSV EELTQFFLRE IQYGIEDTGI RAGIIKVATT GKATPFQELV
    HHHHCCCHHH HHHHHHHHHH HCCCCCCCCC CEEEEECCCCCCCHHHHHHCC

151 LKAAARASLA TGVPVTTHTA ASQRDGEQQA AIFESEGLSP SRVCIGHSDD
    HHHHHHHHHH HCCEEEEECC CCCCHHHHHH HHHHHCCCCC CCEEECCCCC

201 TDDLSYLTAL AARGYLIGLD HIPHSAIGLE DNASASALLG IRSWQTRALL
    HHHHHHHHHH CCEEEECCCC CCCCCCCCCH HHHHHHCCCC HHHHHHHCCC

251 IKALIDQGYM KQILVSNDWL FGFSSYVTNI MDVMDRVNPD GMAFIPLRVI
    HHHHHHCCCC CCEEECCCCC CCEECCCCCH HHHHHHHCCCCCCCHHHHTHH

301 PFLREKGVPQ ETLAGITVTN PARFLSPTLR AS
    HHHHHCCCCH HHHHHHHTHH HHHHHCCCCC CC
```

Figure 1.18 Protein secondary structure of phosphotriesterase from *P. diminuta* (H is helix; E is strand; and C is coil or loop).

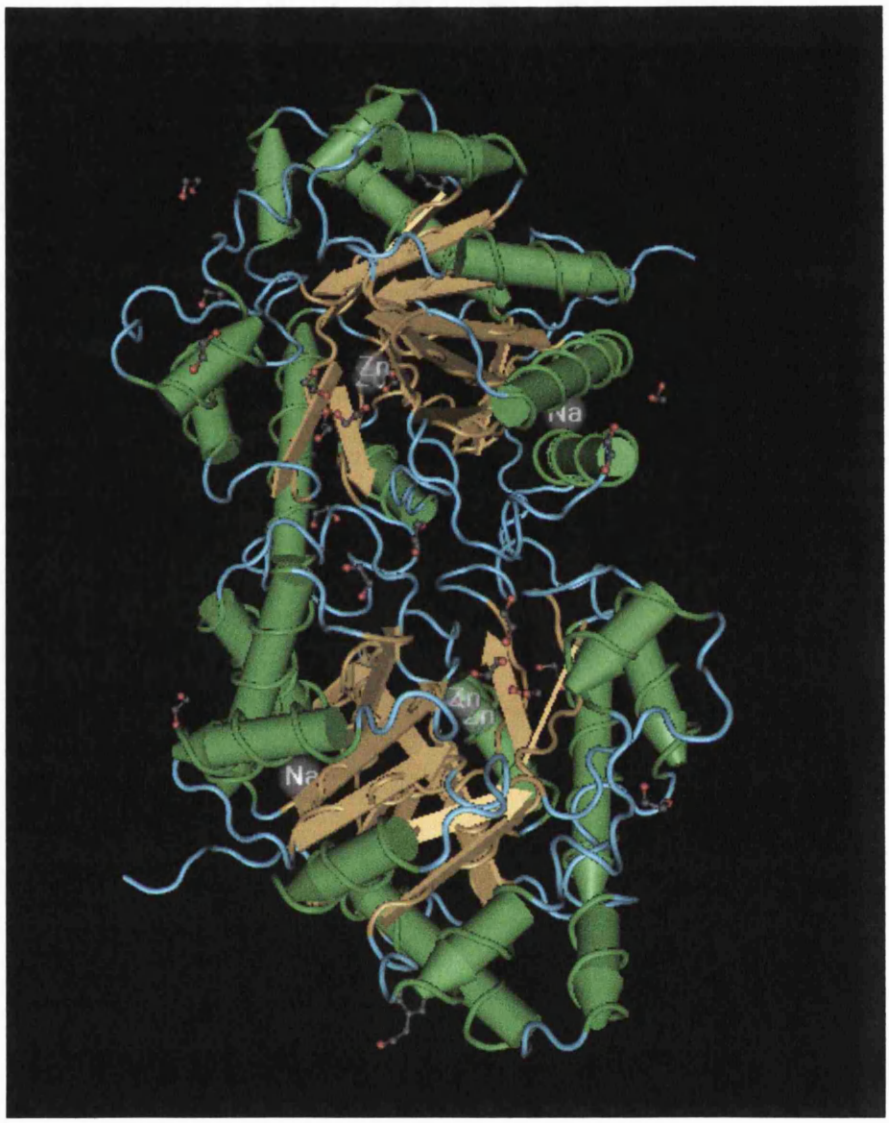


Figure 1.19 Three dimensional structure of phosphotriesterase from *P. diminuta*.

The native enzyme of phosphotriesterase from *P. diminuta* is dimeric, composed of two identical subunits and containing two zinc ions. Yet, the metal ions of Zn^{2+} in the phosphotriesterase can be substitued with Co^{2+}, Ni^{2+}, Cd^{2+}, or Mn^{2+}, which render the level of enzymatic activity of these ion-enzymes as the same as Zn^{2+} bound phophotriesterase (Omburo *et al.*, 1992). In the active centre of the metallo-phosphotriesterase, one Zn^{2+} is co-ordinated to His55, His57 and Asp301, and another Zn^{2+} is coordinated to His201 and His230, and these two zinc ions are bridged by a carboxylated lysine-169 of the phosphotriesterase (Vanhooke, 1996). Nonetheless, these two Zn^{2+} are bridged by a hydroxide ion (OH^-) (see Figure 1.20). This Zinc-bridging hydroxide ion is thought to be a nucleophilic molecule attacking the phosphorous centre in the hydrolytic mechanism of organophosphorous compounds by the phosphotriesterase. The two zinc ions are separated by 3.3 Å, and are found away from the amino acid residues in the active site of phosphotriesterase between 1.9-2.3 Å (see Figure 1.20).

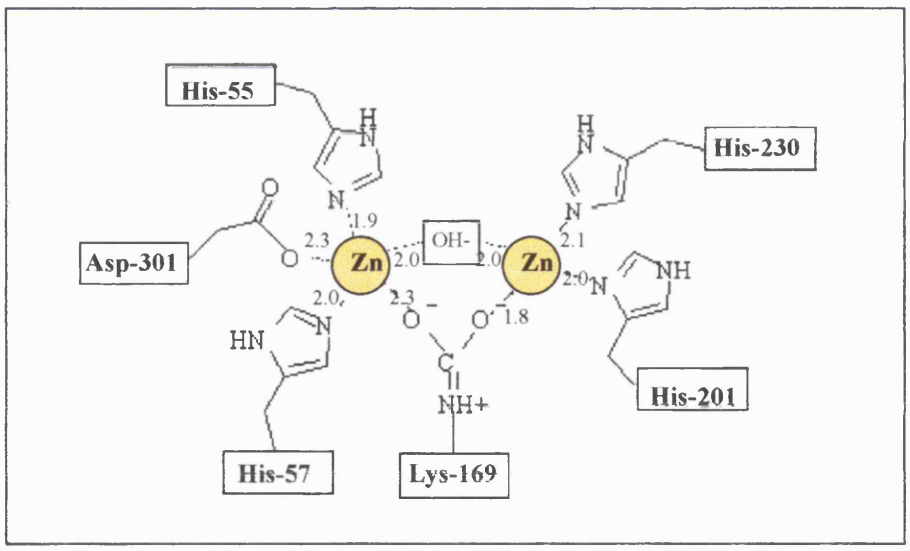


Figure 1.20 Active site pocket of phosphotriesterase from *P. diminuta*. Two Zn^{2+} are bound to the functional amino acids in the active site of phosphotriesterase, and they are bridged by hydroxide.

1.18 The active site of phosphotriesterase

The identity of catalytic residues in binding and reacting with the substrate in the phosphotriesterase from *P. diminuta* has been previously studied (Dumas and Raushel, 1990 and Kuo and Raushel, 1994).

Investigation of amino acids, cysteine, aspartate, glutamate, lysine and arginine of phosphotriesterase has been carried out by inactivation modification using chemical agents (Dumas and Raushel, 1990). The amino acids of, for example, cysteine can be modified by methyl methanethiosulfonate or 2-nitrobenzoate; nucleophilic residues can be modified by iodoacetamide; arginine can be modified by butanedione and histidine can be modified by diethyl pyrocarbonate. The results, using chemical modification, showed the cystene, aspartate, glutamate and arginine appeared not to be involved in the catalytic activity of phosphotriesterase. However, the inactivation of histidine residues resulted in complete loss of the hydrolytic activity of phosphotriesterase to paraoxon. The study by Dumas and Raushel (1990) suggested that the histidine residues were the functional amino acids in the active site of phosphotriesterase.

Site-directed mutagenesis has been employed to alter seven histidine residues (H55, H57, H123, H201, H230, H254 and H 257) in the phophotriesterase to asparagine (Kuo and Raushel, 1994). In addition, by determination of the kinetic activity and the metal content of each histidine-mutant enzymes, the function and significance of these histidine residues have been ruled out. The active site of the zinc ion-binding centre is proposed to be consisted of 4 histidines (H55, H57, H254 and H257). Moreover, it has been suggested that His230 and His201 are the binding ligands between the Zn^{2+} ions and the catalytic base respectively (Kuo and Raushel, 1994). However, the identity of the functional amino acids in the active site of phophotriesterase is not fully understood.

In 1996, the x-ray crytallographic structure of phophotriesterase from *P. diminuta* was examined. The three dimensional structure of a substrate analogue of 4-methyl benzylphophonate bound phosphotriesterase in *P. diminuta* is useful information to understand the catalytic mechanism of phosphotriesterase, based on the bridging hydroxide reacts to the phosphorous centre of a substrate. More importantly, the studies of the different metal-substituted forms and metal-substrate interactions of phosphotriesterase (Hong and Raushel, 1996; Benning *et al.*, 2001; and Raushel, 2002), gives a working model for the catalytic mechanism of paraoxon by the enzyme (see Figure 1.21).

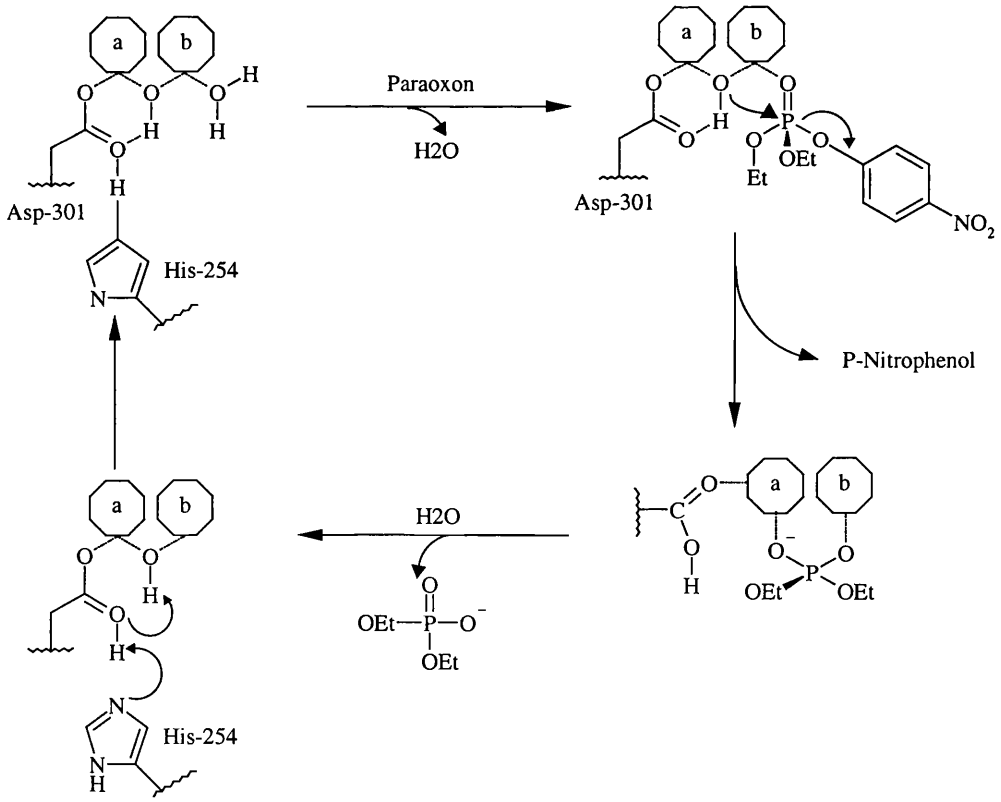


Figure 1.21 The catalytic model for the hydrolytic mechanism of paraoxon by phosphotriesterase from *P. diminuta* (a and b are zinc ions).

1.19 Phosphotriesterase in decontamination and detoxification of organophosphorous compounds

Organophosphorous compounds have been used extensively in agricultural insecticides and pesticides in the U.S. As a result of this wide use, there is increasing concern about spreading and contamination of the organophosphorous compounds in the environment. The organophosphorous compounds are well known in their ability to inhibit acetylcholinesterase, the vital enzyme in neurotransmission, and these compounds are also used to prepare chemical warfare agents. Poisoning from the organophosphorous compounds can result in serious symptoms and death.

Organophosphorous hydrolases in various microorganisms have been studied recently. With broad-spectrum of the phosphotriesterase, which hydrolyses a variety of organophosphorous compounds and phosphorous-ester bonds (P-CN, P-F, P-O and P-S) (Lai *et al.*, 1995), it is consided useful in using phosphotriesterase in the decontamination and detoxification of organophosphorous compounds in the environment. Moreover, the possibility of detoxification of nerve gases used in chemical warfare also can be put into practise as a study of the hydrolysis of Russian-VX, R-VX; *O*-isobutyl *S*-(2-diethylamino) and VX; *O*-ethyl *S*-(2-diisopropylaminoethyl) by organophosphorous hydrolase from *P. diminuta* MG (Rastogi *et al.*, 1997).

The role of the phosphotriesterase hydrolase in decontamination and detoxification of organophosphorous compounds would be important in the future. Improvement of phosphotriesterase by genetic manipulation and cell immobilisation technology (Mulchandani *et al.*, 1998; and Gill and Ballesteros, 2000) will possibly lead to convenient and effective use of the enzyme.

1.20 Aims of the study

A partial N-terminal sequence of 3,6-diketocamphane 1,6-monooxygenase has been obtained by our collaboration group at Exeter University. The designed PCR primers based on this amino acid sequence were used to amplify a 0.7 kb DNA fragment from a total DNA of *P. putida* NCIMB 10007. This DNA fragment was cloned into a pBluescript vector and designated pQR277. We have also cloned a 0.3-kb DNA fragment from the left end of the known DNA sequence of the CAM P450 gene operon. The plasmid containing this DNA fragment is called pQR203. Thus, we will use the pQR277 probe to screen the DNA region of 3,6-diketocamphane monooxygenase gene; and the pQR203 probe to clone the DNA fragment on the left-hand side of *cam* operon from the *P. putida* NCIMB 10007 genomic DNA. In addition, one or two genes from these clones will be characterised. We also expect there will be more monooxygenase genes, probably a 2,5-diketocamphane 1,2-monooxygenase, which might be located on either side of the known DNA sequence neighbouring this *cam* operon. Exploring the CAM plasmid will extend our understanding of the genetic composition of the CAM plasmid and knowledge of the camphor pathway in *P. putida* NCIMB 10007.

Chapter 2

Materials and Methods

2.1 Bacteria and plasmids

Bacteria and plasmids used in this study are listed in Table 1 and 2. Generally, all bacterial strains used in these experiments were maintained in 20% glycerol and kept at –70°C.

Table 2.1 Bacteria, plasmids and their characteristics, that were used in this study.

Bacteria or plasmids	Characteristics	References
P. putida NCIMB 10007	Gram-negative bacteria can grow on camphor as a sole carbon source	NCIMB
E. coli DH5α	F⁻, *endA1*, *hsdR17* (r_k^-, m_k^+), *supE44*, *thi-1*, λ^-, *recA1*, *gyrA96*, *relA1*, Δ(*argF⁻ lacZYA*) *U169*, ϕ80*dlacZΔM15*	Life Technologies
E. coli BL21(DE3)	F⁻ *ompT[lon] hsdS*$_B$ (r_B^- m_B^-; an *E. coli* B strain with DE3 , a λ propharge carrying the T7 RNA polymerase)	Novagen
E. coli BL21(DE3) pLysS	F⁻ *ompT[lon] hsdS*$_B$ (r_B^- m_B^-; an *E. coli* B strain with DE3 , a λ prophage carrying the T7 RNA polymerase), pLysS (Cm^R)	Novagen
E. coli BL21(DE3) CodonPlus-RIL	*E. coli* B F⁻ *ompT hsdS* (r_B^- m_B^-) *dcm* $^+$ Tet^r *gal* λ (DE3) *endA Hte argU ileY leuW* Cam^r	Stratagene
E. coli BL21(DE3) CodonPlus-RP	*E. coli* B F⁻ *ompT hsdS* (r_B^- m_B^-) *dcm* $^+$ Tet^r *gal* λ (DE3) *endA* Hte *argU proL* Cam^r	Stratagene
pUC18/19	Amp^r, ColE1replicon, *lacZ*, cotaining ploylinker cloning sites, plasmid of 2,686 bp	MIB
pET21a	Amp^r, f1 origin, pBR322 origin, *lacI*, T7 promoter, His-tag coding sequence, multiple coloning sites, plasmid of 5443 bp	Novagen
pCR 2.1-TOPO	Amp^r, Km^r, T7 promoter, pUC origin, f1 origin, *lacZ*, containing multiple cloning site, cloning vector (3.9 kb) for the direct insertion of *Taq* polymerase-amplified PCR products,	Invitrogen
pBluescript SK	Amp^r, *lacZ*, f1 origin, ColE1 origin, containg multiple cloning site, phargemid of 2958 bp	Stratagene

Table 2.2 Plasmids constructed in this study.

Plasmids	Characteristics
pQR203	pBluescript SK with an insert of a 222 bp DNA fragment from the left-hand side of the *cam* P450 operon
pQR277	pCR2.1-TOPO with an insert of the 0.7 kb DNA fragment derived from the N-terminus of 3,6-diketocamphane 1,6-monooxygenase
pQR416	pUC19 with a 7.1 kb *Bam*HI insert (*orf1234-camR* and *camD*)
pQR417	pUC19 with a 4.2 kb *Kpn*I insert (*orf5678*), a chormosomal DNA of *P. putida* NCIMB 10007
pQR418	pCR2.1-TOPO with a *orf2* gene insert
pQR419	pCR2.1-TOPO with a *orf2* gene insert with the deletion of its stop codon
pQR420	pCR2.1-TOPO with a *orf4* gene insert
pQR421	pCR2.1-TOPO with a *orf4* gene insert with the deletion of its stop sequence
pQR422	pET21a with *orf2* gene insert
pQR423	pET21a with *orf2* gene that cloned in frame with histidine sequence
pQR424	pET21a with *orf4* gene insert
pQR425	pET21a with *orf4* gene that cloned in frame with histidine sequence
pQR426	pUC19 with *camR-camD*
pQR427	pCR2.1-TOPO with *camC-camA-camB*
pQR428	pUC19 with *camC-camA-camB*
pQR429	pUC19 with *camR-camD-camC-camA-camB*
pQR430	pUC19 with *orf1234*
pQR431	pUC19 with an 192 bp *Eco*RI-*Eco*RV fragment (a DNA fragment at the 5′end of *orf1* gene)

2.2 Chemicals, enzymes and media

Parathion (*O*,*O*-diethyl-*O*-*p*-nitrophenyl phosphorothioate), methyl parathion (*O*,*O*-dimethyl-*O*-*p*-nitrophenyl phosphorothioate), paraoxon (*O*,*O*-diethyl-*O*-*p*-nitrophenylphosphate), *p*-nitrophenol, glucono-δ-lactone, γ-butyrolactone (4-hydroxybutyric acid γ-lactone) , pantolactone (2,4-dihydroxy3,3-dimethylbutyric acid-γ-lactone), γ-caprolactone and δ-valerolactone were purchased from SIGMA. Other routine laboratory chemicals were from either Sigma or BDH.

All restriction endonucleases and calf intestinal alkaline phosphatase (CIP) were purchased from NEB (New England Biolabs). T_4 DNA ligase and Taq DNA polymerase was purchased from Fermentas. The digoxigenin (DIG) non-radioactive nucleic acid labelling and detection system were purchased from Boehringer Mannheim GmbH, Biochemica.

Nutrient broth (NB) and nutrient agar (NA) were purchased from OXIOD (OXIOD Ltd., England). Other media were prepared according to standard methods (see also Appendix).

2.3 Bacteria growth

P. putida NCIMB 10007 cells were grown on either agar plates or liquid media. For agar plate growth, the NCIMB 10007 cells were streaked out on a M9 agar plate (M9 salt, 15% (w/v) agar) supplemented with 2% glucose, and camphor (about 1g) was put on the top lid of the plate. The plate was placed in a closed container and incubated at 27°C or 37°C for 24-48 hours. To grow *P. putida* NCIMB 10007 cells in liquid media, the cells were inoculated into 200 ml of M9 minimal media (Atlas, 1997) supplemented with 2% glucose and camphor (4 gram/litre) with 200 rpm shaking at 27°C for 16-18 hours.

For a plasmid miniprep, a bacteria culture was carried out by inoculation of a single colony into 5 ml nutrient broth and incubation at 37°C with shaking at 100 rpm for 16-18 hours. An about 1.5 ml of the culture was centrifuged at 7,000 rpm for 5 minute, which gave pellet cells approximately 1-2 mg (cell wet weight).

2.4 Gel electrophoresis

For routine works, an agarose gel was prepared as 1% agarose (in separation of a DNA fragment between 0.5-10 kb) in 0.5×TBE buffer (0.045 M Tris-borate, 0.001 M EDTA, pH 8.0), 50 μM ethidium bromide. The DNA sample was added with 10×loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in water), an appropriate amount (approximately 10% v/v), before loading on the gel. The gel was run in 0.5×TBE buffer at approximately 1-20 V/cm using GIB-CO gel electrophoresis apparatus. The DNA markers of λ- *Pst*I (λ DNA cut with *Pst*I), 1 kb ladder (New England Biolabs) and 100 bp ladder (New England BioLabs) were used.

2.5 DNA digestion with restriction endonuclease

In general, the following recipe was used to prepare the DNA digestion with restriction endonuclease (20μl reaction);

1) X μl DNA (100-500 ng DNA in Tris-HCl, pH 8.0)
2) 2 μl 10× restriction buffer (commercially available with the restriction enzyme)
3) 18-X μl sterile water
4) 1 μl restriction endonuclease (1-5 unit/μl)

The reaction mixture was incubated at 37°C (or temperature recommended by the manufacturer) for 1-2 hours. In double digestion, the specific buffer was used as recommended by manufacture.

2.6 DNA ligation

To ligate a DNA fragment into a vector, a typical reaction for ligation was carried out as follows (20 μl reaction);

1) 4 μl of DNA vector (100-200 ng/μl)
2) 8 μl of insert DNA (100-200 ng/μl)
3) 2 μl of 10x T_4 DNA ligase buffer
4) 5 μl of sterile water
5) 1 μl of T_4 DNA ligase (20 units/μl)

In order to give a good result for ligation, the optimal amount of plasmid DNA and ratio of insert and vector DNA (foreign DNA 2:1 plasmid vector according to Sambrook *et al.*, 1989) was determined roughly under a UV transluminator. The ligation reaction was incubated at 20-25°C for 1 hour or at 15°C overnight. For transformation, the DNA from the ligation can be used straight away.

2.7 Removing 5′ and 3′ end of DNA by alkaline phosphatase

To prevent self-ligation of DNA plasmid that had been cut with one restriction endonuclease, alkaline phosphatase from calf intestinal (CIP) (New England Biolabs) was used. The following recipe (20 μl reaction) was set up for the reaction to remove 5′ and 3′ phosphoryl groups from nucleic acids.

1) X μl of DNA (100-200 ng/μl)
2) 2 μl of 10x CIP buffer
3) 17-X μl of sterilised water
4) 1 μl of CIP (10-50 units/μl)

The mixture was incubated at 37°C for 1 hour, and later the DNA was extracted by either phenol/chloroform extraction followed by ethanol precipitation, or by using the QAIGEN gel extraction kit.

2.8 Polymerase chain reaction (PCR)

The DNA primers were supplied by Phamacia Biotech. The following basic recipe was used in PCR experiments; however, some parameters can be altered for optimal reaction conditions. To a 0.5 ml sterile ependorf tube, the following components were added:

1) 10×PCR buffer	10 μl
2) 10 mM dNTPs mixture	2 μl
3) primer mixture (10 μM each)	5 μl
4) DNA template	5 μl
5) *Taq* DNA polymerase (10 U/μl)	1 μl
6) sterile water to	100 μl

The mixture was mixed by pipetting up and down, overlaid with 50 μl of sterilised mineral oil, and briefly centrifuged. The mixture was then subjected to further PCR amplification using Hybaid Omnigene PCR machine, as in the following conditions.

1) a cycle of 1 minute at 94°C;
2) 25 cycles of 30 seconds at 94°C
   30 seconds at 60°C
   3 minutes at 72°C;
3) a cycle of 10 minutes at 72°C.

After the PCR amplification was completed, the PCR products beneath the mineral oil, were removed to a new sterile tube and used immediately (to clone into pCR2.1- TOPO vector) or kept at –20°C until next use.

2.9 Isolation of the CAM plasmid and genomic DNA from *P. putida* NCIMB 10007

Isolation and purification of the total DNA of *P. putida* NCIMB 10007 was performed on cesium chloride (CsCl)-ethidium bromide (EtBr) gradients (Sambrook and *et al*, 1989) using quick seal centrifuge tubes (Beckman; Beckman Instrument, Inc.) Cells of *P. putida* NCIMB 10007 (from a 200 ml culture) were pelleted by centrifugation and the pellet was resuspended with 3 ml of resuspension buffer (P_1 buffer) (50 mM Tris-Cl pH 8.0, 50 mM glucose, 10 mM EDTA, 100 μg/mL). Then the cells were lysed by using 3 ml of P_2 buffer, lysis buffer (0.2 M NaOH, 1% SDS), and the tube was mixed gently by inverting 7-8 times. After that, 3.5 ml of N_3 buffer, neutralisation buffer (3 M potassium acetate, pH 5.5) was added, and the tube was inverted 7-8 time. Then CsCl was added to the mixture as CsCl 1.1g for 1 ml of lysate. The mixture was mixed gently by inverting until the CsCl was dissolved, and then 300 μl of 10 mg/ml ethidium bromide was added. The DNA/CsCl/EtBr solution was centrifuged at 38,000 g for 72 hours or 42,000 g for 24 hours at 18°C using Beckman L7 Ultracentrifuge and Beckman rotor type Ti70.1. Both upper and lower bands located in the gradient after centrifugation were collected. To remove ethidium bromide, an equal volume of isopropanal saturated with water was added to the DNA solution. The upper layer of isopropanol and ethidium bromide was removed and the extraction was repeated 3-4 times. After ethidium bromide was clearly removed (the bottom layer became crystal clear), the DNA solution was dialysed against TE buffer (10 mM Tris-Cl, pH 7.6, 1 mM EDTA), the DNA was precipitated by ethanol precipitation (see below), and finally precipitated DNA was dissolved in 0.5 ml of TE buffer.

2.10 Phenol/chloroform extraction and ethanol precipitation of DNA

In the extraction of DNA, phenol/chloroform/isoamyl alcohol (25:24:1) (from SIGMA) was added to the DNA solution (1: 1 volume), and the mixture was vortexed for approximately 10 seconds. The DNA was then centrifuged at 14,000 rpm at room temperature for 1 minute. The top aqueous phase, which contained the DNA, was removed to a new clean centrifuge tube. NaCl solution was added to the DNA solution (5 M NaCl 1 : 10 DNA solution (v/v)) and mixed gently by inverting. Then, two volume of ice-cold ethanol was added to the DNA solution, mixed gently by inverting and left at –20°C for 2 hours or –70°C for 30 minutes. After that, the tube was spun down at 14,000 rpm for 15 minutes. The supernatant was gently removed, 2 volumes of 70% ethanol were added, and the tube was centrifuged at 14,000 rpm for 10 minutes. The supernatant was discarded, and the pellet was dried under vacuum for about 30 minutes. The dry pellet was dissolved in TE buffer (10 mM Tris-Cl, pH 7.6, 1 mM EDTA) or Tris-Cl pH 8.0 in an appropriate volume.

2.11 Plasmid miniprep

Throughout this study, QIAprep spin miniprep kit (QIAGEN) was used for plasmid minipreps. Briefly, to obtain bacteria cells, a colony was inoculated into 5 ml nutrient broth and incubated at 37°C with 100 rpm shaking for 16-18 hours. Then the culture was poured into a 1.5 ml eppendorf tube, and the cells were spun down at 7,000 rpm for 1 minute. In the next step, the supernatant was discarded, and the pellete cells were resuspended with 250 μl of buffer P_1, resuspension buffer (50nM Tris-Cl pH.8.0, 50 mM glucose, 10 mM EDTA, 100 μg/ml RNase), and vortexed until the cells were completely resuspended. Then 250 μl of buffer P_2, lysis buffer, (0.2M NaOH, 1% SDS) was added, and the tube was inverted 4-6 times to mix. The N_3, neutralisation buffer (3.0 M potassium acetate, pH 5.5) 350 μl was added, and the tube was inverted immediately 4-6 times. Then the tube was centrifuged at 14,000 rpm for 10 minutes in a

microcentrifuge. After the centrifugation, supernatant was applied to the QIAprep column and spun down at 14,000 rpm for 1 minute. The flow-through was discarded from the QIAprep spin column, and 750 μl of buffer PE, wash buffer (1.0M NaCl, 50mM MOPS, pH 7.0 and 15% ethanol), was applied to the QIAprep column and centrifuged for an additional 1 minute. The QIAprep column was removed from the spin column and placed in a clean 1.5 ml microcentrifuge tube. To elute plsmids from the QIAprep column, 50 μl of buffer EB, elution buffer, (10 mM Tris-HCl, pH 8.5) was added. The column was left for 1 minute and then spun down at 14,000 rpm for 1 minute. The eluted plasmid in flow-through EB buffer was collected and kept at –20°C until next use.

2.12 QIAGEN gel extraction

QIAGEN gel extraction kit (QIAGEN Ltd.) was used for the extraction of DNA fragments from agarose gels. First, the DNA fragments in agarose gel were excised with a scalpel, weighed and put into a clean eppendorf tube. Three volumes of buffer QG (binding and solubilization buffer) were added in 1 volume of the gel. The sample was incubated at 50°C for 10 minutes or until the gel was dissolved. Then one volume of isopropanol to the gel was added to the sample. The whole solution was transferred to QIAquick spin column, which was already in the collection tube. Then the QIAquick column was centrifuged at 14,000 rpm for 1 minute, and the flow through from the sample was discarded. 0.75 ml of buffer PE, wash buffer (1.0M NaCl, 50mM MOPS, pH 7.0 and 15% ethanol), was added to the QIAquick column. The column was spun down at 14,000 rpm for 1 minute. After that, the QIAquick column was placed in a clean 1.5 ml eppendrof tube. The 50 μl of buffer EB, elution buffer, (10 mM Tris-Cl, pH 8.5) was pipetted into the column, and the column was left for 1 minute. Finally, the column was centrifuged at 14,000 rpm for 1 minute. The flow through containing DNA fragments was collected and kept at –20°C until next use.

2.13 Preparation of competent cells

E. coli DH5a, BL21(DE3) pLysS, BL21(DE) CodonPlus-RIL and BL21(DE3) CodonPlus-RP were grown overnight in a 5 ml of nutrient broth (except for *E.coli* DH5α, nutrient broth was supplemented with 34 μg/ml chloramphenicol) at 37°C with 100 rpm shaking . Then this culture was inoculated into 200 ml of nutrient broth containing 20 mM $MgCl_2$. The cells were then grown at 37°C with 100 rpm shaking until the mid-log phrase (OD at 600 nm about 0.6-0.8) or about 2 hours. The cells were chilled on ice and transferred into sorvall tube GSA, and centrifuged for 15 minutes at 6,000 rpm (Sorvall RC2-B, GSA rotor) and 4°C. After the spinning, the supernatant was discarded, and the pellet in the sorvall tube was kept on ice. The pellet was then resuspended with 5 ml of ice cold sterile 75 mM $CaCl_2$, 15 % (v/v) glycerol. An aliquot of 100 μl competent cells was pipetted into pre-chilled sterile vials or ependorf tubes. Cells were frozen at –70°C or used immediately if a ligation was ready.

2.14 Transformation

First, the frozen competent cells were thawed on ice. The plasmid or ligation mixture was added to the thawed competent cells, mixed by flicking the tube. The tube was left on ice for about 30- 45 minutes allowing the transformation reaction to complete. Then, the tube was heat shocked at 37°C for 5 minutes, and put in ice immediately after the heat shock. Nutrient broth or SOC medium (400 μl) was added to the tube and the tube was incubated at 37°C with 100 rpm shaking for 45 minutes. Different volumes of media containing transformant cells, 50 μl and 100 μl, were plated out on ampicillin, IPTG and X-gal plates, and grown overnight at 37°C. However, a selective agar plate used in screening other transformants with resistance to tetracycline, kanamycin and chloramphenicol were plated out on a nutrient agar plate supplemented with tetracycline (10 μl/ml), kanamycin (10 μl/ml) and chloramphenicol (25μl/ml) respectively. The remaining transformants, in the nutrient broth, were kept at 4°C and plated at the later

date if necessary. After overnight incubation, the white colonies were selected and streaked out on another AIX plate and kept as a master plate.

2.15 Electroporation

Alternatively, for a highly efficient transformation, electroporation was occasionally used instead of conventional transformation using competent cells made from $CaCl_2$. Host cells were grown in a 200-ml nutrient broth to the log phrase (or OD_{600} of 0.6-0.8). The cells were put on ice for 15 minutes before they were spun down at 5,000 rpm for 10 minutes at 4°C. Then the cells were resuspended in 200 ml ice cold sterile water and centrifuged at 5,000 rpm for 10 minutes at 4°C. The supernatant was removed, and the ice-cold water was added to resuspend the cells to approximately 2×10^{11} cells/ml. The fresh cells (50 μl) were aliquoted into microcentrifuged tubes and used immediately for electroporation by adding 1 μl of plasmid DNA (0.5-1.0 μg/μl). The mixture was transferred to the pre-chilled cuvette (Invitrogen) for electroporation and placed on the electroporation port. The following parameters were set for the electroporation; 1,800 V, 20 μF and 200 Ω. The pulse was then applied to the cuvett, and SOC media was immediately added to the electroporated mixture after the pulse. The cells in the SOC media were transferred into a sterile tube, and incubated at 37°C modulated with 200 rpm shaking for 45 minutes. Finally, the electroporated cells were plated out on appropriated agar plate.

2.16 Preparation of DIG DNA labelled probe

There were two probes used in this study: pQR277, whose fragment is derived from N-terminal of 3,6-diketocamphane 1,6-monooxygenase determined by our collaborator in Exeter; and pQR203 contains a 222-bp DNA fragment from *EcoR*I-*Sma*I digest at the end of *camR* (on the left-hand side of *cam* operon). The plasmids of

pQR277 (pBluescript SK carrying 0.7 kb DNA fragment derived from 3,6-diketocamphane 1,6-monooxygenase sequence) and pQR203 (pBluescript SK carrying 222bp DNA fragment on the left-hand side of *cam* operon) were originally constructed by Dr. John Ward.

To prepare the DIG DNA labelled probe, the pQR277 and pQR203 plasmids were digested with *EcoR*I and *Hind*III (see also section 5) and run on an agarose gel. The fragment of 0.7 kb from pQR277 and 0.2 kb from pQR203 were excised from the gel and purified using the QIAGEN Gel Extraction Kit as previously described. Further DNA labelling by the DIG DNA labelling system was performed according to manufacturer recommendation (Boehringer Mannheim GmbH, Biochemica).

Briefly, the template DNA was diluted to concentration 5-25 ng/ml in a total volume of 15 μl and denatured in boiling water for 10 minutes. The DNA was quickly put in a chilled ice/NaCl. The following reagents were added: 2 μl of hexanucleotide mix, 2 μl of dNTPs mixture and 1 μl of Klenow enzyme. The mixture was incubated at 37°C over night. After that 2 μl of 0.2 M EDTA (pH 8.0) was added to stop the reaction. The labelled DNA was precipitated by the addition of 2.5 μl of 4 M LiCl and 75 μl of pre-chilled ethanol. The mixture was left at –70°C for 30 minutes and then spun down in a microcentrifuge at 14,000 rpm for 15 minutes. The pellet was washed with 50 μl of 70% ethanol and spun again for 10 minutes. The supernatant was removed and the DNA pellet was dried briefly under vacuum. Finally, the DNA pellet was dissolved in 50 μl TE buffer or EB buffer (Tris-Cl, pH 8.0) and kept at –20°C.

2.17 Southern blotting

The recombinant plasmid DNA digested with a restriction endonuclease (*Bam*HI, *Kpn*I, *Sac*I, *Pst*I or *Eco*RI) was loaded in an agarose gel containing 50 mM ethidium

bromide in 0.5×TBE buffer and run slowly at 20V for 16-18 hours. After that, the gel was treated as follows.

1) The gel was equilibrated in 100 ml of depurination solution (0.2 M HCl) and incubated with gentle shaking 15-20 rpm for 10 minutes. The depurination solution was then discarded, and 200 ml of distilled water was added, shaken briefly and poured off.
2) The gel was equilibrated 100 ml of denaturation solution (0.5 M NaOH, 1.5 M NaCl), gently shaking for 15 minutes (repeated twice). Then the gel was washed twice with 200 ml of distilled water.
3) The gel was then immersed in 100 ml of neutralisation solution (1 M Tris-HCl buffer, pH 8.0, 1.5 M NaCl) and shaken gently for 20 minutes.

The gel was treated was inverted and placed on Whatman 3MM paper, which was on the support, and soaked with 2×SSC (0.3 M NaCl, 30 mM trisodium citrate, pH 7.0) (see Figure 2.1). Then, positively charged nylon membrane (Boehringer Mannheim GmbH, Biochemica) was placed on top of the gel. Two pieces of Whatman 3 MM paper were soaked in 2×SSC and placed on the top of nylon membrane. Air bubbles should not be seen between the nylon membrane and the gel. A stack of absorbent paper were placed on top of Whatman 3MM paper, followed by a glass plate and a weight (about 500 g) respectively. The DNA transfer process was then allowed to proceed overnight. The nylon membrane positively charged from the southern blotting was exposed to UV radiation at 254 nm for 1 minute. This would form cross-links between the thymine residues of DNA with amine groups on the membrane. The DNA then fixed on the membrane and could be used for the hybridisation experiment.

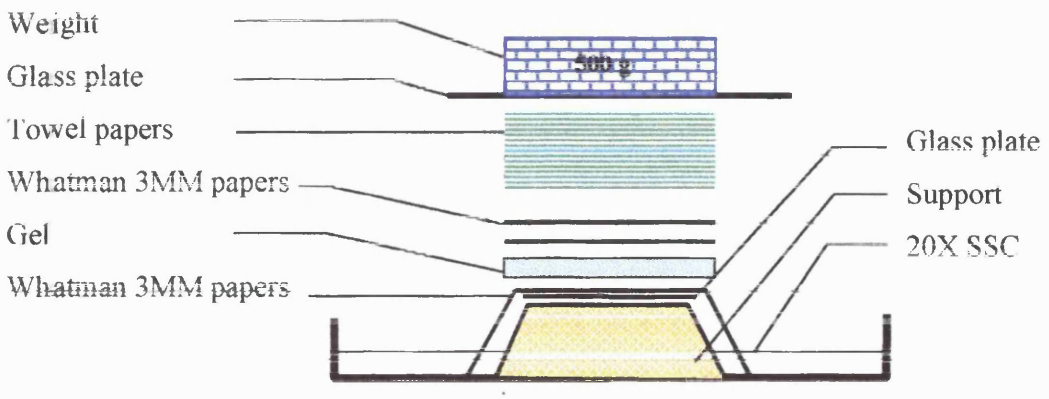


Figure 2.1 The stack of Southern blotting.

2.18 Pre-hybridisation, hybridisation and washing

2.18.1 Pre-hybridisation

After UV-crosslinking, the nylon membrane from the Southern blotting was placed on a nylon muslin mesh soaked with 2×SSC (0.15 M NaCl, 0.15 M trisodium citrate, pH 7.0). The membrane and muslin mesh were rolled and put in a hybridisation tube containing pre-hibridisation solution (5×SSC, 0.1% L-Sarcosine, 1% non-fat dried milk and 0.02% SDS) and incubated at 68°C in a hot-air oven with 5 rpm rotation for 3 hours.

2.18.2 Hybridisation

After pre-hybridisation, the membrane was placed in a hybridisation-plastic bag. To this, DIG DNA labelled probes in hybridisation solution (25 μl of labelled DNA in 5 x SSC, 0.1% L Sarcosine, 1% skimmed milk powder and 0.02 % SDS) kept at 60 C° were

added. Any bubble in the bag was eliminated as far as possible to prevent interference by the bubble in the hybridisation. Then the bag was sealed using a plastic sealer. After that, the membrane was incubated at 68°C with 5 rpm rotation for 16-18 hours.

2.18.3 Washing

The washing procedures were as following. At the end of the hybridisation, the hybridisation solution was discarded and the membrane was washed twice with washing solution 1 (2×SSC containing 0.1 % SDS), 5 minutes each time. Then the washing solution 1 was poured off. The membrane was transferred into the hybridisation tube and washed twice with washing solution 2 (0.1×SSC containing 0.1 % SDS) at 68°C with 11 rpm rotation, 15 minutes for each wash. Then the membrane was ready for colorimetric detection (as below).

2.19 Colorimetric detection of DIG DNA labelling probe with NBT and BCIP for the positive clone

Colorimetric detection was performed with nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) according to manufacturer recommendations (Boehringer Mannheim GmbH, Biochemica). First, the membrane was equilibrated in buffer 1 (0.1 M Tris-HCl buffer, 0.15 M NaCl) with gentle shaking for 5 minutes. The buffer 1 was discarded and blocking solution (0.5 % skimmed powder in washing buffer) was added. The membrane in the blocking solution was gently shaken for 30 minutes. Then the buffer 1 was discarded, and 1: 10,000 of anti-digoxigenin-AP in washing solution was added to the membrane. The membrane and anti-digoxigenin-AP in the buffer 1 were gently shaken at 10 rpm for 30 minutes. Then the membrane was washed twice in buffer 1 with gentle shaking, 15 minutes each time. After washing, the membrane was equilibrated in a detection buffer (0.1 M Tris-HCL pH 9.5, 0.1 M NaCl, 0.05 $MgCl_2$) with gentle shaking 5 minutes. The detection buffer was poured off, and, to

the membrane, detection solution (40mg/ml X-phosphate and 75 mg/ml in 20 % DMF of NBT in detection buffer) was added. The membrane immersed with detection solution was kept in a dark place for approximately 20 minutes. As a result of the precipitation of colorimetric substrate NBT and BCIP, the positive band should be seen as purple or brown. The principle of DNA hybridisation with DIG labelling DNA probe and chromogenic reaction is shown in Figure 2.2.

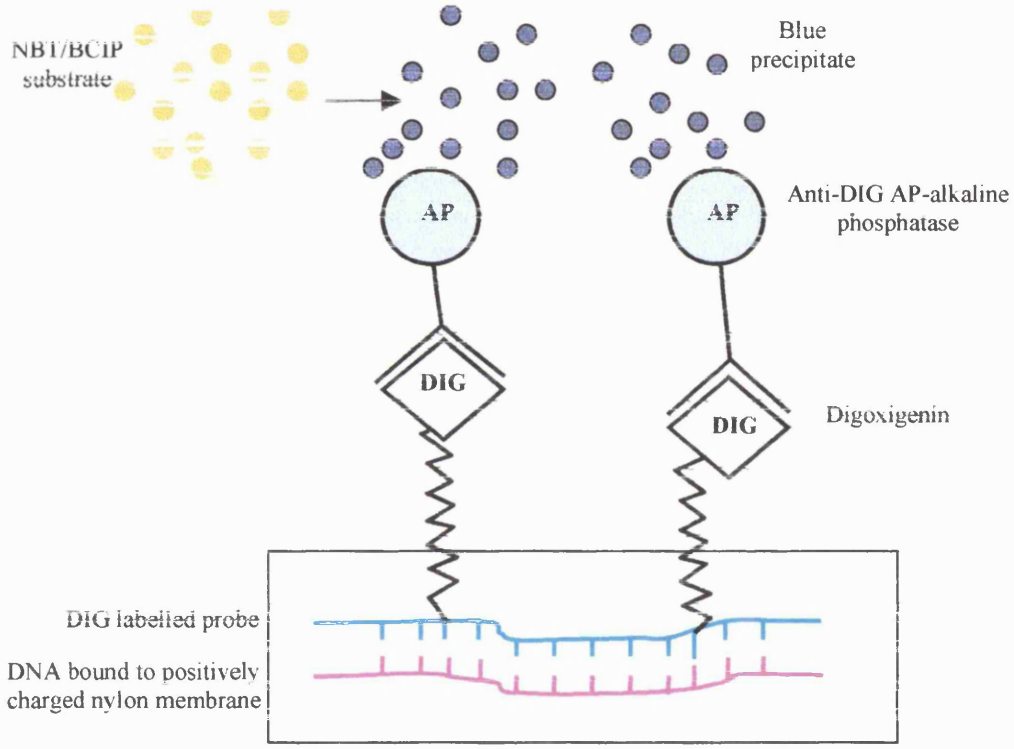


Figure 2.2 Schematic presentation of the colorimetric detection of DIG labelled probe (hybridised to the DNA fragment), with anti-DIG-AP and NBT/BCIP substrate.

2.20 Subgenomic library

A bulk DNA digest of DNA from *P. putida* NCIMB 10007 was performed in 250 μl reaction by using 80 μl DNA (100-200 ng/μl) from a genomic DNA isolation, 25 μl 10 x restriction buffer, 10 μl of restriction enzyme (*Pst*I *Bam*HI, *Kpn*I or *Sac*I) and 135 μl of sterilised water. The reaction was carried out at 37°C for overnight. A pUC19 DNA was digested by the same enzyme that was used for the bulk DNA digest of DNA from *P. putida* NCIMB 10007. The pUC19 restriction endonuclease digest was performed using 5 μl of pUC19 DNA (0.3 μg/μl), 5 μl of appropriate 10×restriction buffer, 5 μl of restriction enzyme and 31 μl of sterilised water. This was incubated at 37°C for 16-18 hours. The pUC19 restriction digest was treated by calf intestinal alkaline phosphatase (Boehringer Mannheim GmbH, Biochemica). After the digestion was completed, the bulk digest of *P. putida* DNA was separated into two parts and loaded in 1% agarose gel (2 lanes) and run. One lane of DNA in agarose gel, after electrophoresis, was excised and prepared for Southern blotting. Then followed the Southern blotting method, hybridisation and detection with the probe to identify the specific site of the desired fragments on agarose gel (see Figure 2.3).

The positive band from DIG DNA labelling detection the nylon membrane was aligned with the original agarose gel that contained other lanes of bulk whole DNA of *P. putida* digest. The section of the gel containing the whole *P. putida* DNA digest was excised at the same site as the hybridising band on the nylon membrane, and the DNA in the gel was extracted using QIAGEN gel extraction kit from QIAGEN Co Ltd.

After the purified genomic was obtained, the DNA fragments were subjected to ligate with appropriate vectors. This gave a result of subgenomic library of *P. putida* NCIMB 1007. This DNA recombinant library was consequently transformed into DH5α competent cells, and plated out on nutrient agar plates supplemented with ampicillin (100 μg/ml), IPTG; isopropylthio-β-D-galactoside (20μg/ml) and X-gal; 5-bromo-4-chloro-3-indolyl-β-D-galactoside (40 μg/ml), and grown overnight at 37°C. The recombinants,

which carried the *P. putida* DNA that were seen as white colonies, were transferred to a new ampicillin-IPTG-X-gal plate. This plate was kept as a master plate.

2.21 Screening of a positive clone

To screen the positive clone from the subgenomic library in the first round, 5 white colonies were pooled into a 10 ml nutrient broth and cultured for 16-18 hours. Next day, 1 ml of pooled culture was spun down at 14,000 rpm for 1 minute, the supernatant was discarded and the pellete was subjected to plasmid miniprep using QIAGEN Spin Miniprep Kit (see section 2.11). The pooled of recombinant plasmids isolated from the plasmid miniprep were then digested with a restriction endonuclease (*Bam*HI, *Kpn*I, *Pst*I or *Sac*I) and run on agarose gel. This gel was later subjected to Southern hybridisation (see section 2.18) and detection with DIG labelling pQR203 or pQR277 probe (see section 2.19). Once a group of 5 transformants was identified as a positive series, next in thre second round screening, members of this positive series were identified individually for a true possitive clone harbouring the recombinant plasmid of interest. This second round screening was as the same as the first round screening except a single colony was inoculated into 5 ml nutrient broth.

2.22 DNA Sequencing

Determination of the sequence of the 4485-bp *Bam*HI and 4201-bp *Kpn*I insert was carried out by subcloning different fragments of the inserts and primer walking (see Figure 2.4). The recombinant plasmids obtained from the subcloning strategy were sequenced by using both M13-21 forward and M13-20 reverse primer. For primer walking strategy, primers were designed and used for DNA sequencing in ambiguous regions. Nucleotide sequencing was done by UCL sequencing centre and the Oswel DNA sequencing laboratory, University of Southampton, UK.

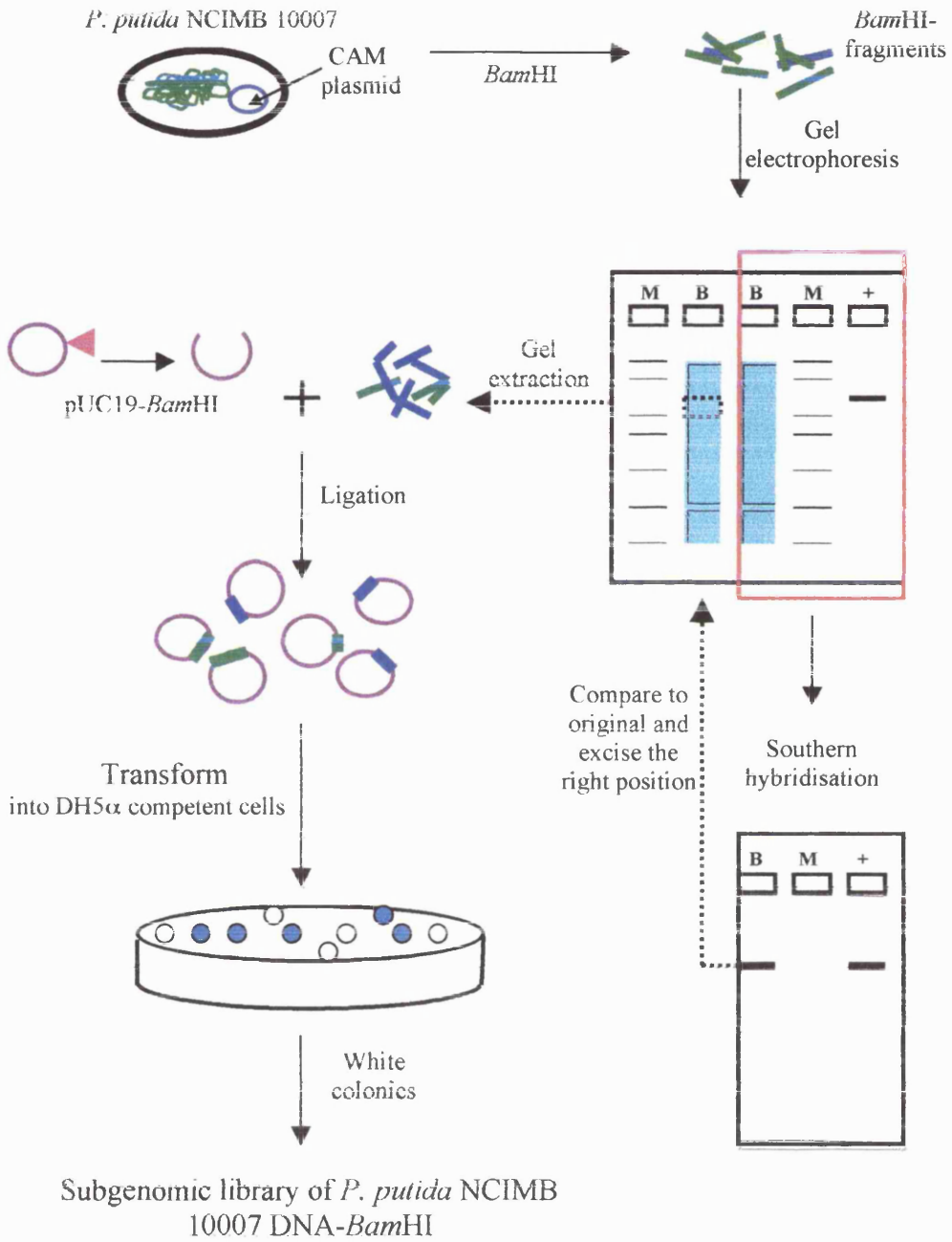


Figure 2.3 The depicted description of subgenomic library of *P. putida* DNA *Bam*HI digest (B).

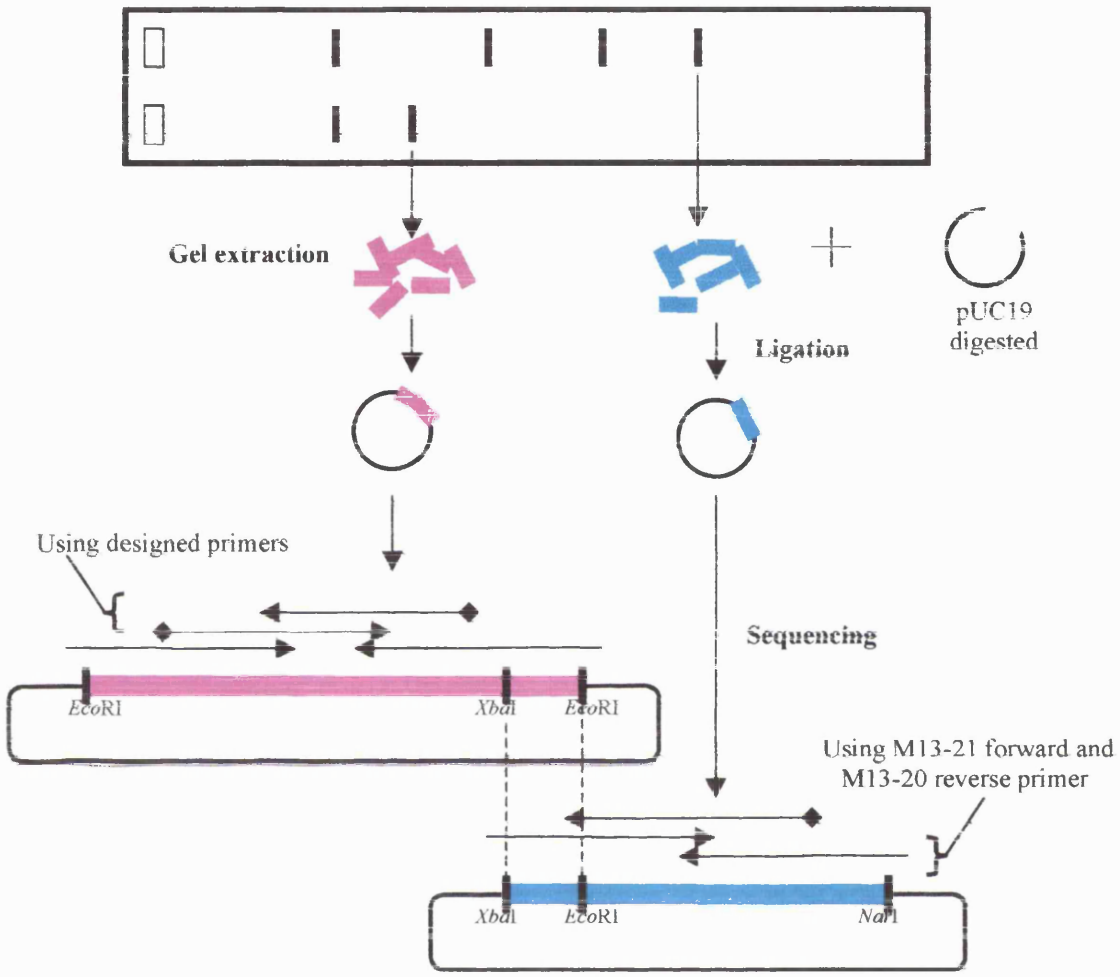


Figure 2.4 Schematic presentation of strategies for the DNA sequencing; using M13-21 forward and M13-20 reverse (in subcloning different fragments) and using designed primers (in primer walking).

2.23 Computer analysis of DNA sequence

The orientation of the fragment from DNA sequence data was aligned and extended from the previous published sequence (Aramaki *et al.*, 1994). Sequence data obtained from the sequencing were assembled by CAP (Contig Assembly Program)

(www.infobiogen.fr) and BLAST2 sequences (www.ncbi.nlm.nih.gov/blast/). A search of DNA and protein similarities of the DNA sequence data was mainly carried out in the BLAST (Basic Local Alignment Search Tool) program of the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/blast/). The nucleotide sequence formatting and analysis were also performed by the Sequence Manipulation Suite (SMS) program (Stothard, 2000) (www.bioinformatics.org/sms). The other excellent Internet tools used in the analysis of DNA and protein sequences in this study are listed in Table3.

Table 2.3 A list of useful Internet tools in the analysis of DNA and protein sequences.

Internet tools	Descriptions
http://psort.nibb.ac.jp/	PSORT; protein localisation prediction
http://molbiol-tools.ca	Molecular biology online analysis tools
www.cbs.dtu.dk/services/SignalP	SignalP prediction
www.cbs.dtu.dk/services/TMHMM2.0/	TMHMM; prediction of transmembrane helics in proteins
www.ccsi.com/firstmarket/cutter/cut2.hlm/	Webcutter 2.0; restriction map analysis
www.ebi.ac.uk/clustalw/	ClustalW; multiple alignment
www.expasy.ch/tools/plotparam.html	Plot Para tool; amino acid composition
www.expasy.org	Web-site based DNA and protein analysis tools
www.ncbi.nih.nlm.gov/gorf	Open reading frame finder
www.ncbi.nlm.nih.gov/blast	Blast homology search
www.bioinfo.rpi.edu	Mfold; RNA and DNA folding analysis
www.tigrblast.tigr.org/cmr-blast/GC	Analysis of G+C content in the third codon position
www.ualberta.ca/~stothard/javascript	Web-site based DNA and protein analysis tools

2.24 Cloning of *orf2* and *orf4* into pCR 2.1-TOPO vector

In order to construct the genes; the *orf2* and *orf4*; and to obtain the His-tagged Orf2 monooxygenase and His-tagged Orf4 hydrolase, *orf2* and *orf4* were individually amplified by polymerase chain reaction (PCR) using the pQR416 DNA as a template.

Two primers were designed to amplify *orf2*, that is to say, PMO1: 5′ GACATATGAAATGCGGATTTTTC-3′ (*Nde*I site underlined) and PMO2: 5′ GTGGATCCTCAGCCCATTCGAAC- 3′ (*Bam*HI underlined). As well as *orf4*, two primers were designed to amplify them as PHY1: 5′ CACATATG CTCACTTCATCACAG-3′ (*Nde*I site underlined) and PHY2: 5′ GTGGATCCTCA GCATGCTCTGCC-3′ (*Bam*HI site underlined). To generate the Orf2 and Orf4 with a histidine-tagged tail, it is necessary to remove the stop codons from the *orf2* and *orf4*. The stop codons from both genes were replaced with *Xho*I site, and the genes were cloned into pET21a expression vector. The genes were then translated in frame with the six histidine residues and the terminate condon down stream of pET21a vector (Novagen). Two designed primers which were used as a downstream primer (to replace PMO2 and PHY2) to generate the enzymes with His-tagged were PMO3: 5′-GTCTCGAGGCCCATTCGAACCTTC-3′ (*Xho*I site underlined) and PHY3: 5′-GTCTCGAGGCATGCTCTGCCGTG-3′ (*Xho*I site underlined). The PCR were carried out in the following conditions;

1) a cycle of 1 minute at 94°C;
2) 25 cycles of 30 seconds at 94°C, 30 seconds at 60°C and 2 minutes at 72°C;
3) a cycle of 10 minutes at 72°C.

The PCR products were run on 1% agarose gel. DNA well-distinctive bands at approximately 1.1 and 1.0-kb were excised from the gel, and subjected to gel extraction using QIAGEN kit to isolate the DNA fragments. Then these four fragments were ligated to pCR 2.1-TOPO vector (Novagen) and transformed into DH5α competent cells. The recombinant plsmids were designed as pQR418, pQR419, pQR420 and pQR421, pCR2.1-TOPO vector containing *orf2*, *orf2* with the deletion of its stop codon, *orf4* and *orf4* with the deletion of its stop codon respectively (see also Figure 2.5, 2.6, 2.7 and 2.8). These recombinant vectors were also sequenced to confirm that there was no silent mutation during PCR amplification.

2.25 Subcloning of *orf2* and *orf4* into pET21a expression vector

Cultures of the trasformants harbouring pQR418, pQR419, pQR420 and pQR421 were carried out in a 5 ml nutrient broth with 100 rpm shaking at 37°C for 16-18 hours. The cells were pelletted by centrifugation and subjected to plasmid minipreps using the QAIGEN miniprep kit. The plasmid of pQR418 and pQR420 were digested with *Nde*I and *Bam*HI; the plasmid of pQR419 and pQR421 were digested with *Nde*I and *Xho*I. After the digest reactions were completed, all DNA samples were run on 1% agarose gel. The DNA fragment at 1.1 kb from pQR418 and pQR420-*Nde*I and *BamH*I digests, and 0.95 kb from pQR419 and pQR421-*Nde*I and *Xho*I digests on the gel were excised and extracted by the QAIGEN gel extraction kit. The two isolated fragments of *orf2* and *orf4* were then ligated into pET21a-*Nde*I and *Bam*HI digest. In contrast, two isolated fragments of *orf2* and *orf4* with the deletion of their stop codons were ligated into pET21a-*Nde*I and *Xho*I digests. All four recombinant plamids were transformed into DH5α competent cells. These four recombinant plasmids were named pQR422 (pET21a-*orf2*), pQR423 (pET21a-*orf2* with the deletion of its stop codon), pQR424 (pET21a-*orf4*) and pQR425 (pET21a-*orf4* with the deletion of its stop codon) (see Figure 2.5, 2.6, 2.7 and 2.8). The DH5α cells carrying these expression plasmids were also maintained in 20% glycerol and kept at -70°C. Isolation of the recombinant plasmids of pQR422, pQR423, pQR424 and pQR425 for other purposes was done by using the QAIGEN miniprep. The pQR424 and pQR425 were transformed into different hosts BL21(DE3), BL21(DE3)pLysS and BL21(DE3)CodonPlus-RP for an expression purpose and to obtain satisfactory levels of protein expression.

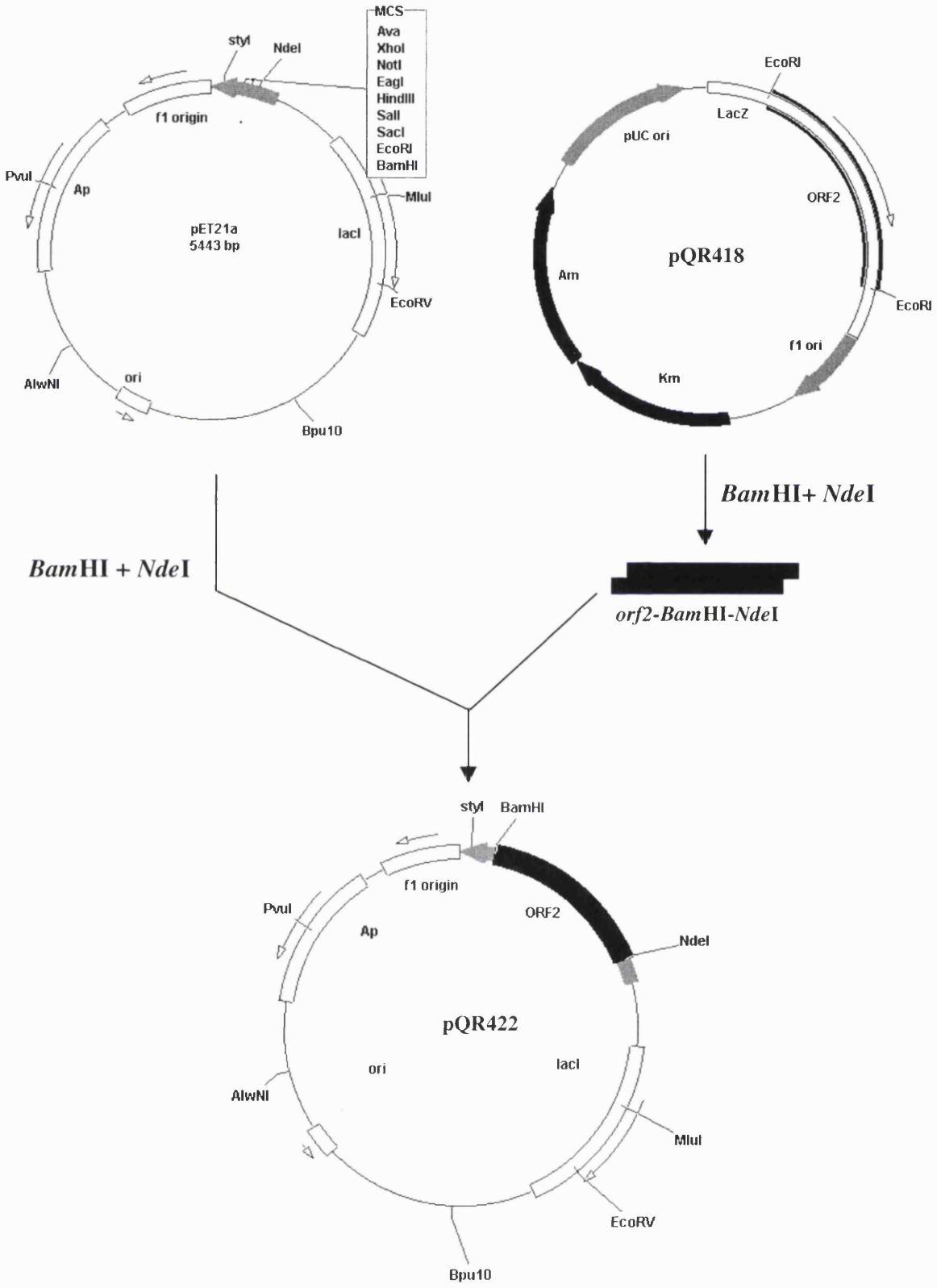


Figure 2.5 Construction of pQR422.

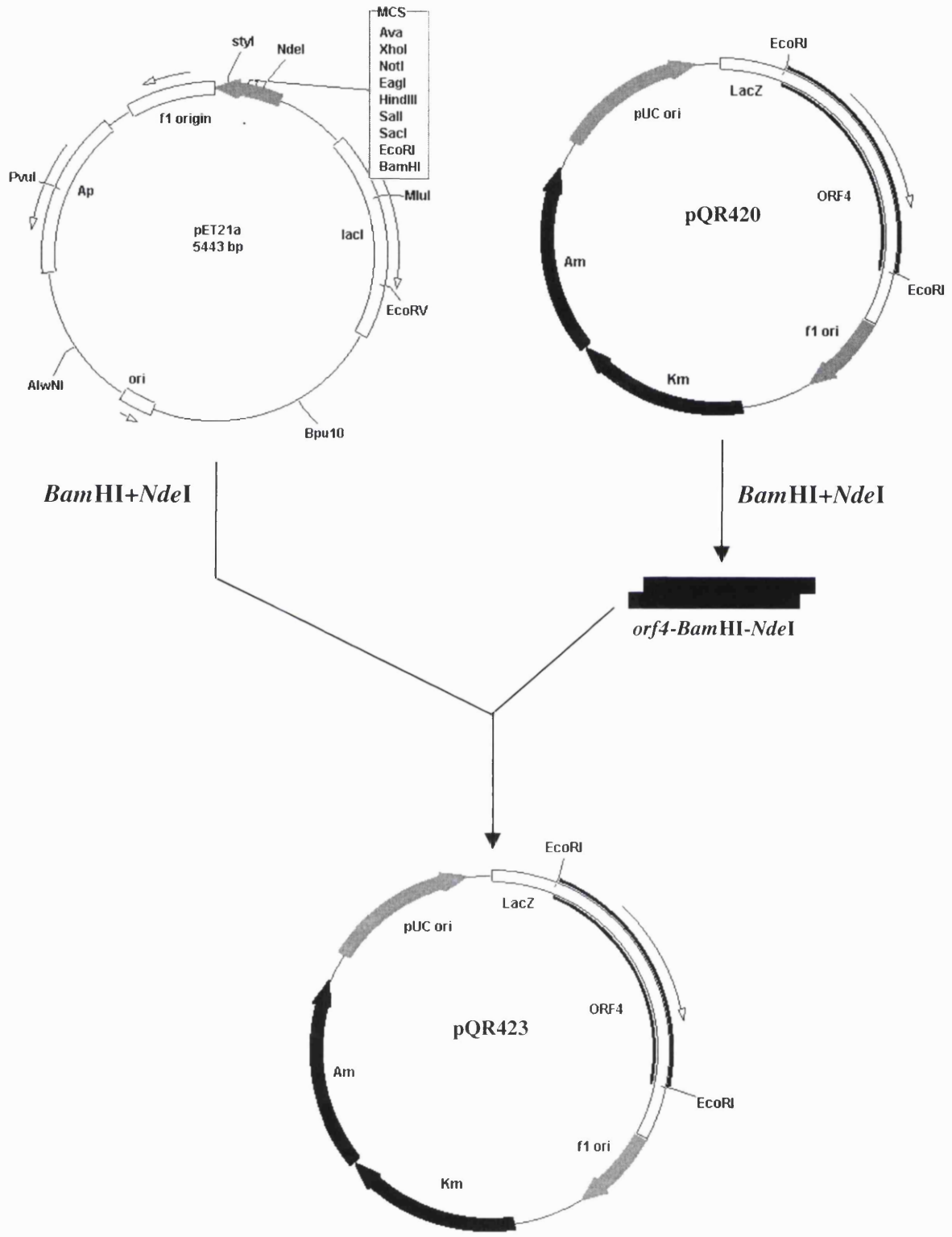


Figure 2.6 Construction of pQR423.

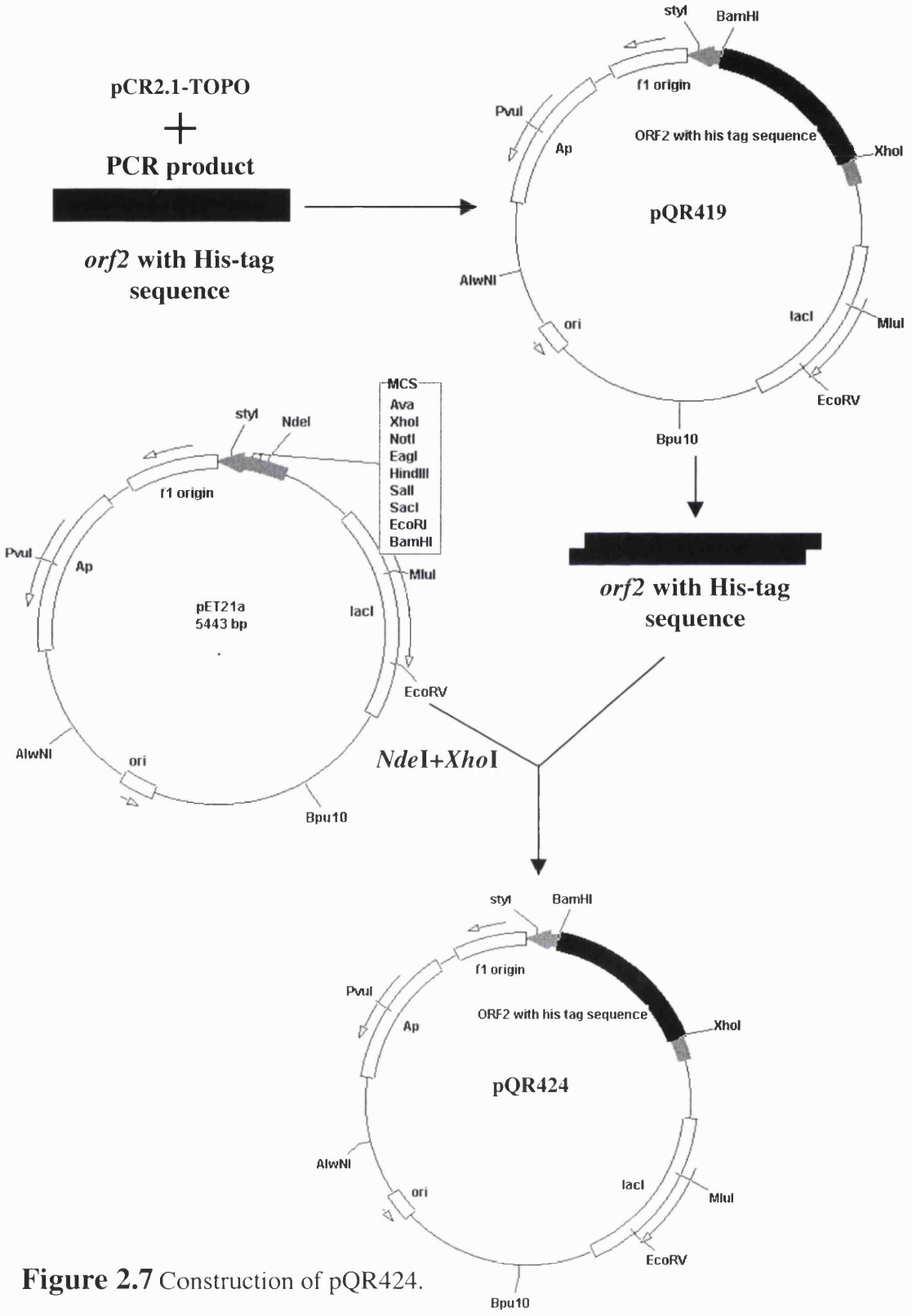


Figure 2.7 Construction of pQR424.

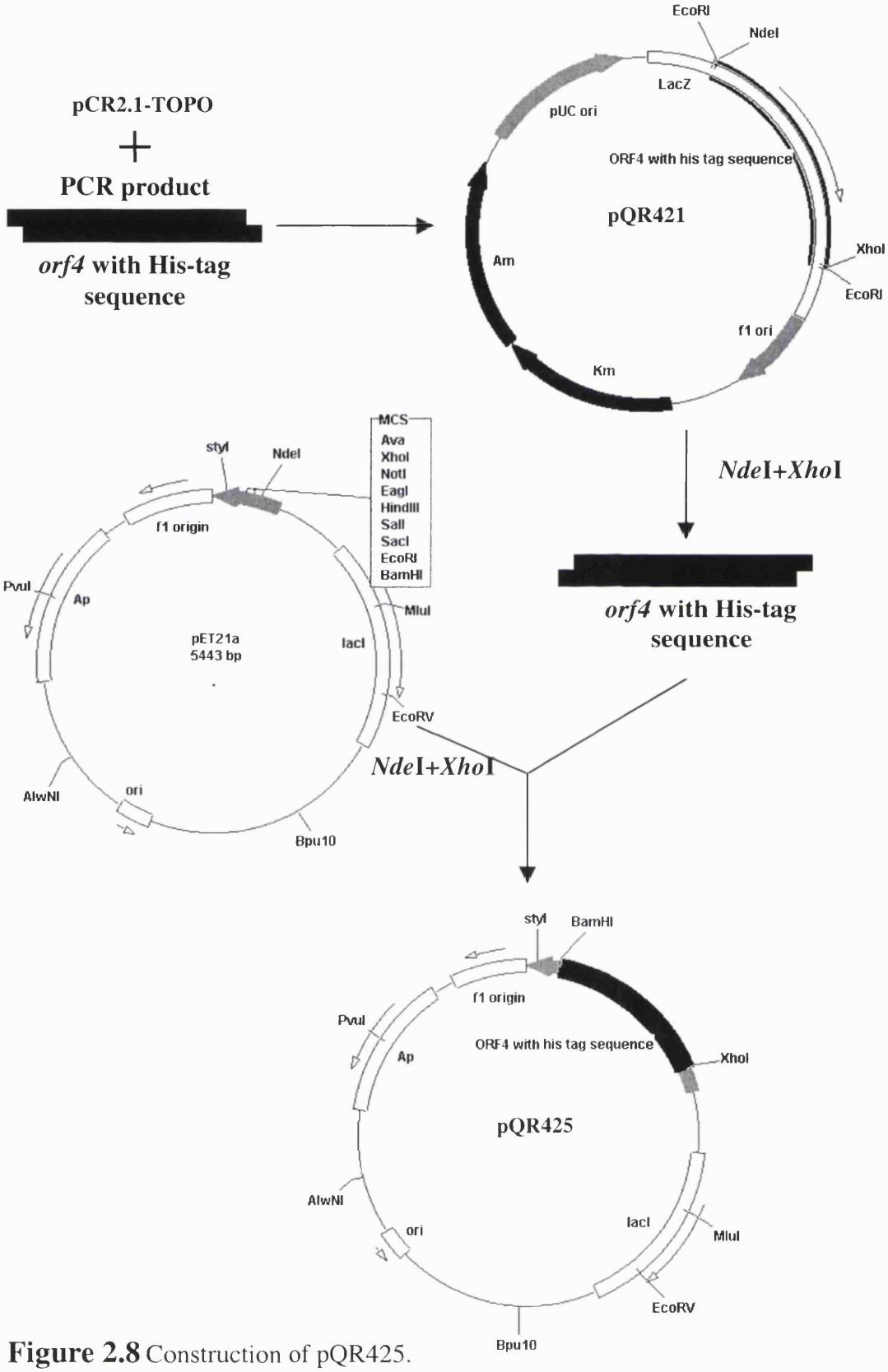


Figure 2.8 Construction of pQR425.

2.26 Costruction of *camRDCAB* in array

2.26.1 Construction of pQR426 (pUC19-*camR-camD*)

camRD gene was costructed by the deletion of genes, *orf1234*, in pQR416 clone. The pQR416 was digested with *Eco*RI yielding ~4.3 kb gene array of *orf1234* and ~5.6 kb pUC19-*camRD*, pUC19 vector containing *camRD* gene. The fragment of 5.6 kb-pUC19-camRD was re-ligated with T_4 DNA ligase and transformed into DH5α competent cells. The trasformant from coloremetric screening, a white colony, harbouring 5.6 kb *camRD,* was designed as pQR426 (see Figure 2.9).

2.26.2 Cloning of *camCAB*

The genes of *camC*, *camA* and *camB* were amplified from the genomic DNA of *P. putida* NCIMB 10007 using PCR protocol. Two designed primers were used to amplify ~3.0 kb *camCAB* from the genomic DNA of *P. putida.* These two primers were PCAMN: 5′-GGCATATGACGACTGAAACCATACAAAGC-3′ (*Nde*I restriction site underlined) and PCAMC: 5′GCGAATTCGGTTTACCATTGCCTATCGGGAAC-3′ (*Eco*RI restriction site underlined). The PCR condition was as follows:

1) denaturation at 94°C for 3 minutes for 1 cycle;
2) denaturation at 94°C for 30 seconds; anealing at 60 °C for 30 seconds; and polymerisation at 72°C for 6 minutes (for 30 cycles).
3) polymerisation at 72°C for 10 minutes.

The PCR product of *camCAB* was then cloned into pCR2.1-TOPO vector (Invitrogen). The ligation mixture contained 2 μl of fresh PCR product, 1 μl of salt solution (1.2 M NaCl, 0.06M $MgCl_2$), 2 μl of sterile water and 1 μl of pCR2.1-TOPO vector. The mixture was incubated at room temperature for 5 minutes, and then pipetted into the tube containing DH5α competent cells. The tube was left on ice for 30-45

minutes. After that, the cells were heat shocked at 37°C for 30 seconds and placed on ice immediately. 400 μl of nutrient broth was added to the tube, and it was incubated at 37°C with 100 rpm shaking for 1 hour. After incubation, 100 μl of transformant culture was spread out on ampicillin, X-Gal and IPTG agar plate and incubated at 37°C for 16-18 hours. White colonies from the plate were picked and subjected to the plasmid miniprep to isolate the recombinant plasmid of pCR2.1-TOPO-*camCAB*. The recombinant plasmid was designed as pQR427. The pQR427 was then digested with *Eco*RI to cut off the *camCAB* gene. The 3 kb *camCAB* gene was then subcloned into pUC19-*Eco*RI digest. The clone of *camCAB* in pUC19 was called pQR428 (see Figure 2.10).

2.26.3 Construction of plasmid containing *camRDCAB*

First, pQR426 (pUC19-*camR*-*camD*)was digested with *Acc*65I and *Bsr*GI. Then the digest was run in agarose gel. The fragment of 2.1 kb *camRD* was excised from the gel and subjected to the QIAGEN gel extraction. Then the fragment of 2.1 kb *camRD*-*Acc*65I-*Bsr*GI was ligated into pQR428-*Bsr*GI (*Acc*65I digest gives compatible cohesive ends to *Bsr*GI digest; *Acc*65I: 5′-G/GTACC-3′ and *Bsr*GI: 5′-T/GTACA-3′). The recombinant plasmid containing *camR*, *camD*, *camC*, *camA* and *camB* was designed as pQR429, pUC19 harbouring *camR*, *camD*, *camC*, *camA* and cam*B* as an array as in *cam* operon (see Figure 2.11).

2.27 Construction of pQR430 (pUC19/ORF1234)

The recombinant plasmid containing novel genes of *orf1*, *orf2*, *orf3* and *orf4* was constructed from pQR416 plasmid. First, pQR416 was digested with *Bgl*II and *Bam*HI and run on a agarose gel. The fragment of about 6.7 kb was excised from the agarose gel and purified by QAIGEN gel extraction kit. Such fragment containg pUC19 DNA, *orf1*, *orf2*, *orf3* and *orf4* was subjected to self-ligation, and this recombinant plasmid was designated pQR430 (see Figure 2.12).

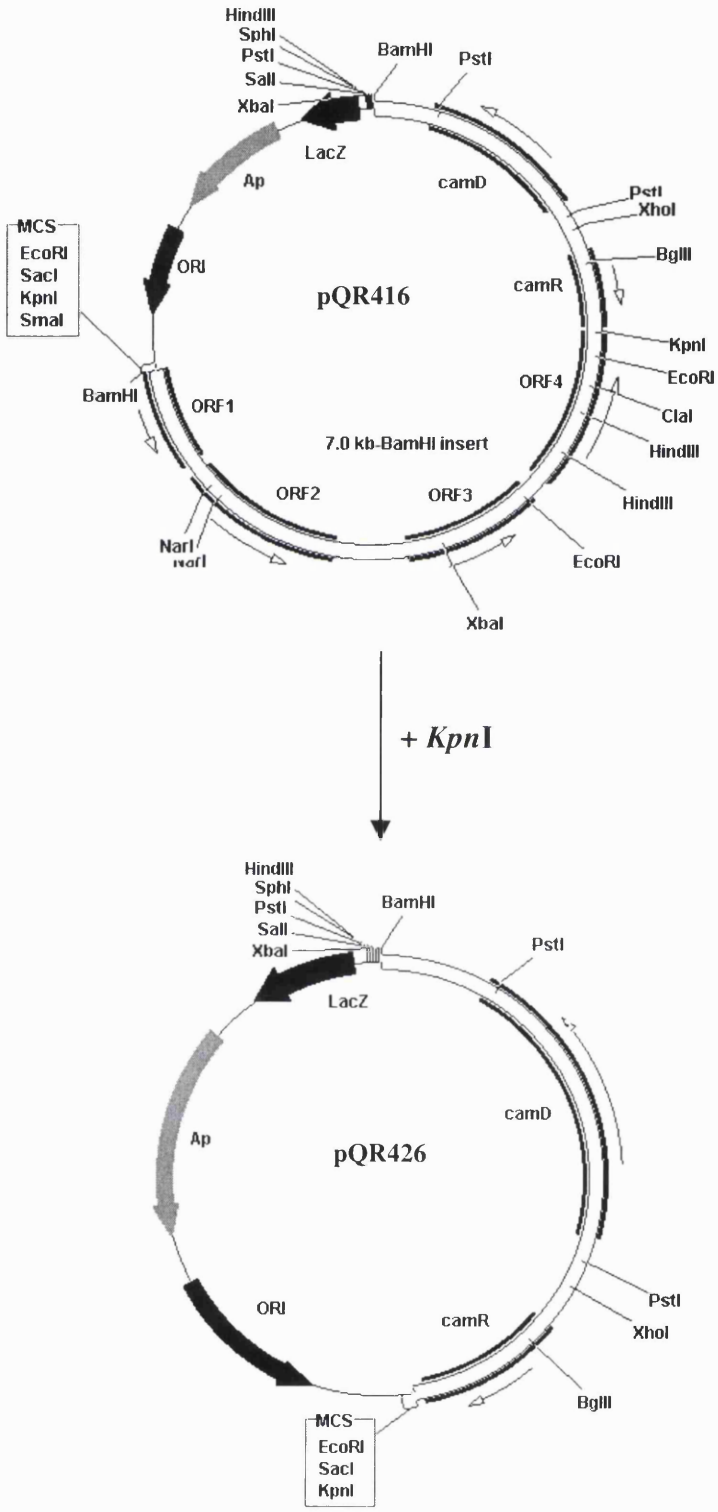


Figure 2.9 Construction of pQR426.

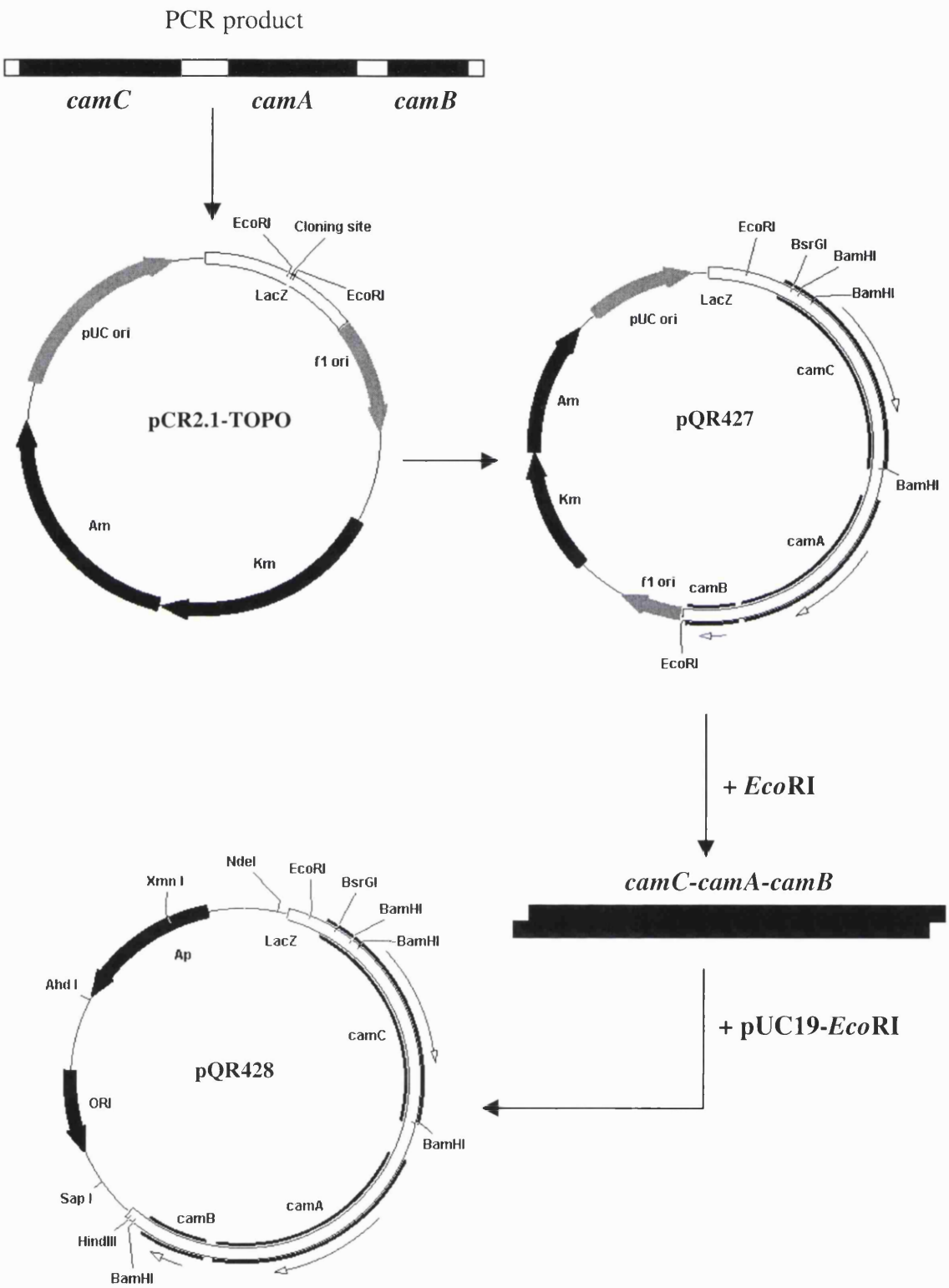


Figure 2.10 Construction of pQR427 and pQR428.

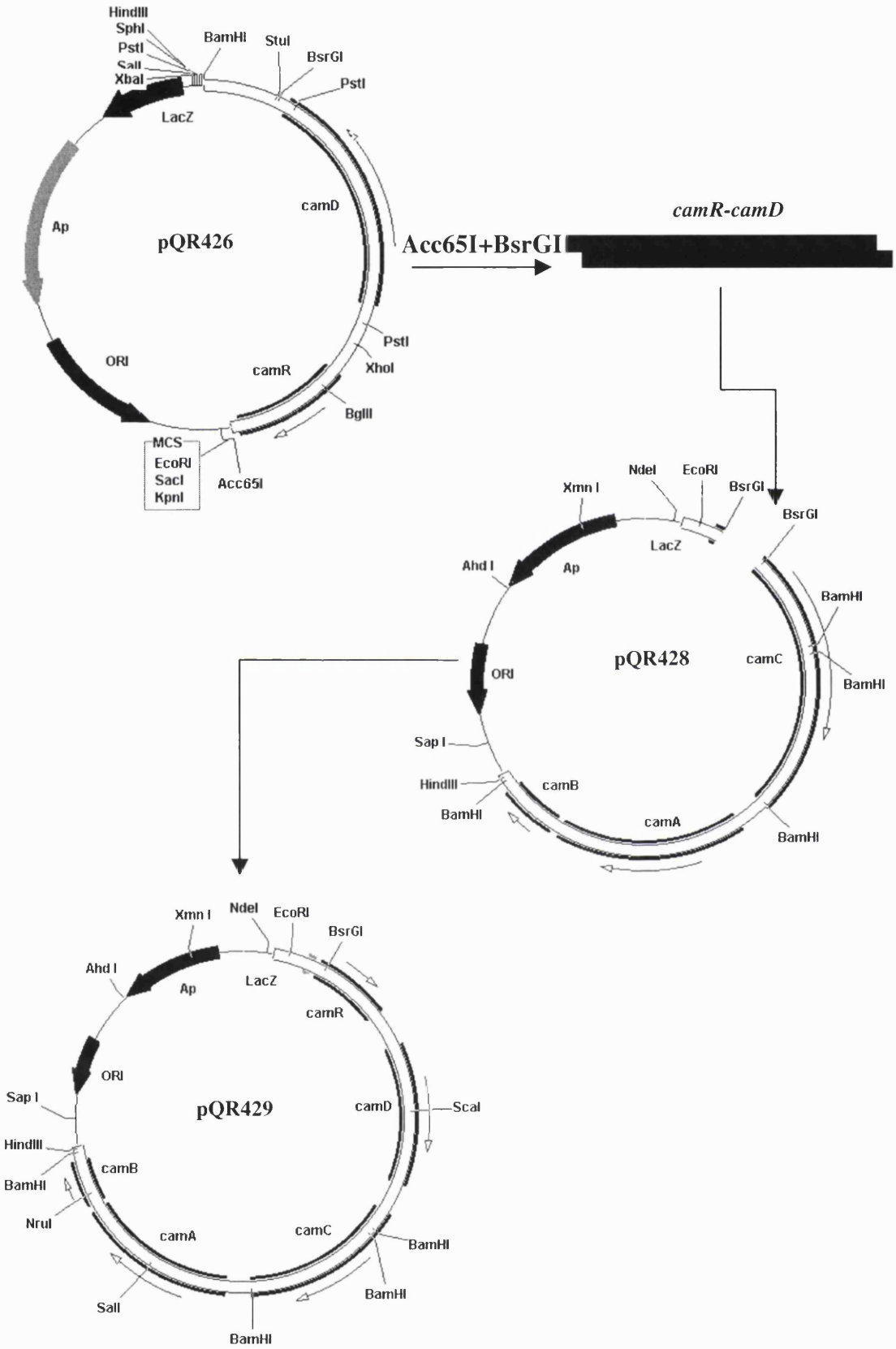


Figure 2.11 Construction of pQR429.

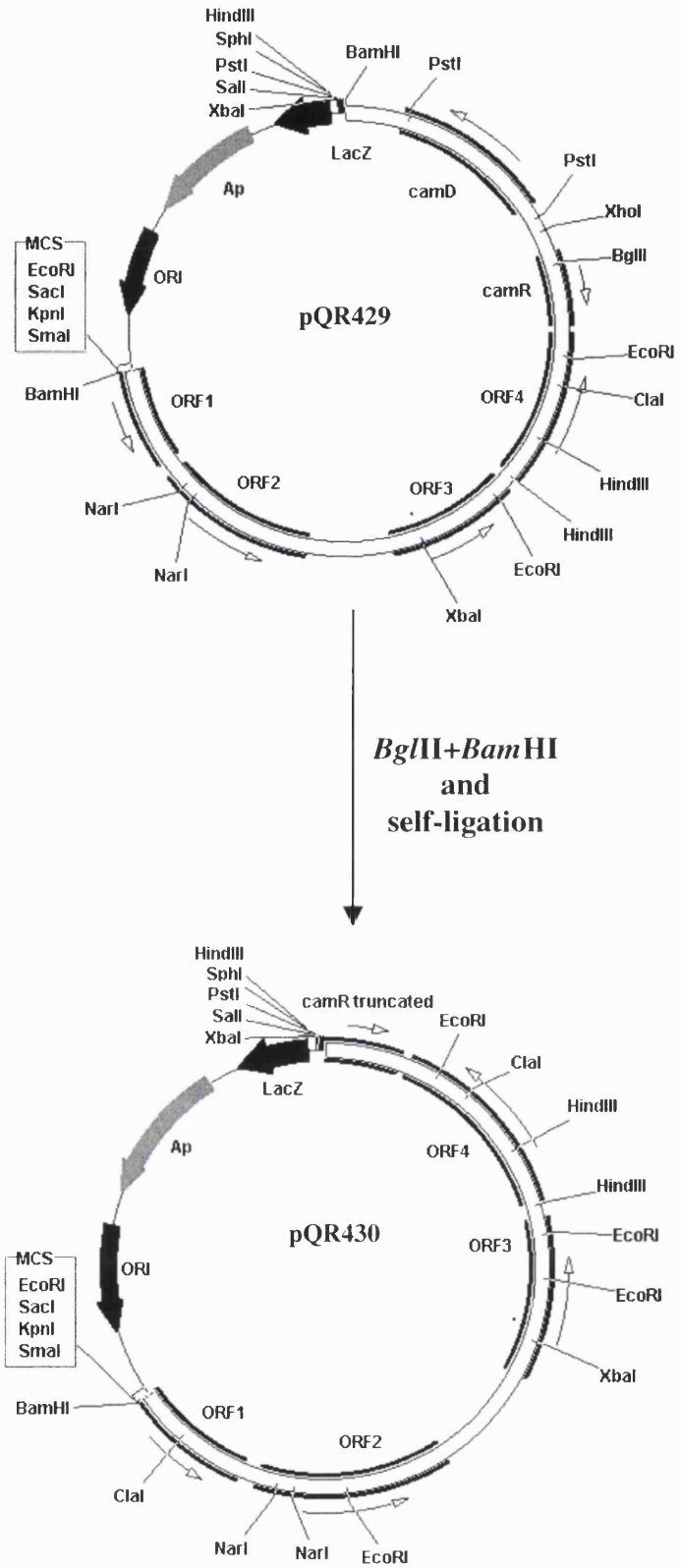


Figure 2.12 Construction of pQR430.

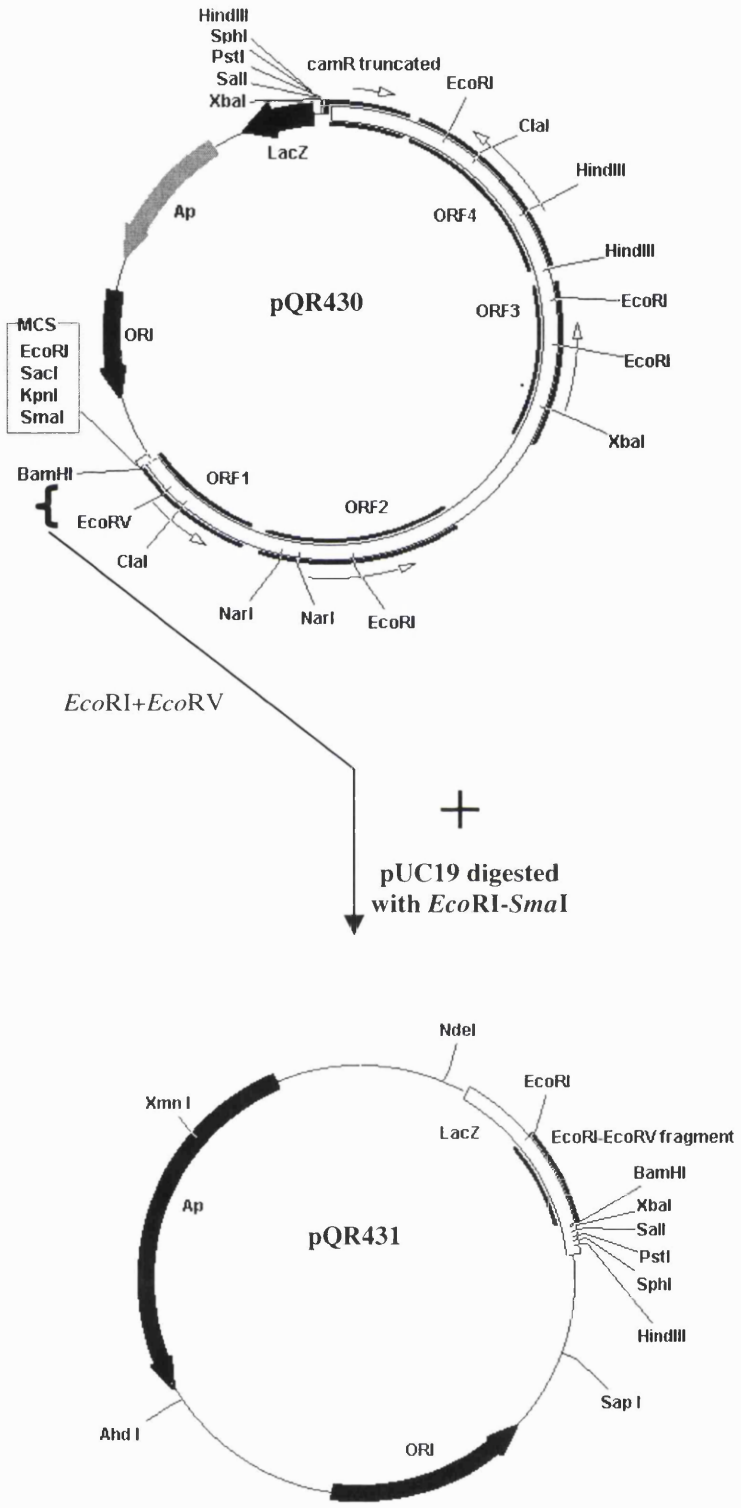


Figure 2.13 Construction of pQR431.

2.28 Construction of pQR431

The pQRX14 is recombinant plasmid of pUC19 with a 192 bp *Eco*RI-*Eco*RV fragment (a DNA fragment at the 5′end of *orf1*). This construction was constructed by subcloning of the 192 bp *Eco*RI-*Eco*RV fragment from the pQR430 intopUC19-*Eco*RI-*Sma*I digest (see Figure 2.13). First, pQR430 plasmid was digested with *Eco*RI and *Eco*RV. This DNA was then run on agarose gel, and a band of 192-bp *Eco*RI-*Eco*RV fragment was excised from the gel and extracted using QIAGEN gel extraction kit. Finally, the 192 bp *Eco*RI-*Eco*RV fragment was ligated with pUC19 *Eco*RI-*Sma*I digest.

2.29 Induction of the protein expression

E. coli host cells for protein expression, containing the recombinant, were inoculated into 5 ml nutrient broth, supplemented with 100 mg/ml ampicilin and 34 mg/ml chloramphenicol, at 37°C with 100 rpm shaking for 16-18 hours. One millilitre of the overnight culture was added to 20-ml nutrient broth supplemented with antibiotics as in the starter culture. The culture was incubated at 37°C with 250 rpm shaking until the optical density (OD_{600}) of the culture reached 0.6-0.8 or approximately 2 hours. Then IPTG was added to a final concentration of 1 mM to induce protein expression. The cells were grown for a further 3 hours with 250 rpm shaking at 37°C. In order to collect the cells, the culture was centrifuged at 7,000 rpm for 5 minutes.

2.30 Harvesting and purification of His-tagged proteins

The host cells harbouring pQR423 and pQR425 were inoculated in 5 ml nutrient broth and incubated at 37°C with 200 rpm shaking for 16-18 hours. This inoculumn was added into 200 ml nutrient broth supplemented with 100 mg/ml ampicilin and 34 mg/ml choramphenical. The culture was incubated at 37°C with 200 rpm shaking until OD_{600}

0.6-0.8, then added IPTG to a final concentration of 1.0 mM. Then the culture was incubate at 37°C with 200 rpm shaking until it reached late exponential phrase or OD_{600} about 4.0-5.0. Cells were spun down by Sorvall centrifuge SB3, 11,000 rpm for 10minutes. The supernatant was discarded and pelletted cells were collected for soluble fraction purification (to purify cytoplasmic proteins) or membrane fraction purification (to purify membrane bound proteins).

2.30.1 Soluble fraction purification

To purify soluble fraction, the pelletted cells were resuspended in 4 ml ice-cold binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) per 100 ml culture volume. The 50 mg/ml of lysozyme was added to the sample to a final concentration of 200 μl/ml, and incubated at 30°C for 15 minutes. The cells were disrupted by sonication on ice by MSE sonicator. To do this, the cells were sonicated for 10 seconds and interspersed with 10 seconds on a pre-chilled NaCl/ice(8-10 cycles). The disrupted cell sample was centrifuged at 10,000 g for 10 minutes. The supernatant containing soluble protein of His-tagged proteins was removed to a new tube for further protein purification. The pellet in this step can be used in membrane fraction purification.

2.30.2 Membrane fraction purification

To purify membrane fraction, the pellet from the previous section was used. This pellet was resuspended in 20 ml binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) per 100 ml culture volume, and spun down at 5,000 g for 15 minutes. Then the supernatant was removed, and the pellet was suspended in 5 ml binding buffer containing 6 M urea (per 100 ml culture volume). The mixture was incubated on ice for an hour, and then spun down at 16,000 g for 30 minutes. After that, the supernatant was separated and filtered through a sterile 0.45 micron filter. The flow-through containing membrane bound proteins was then ready for his-tagged protein purification.

2.30.3 His-tagged protein purification

The protein purification of the His-tagged proteins was carried out using Quick 900 Cartridges; Ni^{+2} charged resin column (Novagen). First, the cartridge was attached to a 10-ml syringe and equilibrated with 6 ml binding buffer. The supernatant from previous sonication was loaded onto the nickel column at a rate of approximately 2 drops per second. The cartridge was washed with 20 ml binding buffer and then followed with 10 ml of wash buffer (60 mM imidazole, 0.5 M NaCl, 20nM Tris-HCl pH 7.9). In the final step, the protein was eluted with 40 ml elute buffer (1M imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9) in the last step. The protein solution of the his tagged proteins was dialysed with 50 nM Tris-HCl pH 7.5, 1 mM 2-mercaptoethanol and 15% glycerol. The protein solution was stored at –20°C.

2.31 Determination of protein by Bradford method

The protein determination was done according to the Bradford method using a Protein Assay Kit (Bio-Rad Laboratories) and Bovine serum albumin as a standard. The Bio-Rad solution was diluted 5 times (mixed Bio-Rad solution 1: 4 distilled water), and unknown protein (100-200 μl) was pipetted in the diluted solution. The amount of unknown protein was determined at λ_{595} spectrometrically. The absorbance was then compared to the standard curve to calculate the protein concentration of the unknown.

2.32 Protein analysis by SDS-PAGE

Analysis of protein expressions was carried out by SDS-polyacrylamide gel electrophoresis (Laemmli, 1970). ProtoGel (1.5 M Tris-HCl, 0.384% SDS, pH 8.8) (National diagnostics, Kimberly Research Atlanta, Georgia) and acrylamide were used to prepare 12.5% resolving gel. Ammonium persulfate and TEMED (N,N,N′, N′-tetramethylethylenediamine) were added to the gel at 5.0% and 0.05% (v/v) respectively. Stacking gel 6.0% was prepared by using ProtoGel, ProtoGel Stacking buffer (0.5M Tris-HCl, 0.4% SDS, pH6.8) (National diagnostics, Kimberly Research Atlanta, Georgia), 10% ammonium persulfate and 0.1% TEMED: total volume (v/v) of stacking gel.

Cells (1 ml) were pelletted by centrifugation. The cells were resuspended with appropriate amount of 2×sample buffer (0.5 M Tris-HCl pH 6.8, 10% (v/v) SDS, 20% (v/v) SDS, 20% glycerol, 2% (v/v) 2-mercaptoethanol and 0.02 mM bromphenol blue). For example, if the A_{600} of cells is 1.5, 150 μl of the sample buffer will be added to the cells. Then the sample was boiled at 95°C for 10 minutes. After that 20 μl of sample was loaded on a SDS-PAGE gel in 10×10 cm vertical electrophoresis unit. The SDS-PAGE gel was stained with Coomassie brilliant blue (0.25% w/v of Coomassie Brilliant Blue in 45% methanol and 10% of glacial acetic acid) for about 45 minutes. The stain was removed from the gel by destaining in methanol:water (1:1 v/v) 10% acetic acid solution for 2 hours.

2.33 Stardard curve of *p*-nitrophenol

A standard curve of *p*-nitrophenol was determined by plotting graph of the absorbance of *p*-nitrophenol at 400 nm in various concentrations. The concentrations of *p*-nitrophenol used were 0, 0.03, 0.43, 0.066, 0.096, 0.13, 0.19, 0.33, 0.41, 0.45, 0.51, 0.6, 0.78, 0.91, 1.07, 1.36 and 1.63 mM (in Tris-Cl buffer, pH 7.9). From the graph, we can calculate the concentration of *p*-nitrophenol accurately between the reading absorption

(A_{400}) 0-0.78, but the reading absorption greater than 0.78 would not be reliable. An extinction coefficient of *p*-nitrophenol is 13.5 $mM^{-1}cm^{-1}$ in 0.2 M Tris-Cl buffer, pH 7.9, which determined from the slope of the linear plot of the absorbance of *p*-nitrophenol versus the concentration (see Figure 2.15).

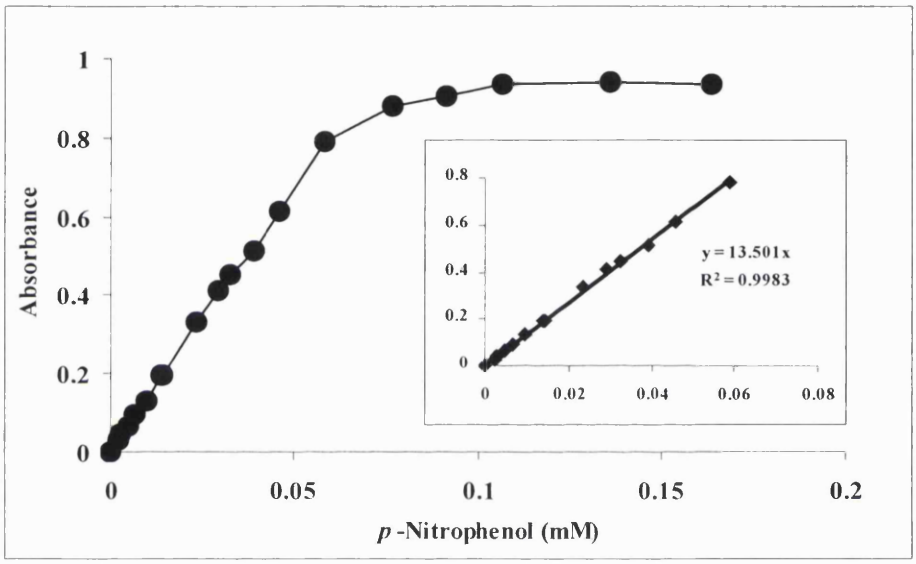


Figure 2.14 Standard curve of *p*-nitrophenol in 0.2 M Tris-HCl buffer, pH 7.9 with an extinction coefficient of 13.5 $mM^{-1}cm^{-1}$.

2.34 Determination of the hydrolysed paraoxon by ORF4 hydrolase

The hydrolysis of paraoxon yields *p*-nitrophenol and diethyl phosphorate. The product of *p*-nitrophenol can be monitored spectrometrically at 400 nm. The molar coefficient, ε, for *p*-nitrophenol at λ_{400} is 13.5 $mM^{-1}cm^{-1}$. In determination of an unknown *p*-nitrophenol, if the concentration of unknown *p*-nitrophenol is too high, it can be quantified by using a dilution of the unknown.

Determination of *p*-nitrophenol concentration can be calculated as in the following equation.

$$A = \varepsilon c l$$

A: absorbance

ε: extinction coefficient ($mM^{-1}cm^{-1}$)

c: concentration of sample (mM)

l : path length of sample (cm)

This fomula is Beer 's law; the absorption of a molecule at a particular wavelength is related to the concentration of the molecule in the solution.

An initial rate of the hydrolysis of the substrate by hydrolase enzyme can be determined as following;

$$\nu = -\Delta[S]/\Delta t$$

$$\text{or} \quad \nu = -\Delta A/\varepsilon l \Delta t$$

Δ[S]: concentration of substrate determined at $T_0 - T_{final}$ (mM)

ΔA: absorbance of substrate determined at T_0-T_{final} (mM)

Δt; $T_0 - T_{final}$ (min)

ν; initial rate ($mM\ min^{-1}$)

2.35 The whole cell activity of Orf4 hydrolase towards paraoxon, parathion and methyl parathion

E. coli cells harbouring pET21a-*orf4* or *E. coli* pET21a (as a control) was grown in 5 ml nutrient broth supplemented with 100 mg/ml of ampicillin and 34 mg/ml of chloramphenicol at 37°C with 200 rpm shaking overnight. Then 1 ml of the overnight culture was inoculated into 20 ml nutrient broth, shaking 200 rpm at 37°C until OD_{600} 0.6-0.8 (or about 2 hours). IPTG was added to final concentration 0.5 mM, and the culture was incubated at 37°C and shaken at 200 rpm for a further 3 hours or until the cells reached mid-log phrase orOD_{600} of the cells reached about 2.5-2.8.

The cells were divided into two tubes; the first tube contained 15 ml of culture and the second tube contained 5 ml of culture. The pellet cells were harvested by centrifugation at 7,000 rpm for 5 minutes. The pellet cells from 5-ml culture were then used to determine dry cell weight. The pellet cells were dried in a hot air oven for 16-18 hours, whereas the pellet cells from 15-ml culture were subjected to further phosphotriesterase assay using a whole cell.

The assay tube consisted of the cell pellete resuspended in 1 ml of 0.2 M Tris-Cl pH 7.9, and parathion, methyl parathion or paraoxon in 20% methanol was added to a final concentration of 2.0 mM. The reaction tube was incubated at 37°C with 200 rpm shaking.

To determine the formation of *p*-nitophenol product, 50-μl sample from the reaction tube was taken at various times and spun down at 7,000 rpm for 30 seconds. Then 40 μl of sample was pipetted into 2-ml quartz cuvette (1-cm-path-length) containing 1 ml of 0.2 M Tris-Cl, pH 7.9. The absorbance of *p*-nitrophenol at λ_{400} was determined by using Beckman DU 5000 spectophotometre.

Each assay tube was paired with the control, BL21(DE3)pLysS cells harbouring pET21a incubated with tested substrates. Moreover, the *p*-nitrophenol formation from the control tube was also determined, as the same as that from the assay tube.

2.36 Enzyme assay of lactone hydrolase; a whole cell activity

Lactonase activity of Orf4 hydrolase towards γ-butyrolactone, gluconolactone, pantolactone and δ-valerolactone was determined according to Fishbein and Bessman (1966). The reaction tube contained 2.0 mM lactone and 10 mg (calculated from the dry cell weight) of whole cells suspended in 1 ml of 0.2 M Tris-Cl buffer, pH 7.9 (the recombinant cells were obtained as described insection 2.35). The reaction was incubated at 37°C with 100 rpm shaking for 30 minutes. After the incubation, the tube was centrifuged at 14,000 rpm, and the supernatant was transferred to a new tube. To this, 1 ml of alkaline hydroxylamide reagent (1:1 volume of 2 M hydroxylamide hydrochloride and 3.5 M sodium hydroxide) followed by 2 ml of ethanoloic ferric chloride reagent (1:1 volume of 10% ferric chloride in 4 M HCl and 95% ethanol). Both reagents were freshly prepared. The tube was spun down at 14,000 rpm for 5 minutes. The solution was applied to a plastic cuvette and determined spectrophotometrically at λ_{max} of 520 nm. This absorbance is a colour complex formed by a hydroxamic acid reacted with ferric ion (see Figure 2.14). Each lactone hydrolysis assay was paired with the lactone incubated in the 0.2 M Tris-Cl buffer, pH 7.9 alone and the lactone incubated with *E. Coli* harbouring pET21a. The decreas of the reading absorption at 520 nm of colour complex was calculated as the percentage of the amount of lactone hydrolysed.

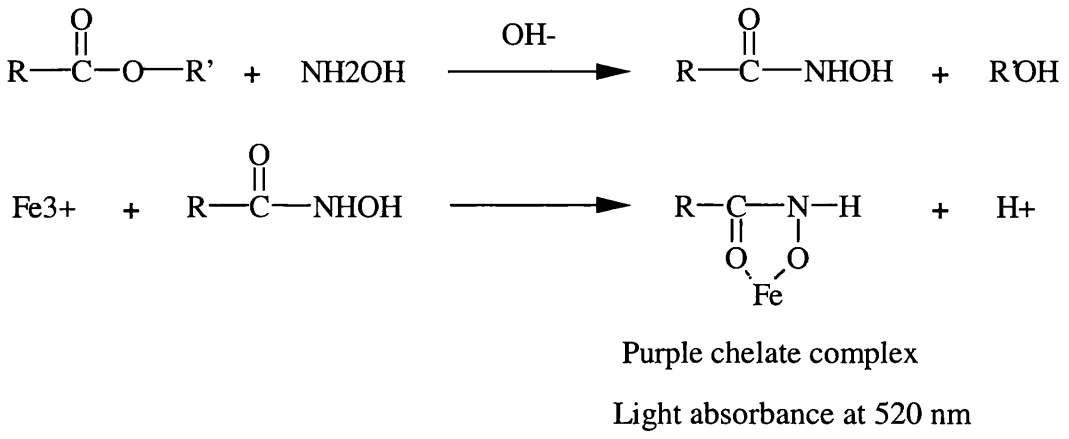


Figure 2.15 Formation of the colour complex of hydroxamic acid (the reaction of lactone with hydroxylamide in alkaline solution) and ferric ion.

Chapter 3

Southern hybridisation, DNA sequencing and open reading frame analysis

This chapter will present the results in determination of the novel sequence on the left-hand side of *cam* operon and the novel sequence obtained from the Southern hybridisation with the pQR277 probe derived from 3,6-diketocamphane 1,6-monooxygense.

3.1 Southern hybridisation, DNA sequencing and ORF analysis of a 7.0 kb *Bam*HI fragment

3.1.1 Southern hybridisation: identification of a 7.0 kb *Bam*HI fragment on the left-hand side of *cam* operon

The nucleotide sequence of 5′ end of *camR*, which was reported by Aramaki *et al.*, 1994, is useful data to make a probe by PCR amplification of a 222 bp segment constituted on the far left of *cam* operon. The recombinant plasmid containing this 222-bp segment is pQR203, which is constructed by Dr. Sejal Patel. This 222 bp probe in pQR203 can then be used to find DNA fragments that are adjacent to the 5′ end of the *cam* operon. Data from a Southern hybridisation analysed by Dr. John Ward showed that a 4.7 kb *Pst*I, 7.0 kb *Kpn*I, 4.75 kb *Xho*I and 7.0 kb *BamH*I fragment from genomic digests hybridised to the pQR203 probe. In the study described in this thesis, total DNA of *P. putida* NCIMB 10007 digested with *Pst*I, *Kpn*I and *Bam*HI were used to construct subgenomic libraries of *P. putida* NCIMB 10007 DNA-*Pst*I, -*Kpn*I and -*Bam*HI respectively. However, the 7.0 kb *Bam*HI fragment was chosen and focused on because we could extend the sequence information from the left hand side of *cam* operon in more detail.

In the study, a comparative strategy was used to narrow down a number of possible clones in the construction of this library. Because there is no procedure in isolating a large plasmid and obtain it in sufficient amounts, instead the total DNA is used. First, the whole cell DNA from *P. putida* NCIMB 10007 was digested with *Pst*I, *Kpn*I or *Bam*HI. After the digestion was completed, the bulk digest of *P. putida* DNA was divided into two parts and loaded on 1% agarose gel (2 lanes) and run. After electrophoresis, one lane of the DNA in the agarose gel was excised and subjected to a Southern blotting. After blotting, the DNA bound nylon membrane was hybridised with the probe from pQR203 to identify the specific site of desired fragments (the DNA *Bam*HI-fragment on the left-hand side of *cam* operon) on the agarose gel. Then the nylon membrane with the positive band from the DIG DNA labelling detection was aligned with the original agarose gel that contained another lane of the *P. putida-Pst*I, *Kpn*I or *Bam*HI digest. The DNA in the gel at the site identified by Southern hybridisation was excised. The DNA was purified from the gel and used for ligation into pUC19 *Pst*I, *Kpn*I or *Bam*HI to form the *Pst*I, *Kpn*I or *Bam*HI library respectively.

For example, in the *Bam*HI library, a 7.0 kb *Bam*HI fragment from the whole *P. putida* DNA was extracted and purifed. This DNA was ligated into alkaline phosphatase treated pUC19 cut with *Bam*HI , and the ligation mixture was used to transform to *E. coli* DH5α competent cells. The white-blue colony screening of recombinant cells was performed. The recombinant cells (the white colonies) were picked and streaked on new AIX (containing ampicillin, IPTG and X-Gal) plates. In the first round of screening, these cells were selected and pooled in groups of 5 colonies into one tube of 10 ml nutrient broth. After an overnight, the cultures were spun down, and the cells were pelleted and subjected to the miniprep procedure. Recombinant DNA from these minipreps was digested with *Bam*HI, run on an agarose gel, and screened by Southern hybridisation using the pQR203 probe labelled with DIG to identify the group of pooled culture that contained a positive clone. Next in the second round, the members in a group of pooled culture, which was identified as positive in the first round of screening, was identified individually (repeated procedures as the first round) for the true positive clone (see also the Chapter 2 section 2.21).

More than 400 transformants from the *Pst*I library, 300 transformants from the *Kpn*I library and 1000 transformants from the *Bam*HI library were identified. Only one transformant obtained from *Bam*HI library was identified as a positive clone. The recombinant plasmid from series number p138 was digested with *Bam*HI, and showed an insert of about 7.0 kb (see Figures 1 and 2). The recombinant plasmid with a 7.0 kb *Bam*HI insert was designated pQR416 and kept in our collection. Although the *Bam*HI insert was identified as a large insert, an approximately 3 kb of the 7.0 kb sequence is a known sequence of *camRD*. This was verified by restriction mapping of the *Bam*HI-insert (see next section).

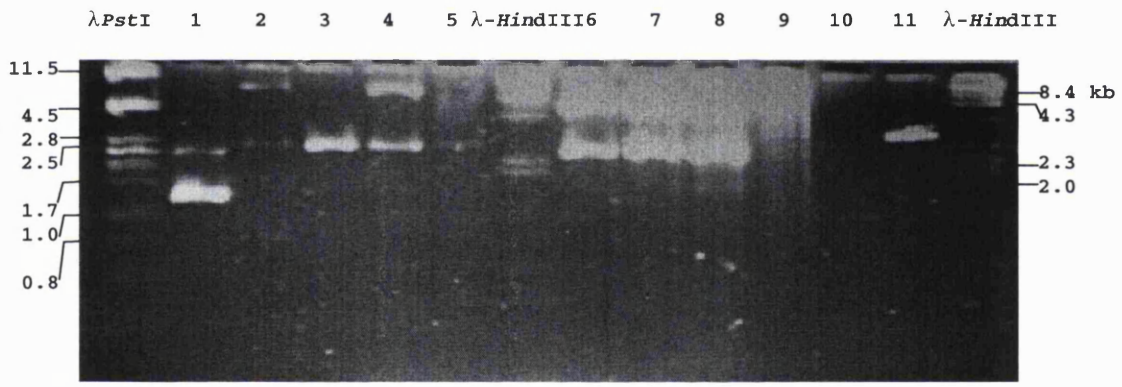


Figure 3.1 Agarose gel of recombinant plasmids (series number p132-139) from the *Bam*HI library and *P. putida* genomic DNA digested with *Bam*HI. λ DNA digested with *Pst*I and *Hind*III were used as a marker. From left to right, Lane 1: p132; 2: p133; 3: p134; 4: p135; 5:p136; 6: p137; 7: p138, 8: p139; Lane 9 and 10: *P. putida* genomic DNA *Bam*HI digest; and Lane 11: pQR203 *Hin*dIII digest. This was the second round of screening (from the positive pool). The picture shows that the bands of the 7.0 kb *Bam*HI inserts migrates between DNA molecular weight markers of 4.5 kb of λ-*Pst*I and 8.4 kb of λ-*Hin*dIII (pUC19 vector has a DNA fragment of 2.68 kb).

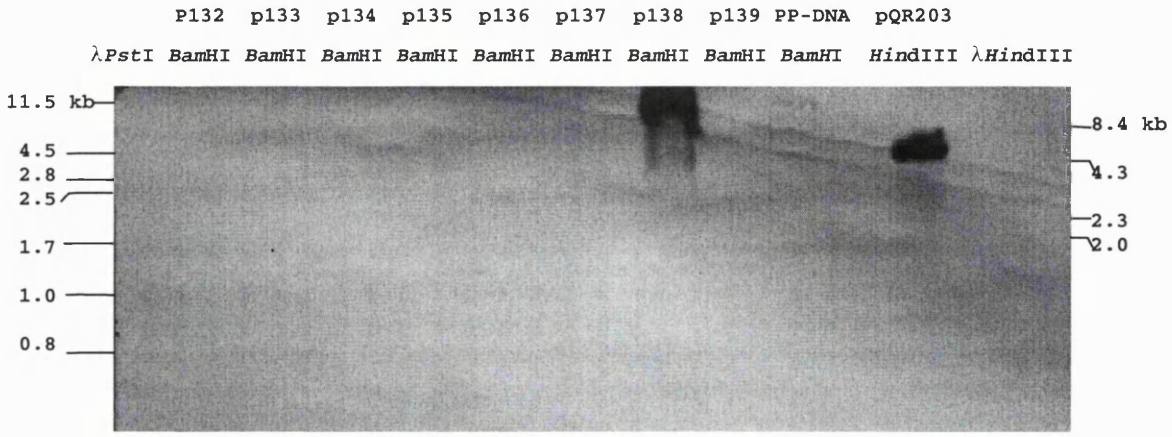


Figure 3.2 The Southern hybridisation of the agarose gel in Figure 1. The p132-p139 *Bam*HI: the recombinant plasmids series p132-p139 digested with *Bam*HI; PP-DNA-*Bam*HI: *P. putida* NCIMB 10007 digested with *Bam*HI; pQR203-*Hin*dIII: pQR203 plasmid linearised with *Hin*dIII; λ-*Pst*I and λ-*Hin*dIII: λ-*Pst*I and λ-*Hin*dIII DNA markers. The p138 shows a positive hybridisation with the 222-bp probe from pQR203.

In Figure 3.2, the series of recombinant plasmids (p132-p139) from the *Bam*HI library were digested with *BamH*I, run gel electrophoresis, subjected to southern blotting, hybridised with the digoxigenin-labelled 222 bp pQR203 probe (the segment on the far left of the known *cam* operon), and detected with colorimetric substrates NBT and BCIP. A positive band was seen on the insert fragment of p138 *Bam*HI located at the same size as *P. putida* NCIMB 10007 genomic DNA digested with *Bam*HI. The brown colour of the positive band (p138 *Bam*HI) appeared within about 2 hours, and showed as intense as the control of pQR203 *Hind*III digest (about 3.15 kb).

3.1.2 Restriction map of the *Bam*HI insert

To verify that the 7.0 kb *Bam*HI insert of p138 is an extention of the left end of *cam* operon, we carried out restriction mapping of this 7.0 kb *Bam*HI insert. The restriction mapping was carried out using different restriction enzymes to digest the p138 recombinant plasmid. *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Xho*I, *Cla*I, *Kpn*I, *Sma*I and *Xba*I were used for this restriction mapping. Most of these restriction endonucleases can cut the DNA region on the left-hand site of *cam* operon (based on previous publication), except *Cla*I and *Xba*I. The agarose gel of these digests is shown in Figure 3.3 and 3.4.

To determine the site where the pQR203 probe bound to the 7.0 kb *Bam*HI insert, the electrophoresis gel of the previous digests was also carried on further investigation in a Southern hybridisation (see Figures 3.5 and 3.6).

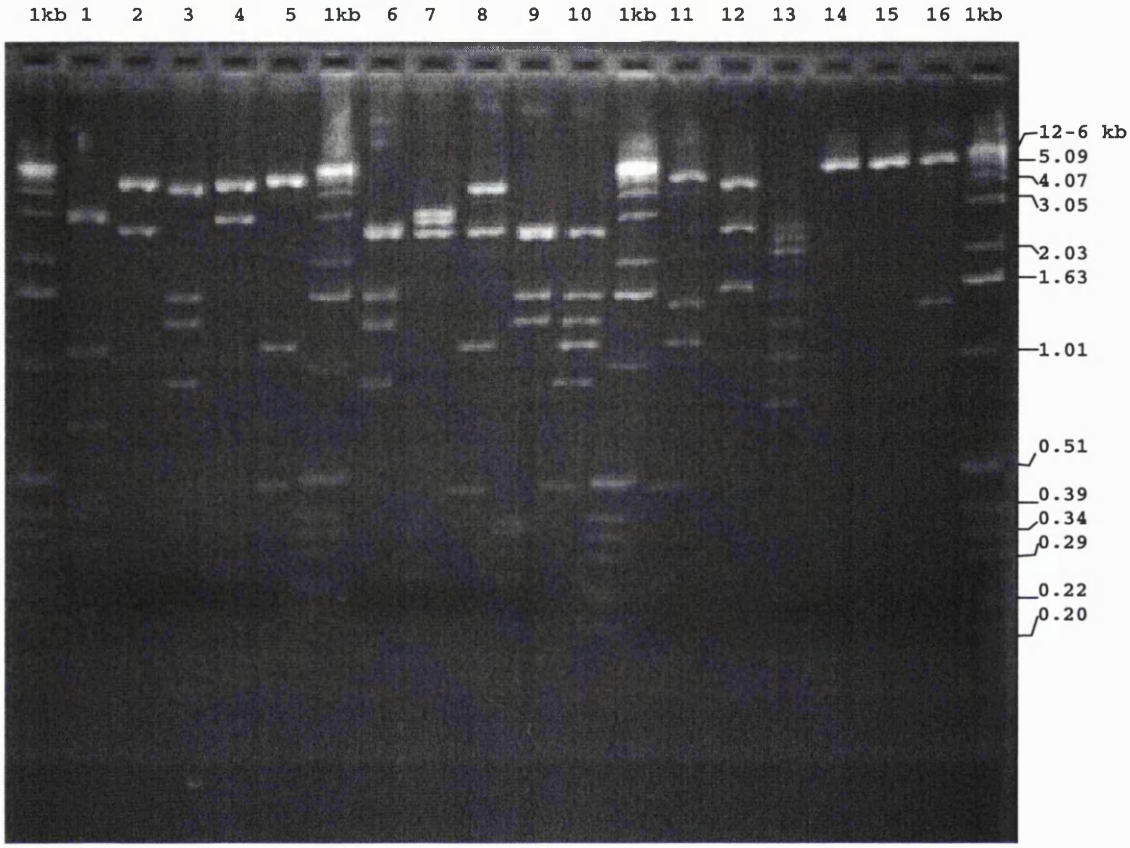


Figure 3.3 Agarose gel of p138 digests. The p138 was digested with different restriction enzymes, and with single and double digest for restriction mapping. The p138 digested with lane 1: *Nar*I; 2: *Bam*HI; 3: *Eco*RI; 4: *Hin*dIII; 5: *Pst*I; 6: *Bam*HI-*Eco*RI; 7: *Bam*HI-*Hind*III; 8: *Bam*HI-*Pst*I; 9: *Eco*RI-*Hin*dIII; 10: *Eco*RI-*Pst*I; 11: *Hin*dIII-*Pst*I; 12: *Bam*HI-*Xho*I; 13: *Bam*HI-*Cla*I; 14: *Cla*I; 15: *Xba*I; 16: *Cla*I-*Xba*I ; and 1 kb: 1 kb DNA ladder marker.

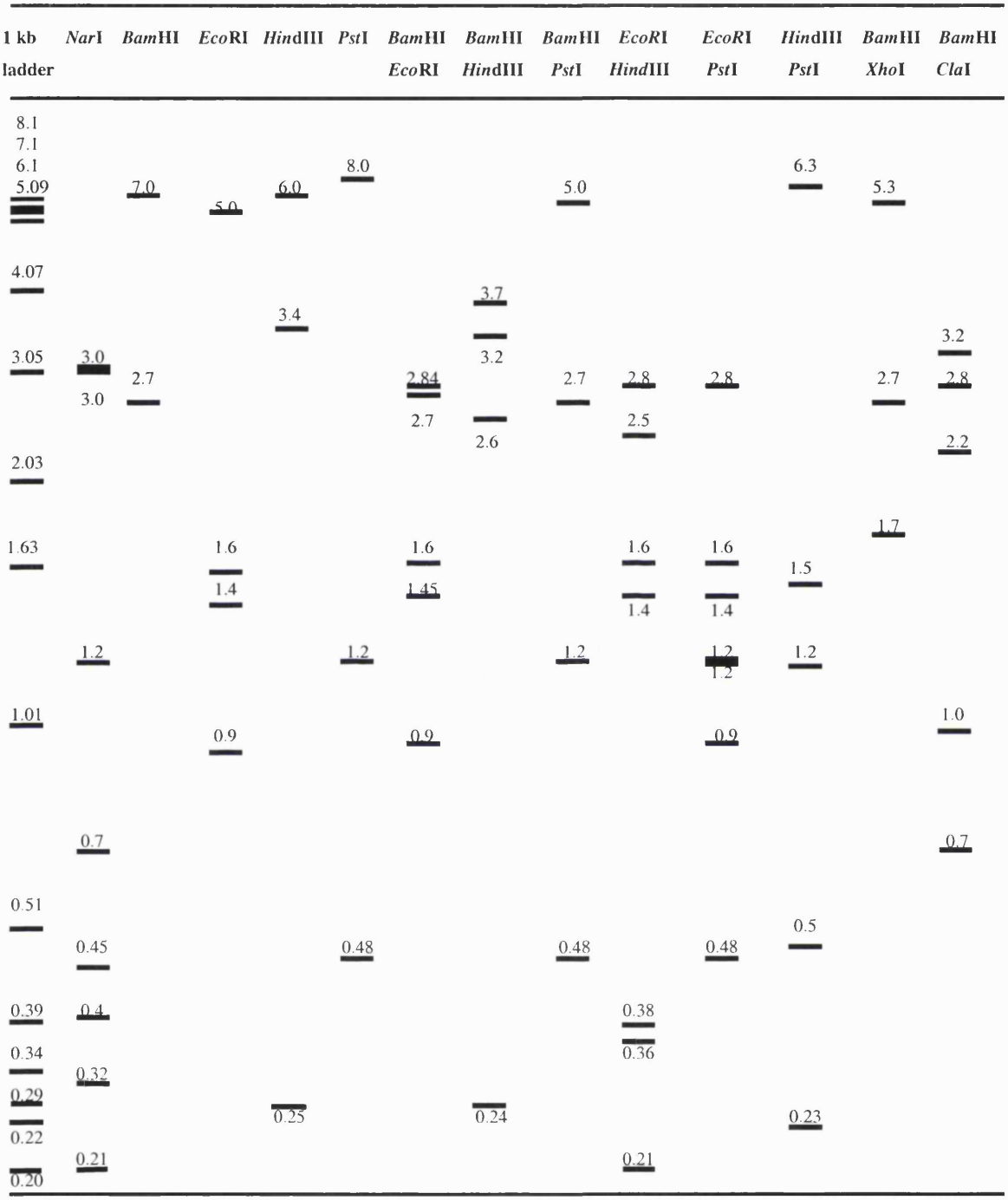


Figure 3.4 Graphic picture representing the agarose gel of p138 digested with different restriction endonucleases in Figure 3.3. Each small lines represent the DNA fragments of p138 digested with different restriction endonucleases, and the numbers represent approximate DNA fragment lengths in kilobases.

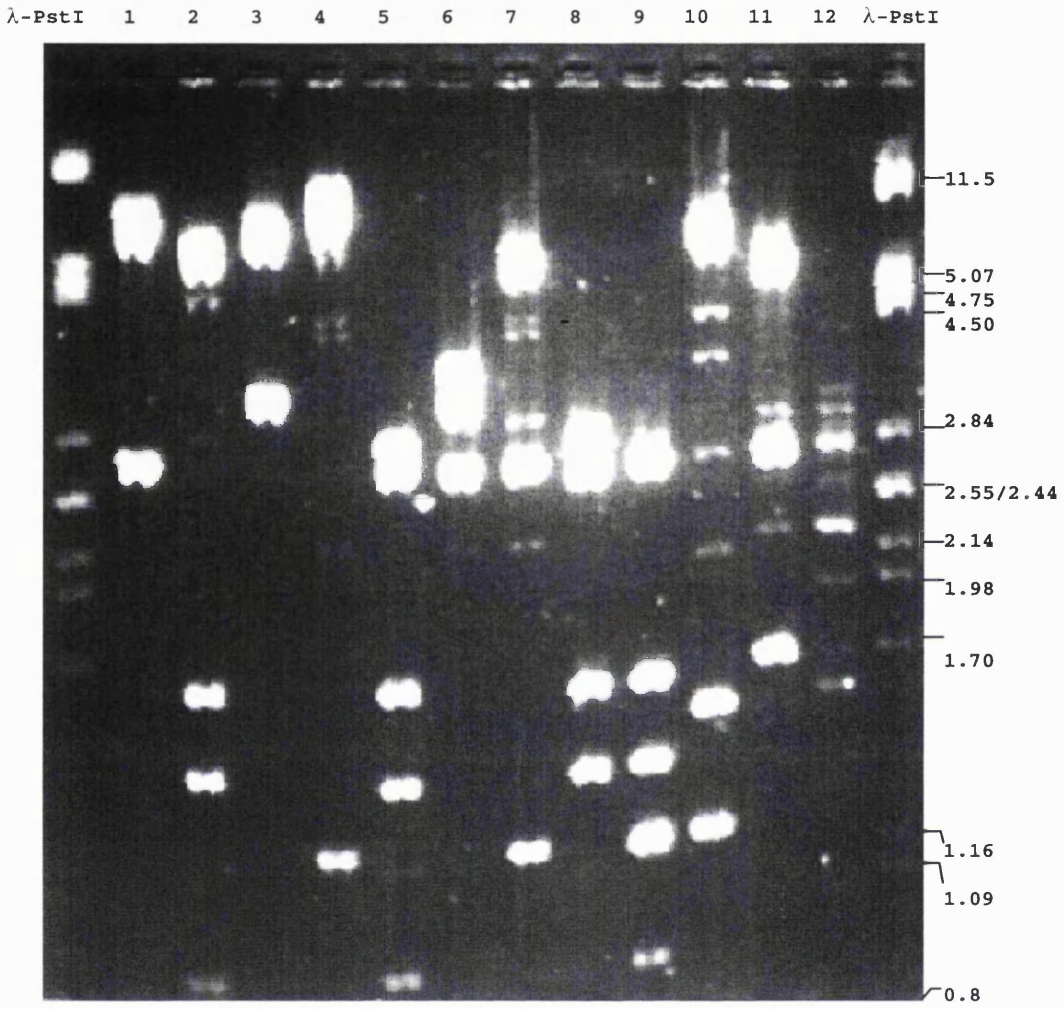


Figure 3.5 The agarose gel of p138 digested with different restriction endonucleases. From left to right, p138 was digested with 1: *Bam*HI; 2: *Eco*RI; 3: *Hind*III; 4: *Pst*I; 5: *Bam*HI-*Eco*RI; 6: *Bam*HI-*Hind*III; 7: *Bam*HI-*Pst*I; 8: *Eco*RI-*Hind*III; 9: *Eco*RI-*Pst*I; 10: *Hind*III-*Pst*I; 11: *Bam*HI-*Xho*I; 12: *Bam*HI-*Cla*I; and λ-*Pst*I: λ DNA digested with *Pst*I, as a marker.

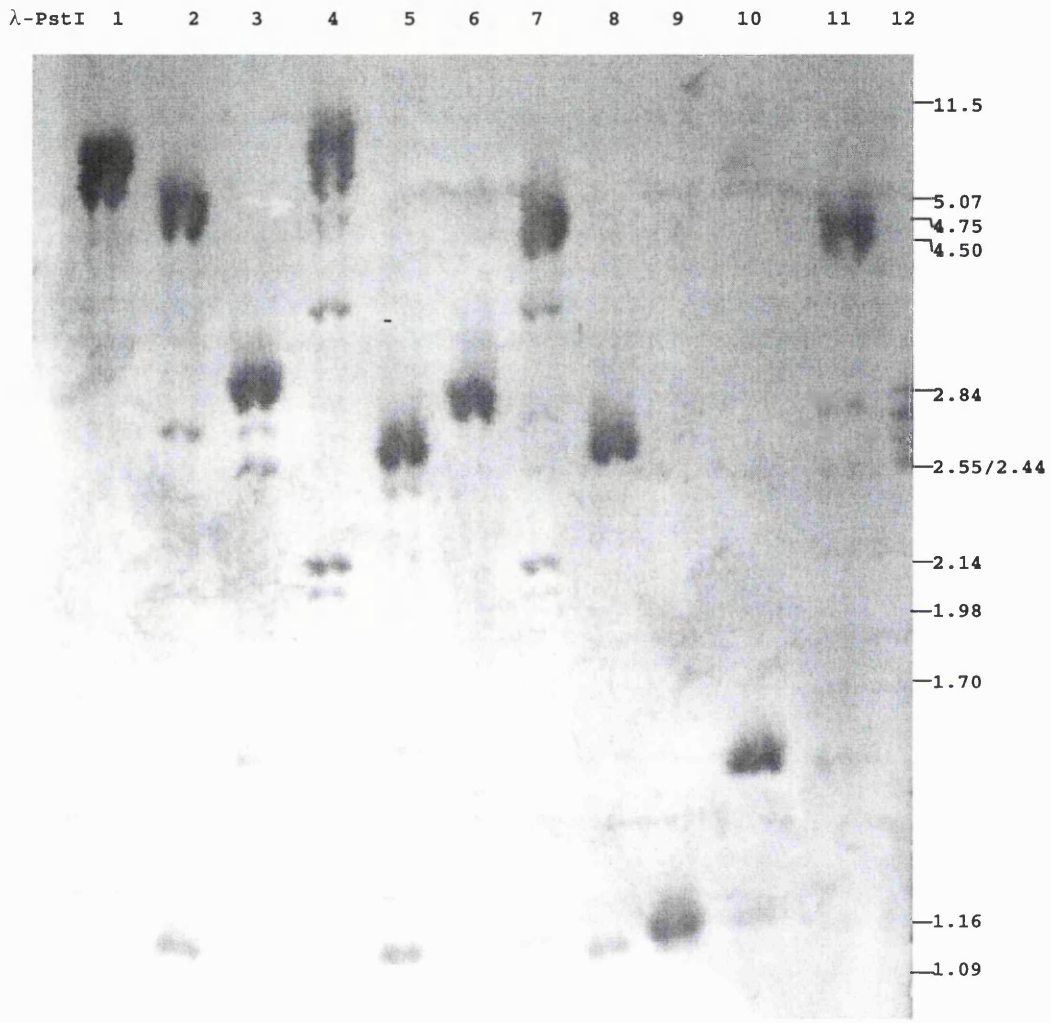


Figure 3.6 Southern hybridisation of gel in figure 3.4 with 222-bp fragment of pQR203 probe. From left to right, p138 digested with lane1: *Bam*HI; 2: *Eco*RI; 3: *Hin*dIII; 4: *Pst*I; 5: *Bam*HI-*Eco*RI; 6: *Bam*HI-*Hin*dIII; 7: *Bam*HI-*Pst*I; 8: *Eco*RI-*Hin*dIII; 9: *Eco*RI-*Pst*I; 10: *Hin*dIII-*Pst*I; 11: *Bam*HI-*Xho*I and 12: *Bam*HI-*Cla*I.

In Figure 3.6, the precipitation of colorimetric substrates NBT and BCIP resulted in purple or brown. The oligonucleotide probe of the 222 bp DNA fragment, bound to a 7 kb *BamH*I, 5 kb *EcoR*I, 3.4 kb *EcoR*I, 8 kb *Pst*I, 2.8 kb *BamH*I-*EcoR*I, 3.2 kb *BamH*I-*Hind*III, 5 kb *BamH*I-*Pst*I, 2.8 kb *EcoR*I-*Hind*III, 1.2 kb *EcoR*I-*Pst*I, 1.5 kb *Hind*III-*Pst*I and 5.3 kb *BamH*I-*Xho*I DNA fragment in previous agarose gel(Figure 3.5). However, the binding of the probe on p138 *BamH*I-*Cla*I digest (lane 12) was unclear.

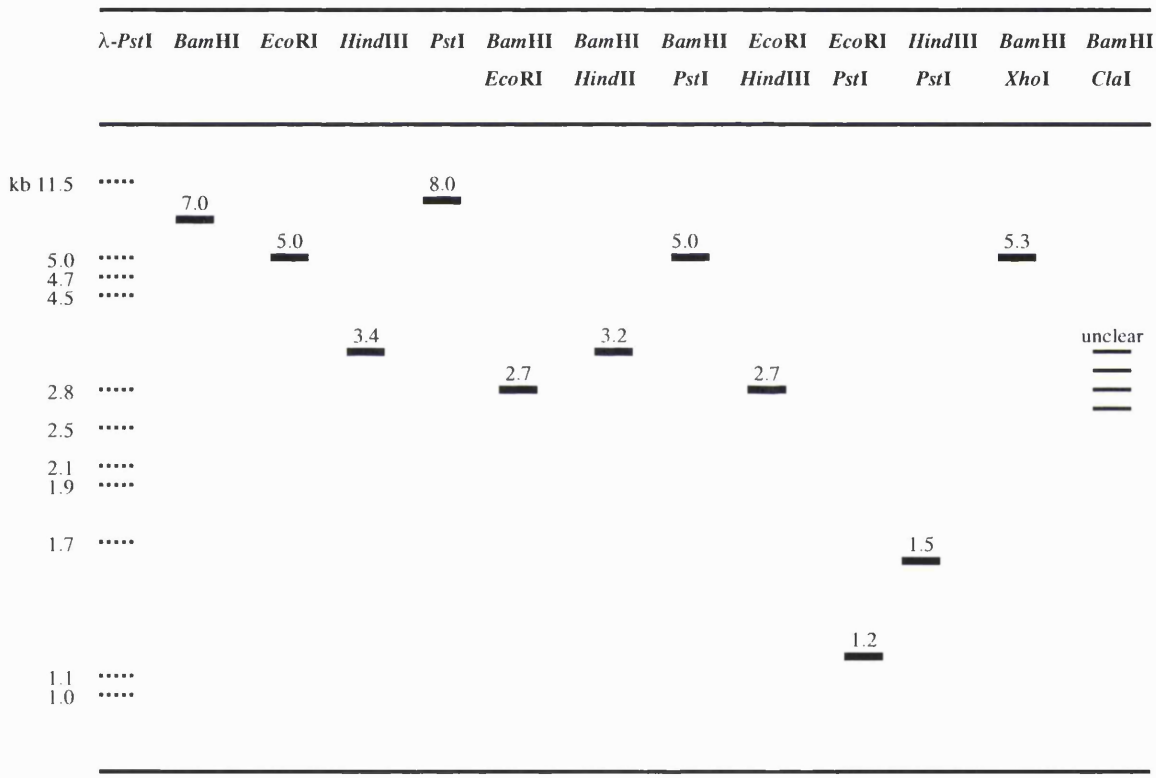


Figure 3.7 The graphic representation of the Southern hybridisation of p138 digests in Figure 3.6. The heavy bands show the fragments that bound with the pQR203 probe. The numbers represent approximate DNA fragment lengths in kilobases (kb). λ-PstI is λ-PstI DNA marker; and the others are p138 with single and double digests.

By combining the restriction results and southern hybridisation in Figures 5 and 6 together, a rough restriction map of p138 can be produced.

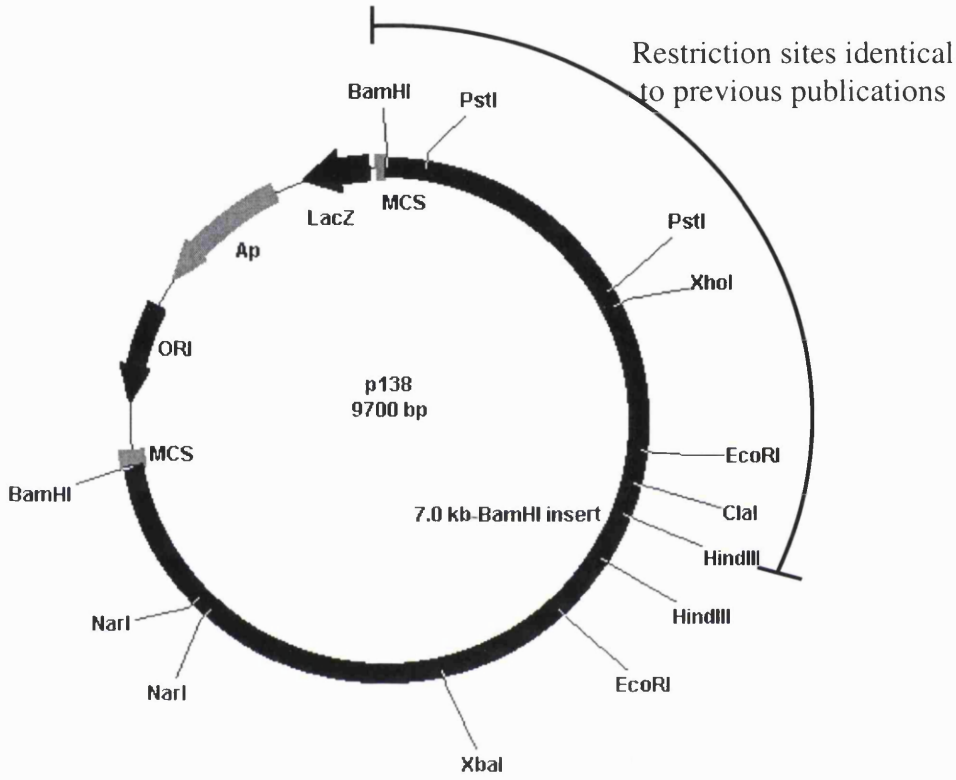


Figure 3.8 The preliminary restriction map of p138, harbouring the 7.5 kb *BamH*I insert in pUC19. Total length of the plasmid is approximately 9.7 kb and the 7.5 *Bam*HI insert is about 7.1 kb. The eight restriction sites on the insert are *Bam*HI, *Nar*I, *Xba*I, *Eco*RI, *Hin*dIII, *Cla*I, *Xho*I and *Pst*I.

We have able to show that the restriction map of about 3.0 kb nucleotides on the right hand side of the *Bam*HI insert (Figure 3.8) is in agreement with previous publications (Koga *et al.*, 1989 and Aramaki *et al.*, 1993). The restriction sites on the right-hand side of the *Bam*HI insert are *Bam*HI, *Pst*I, *Pst*I, *Xho*I, *Eco*RI *Cla*I and *Hin*dIII respectively. These restriction sites are identical in both restriction site order and sizes to

previous publications by Koga (1989) and Aramaki (1993). This result left the work in determination of the novel nucleotide sequence from the *Bam*HI insert at only about 4 kb. The recombinant plasmid (p138) carrying 7.0 kb *Bam*HI fragment was then renamed pQR416.

3.1.3 DNA sequencing of the *Bam*HI insert

Determination of the novel sequence on the *Bam*HI insert was carried out by both subcloning of the different fragments of the *Bam*HI insert and primer walking. There were 6 fragments: 1, 1.4 and 1.6 kb *Eco*RI-*Eco*RI, 1.4 kb *Nar*I-*Stu*I, 1.5 kb *Nar*I-*Xba*I, 1.8 kb *Xba*I-*Sac*I and 0.5 kb *Eco*RI-*Kpn*I, which were subcloned into pUC19 vectors. The recombinant plasmids obtained from these subclonings were sequenced by using both M1-21 forward and M13-20 reverse primer. Moreover, 10 primers were designed and used for DNA sequencing in ambiguous regions. These primers are E21: 5′-GCCCCGAGCCTGAAGGCCG-3′, E22: 5′-GCCAGCGTGGAATGCTCGC-3′, E23: 5′-GCCACCCGGCGCCACGCCG-3′, E24 5′-CCCACTCCGCCTGCAATCC-3′, E2E2: 5′-CCAGGGGCTGTTGTCGGCG-3′, E3E3: 5′-CAATTTGCGGCACTTGCCC-3′, SXSX: 5′-GCCGCTCTGGAGCTAAGGC-3′, XNXN: 5′-GAGAGGTTGGACCACGCCG-3′, primer1: 5′-CCCTTCGAGATCGCAATTGG-3′, and Primer2: 5′-GGCCTGTGCTGTGCTCCTC-3′. The sequence strategy in Figure 3.9 shows the DNA sequences obtained using these primers to determine the nucleotide sequence in ambiguous regions.

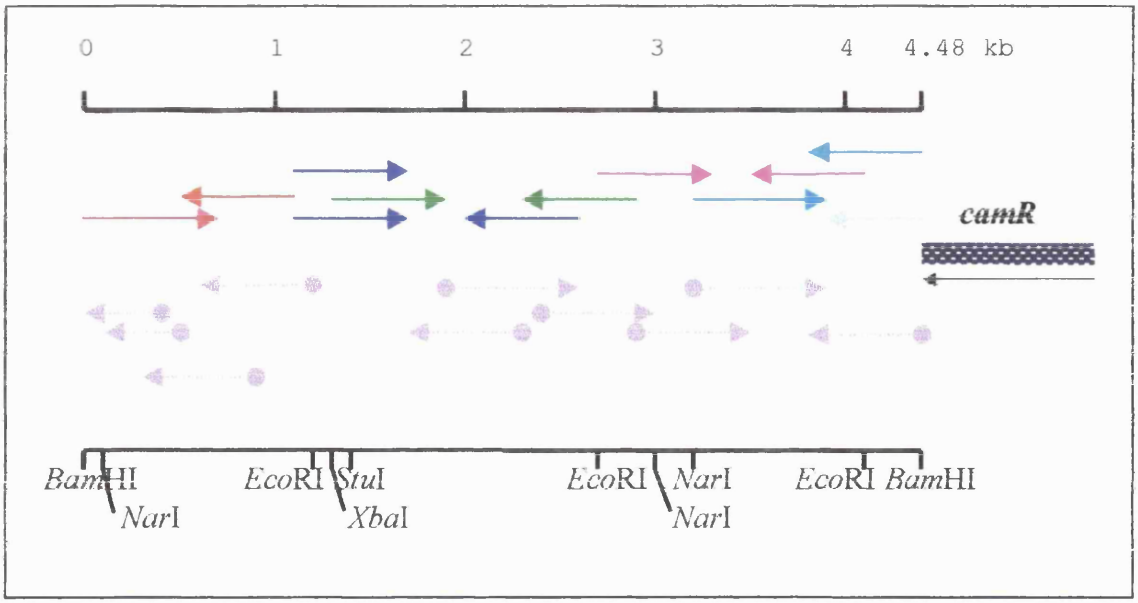


Figure 3.9 Sequencing strategy in the *Bam*HI insert of pQR416 (the *camR* and its direction of transcription were shown on the right hand side in the picture). DNA sequences were determined from the 5' end of the *camR* (→ represents the DNA sequences obtained from the subcloning strategy of various fragments and •--> represents the DNA sequence obtained from the primer walking strategy).

The orientation of the fragment from DNA sequence data was compared and aligned correctly to establish a DNA sequence that extended from the previous published sequence. Sequence data obtained from the sequencing were assembled by CAP (Contig Assembly Program) (www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl) and BLAST2 sequences (www.ncbi.nlm.nih.gov/blast/). CAP is a sequence assembly program which can assemble overlapping DNA fragments into one long contig by constructing multiple alignment of the overlapping DNA fragments and generating their consensus sequence (Huang, 1992). BLAST2 sequences is a BLAST program that can align two sequences (in this case nucleotide sequences) against each other. By using these two programs, the nucleotide sequence of the *Bam*HI insert was assembled into one contig of 4485 base pairs (see Figure 3.10).

1 - GGATCCCGTT GTCGCTGAAA AACTGATTCC CAAGGATCAT CCCTTCGGTG CTAAGCGCGT - 60

61 - GCCGATGGAA ACCAATTATT ACGAGACCTA CAACCGCGAT AACGTCCATC TGGTCGATAT - 120

121 - CCGTGAGGCA CCGATTCAGG AGGTCACGCC GGAAGGGATC AAAACGGCTG ACGCAGCCTA - 180

181 - CGATCTTGAT GTGATCATCT ATGCCACGGG CTTTGATGCG GTCACTGGTT CACTCGACCG - 240

241 - GATCGACATC AGGGGCAAGG ACAACGTCCG GCTGATCGAT GCCTGGGCTG AAGGCCCAAG - 300

301 - CACTTATCTC GGCCTTCAGG CTCGGGGCTT CCCGAACTTC TTCACCCTTG TCGGCCCGCA - 360

361 - CAACGGCTCG ACCTTTTGCA ACGTCGGTGT ATGTGGAGGA TTGCAGGCGG AGTGGGTGCT - 420

421 - CCGAATGATC TCCTACATGA AGGATAACGG TTTCACCTAT TCCGAACCGA CCCAAGCAGC - 480

481 - AGAGAACCGG TGGACCGAGG AAGTCTATGC CGACTTCTCC CGCACTCTGC TTGCAGAGGC - 540

541 - CAATGCCTGG TGGGTCAAGA CCACGACCAA ACCGGATGGC TCGGTCGTGC GCCGCACGCT - 600

601 - GGTGCATGTC AGTGGTGGAC CGGAATACCG CAAGCGCTGC GAGCAGGTCG CTTATAATAA - 660

661 - CTACAACGGA TTTGAACTCG CCTAATAACC AGAATTGGCT ACTTGCCTAG TGTGCGAACG - 720

721 - CACGTTCGCT CTCCGCGAAC GTGCAACCAA TAAGACCGAT AAGAGGACAC ACTATGAAAT - 780

781 - GCGGATTTTT CCATACCCCA TACAACTTGC CGACCCGTAC CGCTCGGCAG ATGTTCGACT - 840

841 - GGTCCCTCAA GCTGGCGCAG GTTTGTGACG AGGCCGGTTT CGCCGACTTC ATGATCGGCG - 900

901 - AGCATTCCAC GCTGGCCTGG GAAAATATCC CCTGCCCGGA AATCATCATC GGCGCCGCAG - 960

961 - CACCGCTGAC CAAGAACATC CGCTTTGCAC CGATGGCGCA TTTGCTGCCT TACCACAACC - 1020

1021 - CGGCTACCCT GGCGATCCAG ATCGGCTGGC TGTCGCAGAT TCTCGAAGGC CGCTACTTCC - 1080

1081 - TCGGCGTGGC GCCGGGTGGC CACCATACCG ATGCCATCCT GCATGGCTTC GAAGGCATTG - 1140

1141 - GCCCGCTACA GGAGCAGATG TTCGAATCCC TGGAGCTGAT GGAAAAAATC TGGGCCCGCG - 1200

1201 - AGCCCTTCAT GGAGAAAGGC AAGTTCTTCC AGGCTGGCTT CCCCGGCCCG GACACCATGC - 1260

1261 - CCGAGTACGA TGTGGAGATC GCCGACAACA GCCCCTGGGG CGGACGCGAG TCGATGGAAG - 1320

1321 - TCGCGGTCAC CGGCCTGACC AAGAATTCCT CGTCGCTGAA GTGGGCGGGT GAGCGCAACT - 1380

1381 - ACAGTCCGAT CTCCTTCTTC GGCGGTCACG AAGTCATGCG CTCGCATTAC GACACCTGGG - 1440

1441 - CGGCGGCTAT GCAGTCGAAA GGCTTCACTC CCGAGCGTTC CCGCTTCCGT GTCACCCGTG - 1500

1501 - ACATCTTCAT TGCCGACACC GATGCCGAAG CGAAGAAGCG TGCCAAGGCC AGTGGCCTGG - 1560

1561 - GGAAAAGTTG GGAGCACTAT CTGTTCCCGA TCTACAAGAA GTTCAATCTG TTCCCCGGCA - 1620

1621 - TCATCGCCGA TGCCGGCCTC GACATCGATC CGAGCCAGGT GGACATGGAT TTCCTCGCTG - 1680

1681 - AGCATGTCTG GCTTTGTGGC TCGCCGGAAA CGGTGAAAGG CAAGATCGAG CGCATGATGG - 1740

1741 - AGCGTAGCGG TGGCTGTGGG CAGATAGTCG TCTGCTCCCA CGACAATATC GACAACCCGG - 1800

1801 - AACCTTATTT CGAATCGCTA CAGCGCCTTG CCAGCGAAGT GTTACCGAAG GTTCGAATGG - 1860

1861 - GCTGAGGGAA CACCAATTCG GGAGAGGTTG GACCACGCCG CTATGCGGCG TGGCTCCATT - 1920

1921 - TCTCTCTGCA ATCGGGCATT GCGATCACCT CAGCCGAAAC TGGTGGCAGT GGCCTGTGCT - 1980

1981 - GTGCTCCTCG TCGCAATGCT GTGCGGTTCT GACATATAGC ATGTTCACCA GCATCGCAGA - 2040

2041 - GAAACCGTCC ATGAGCGTCT CTACTCCTGC GTAGCCCTCC CCTCGAACCA ACGCACCGCC - 2100

2101 - CTATACCGCA CGGAAGGCAC ATTCCCGTGC CAGCTGAGGA TCAGCGGCCA CAGCACTAGC - 2160

2161 - TGACAGAAAT CTGCCATCAA TCGCCGGATC ACCGATCCAA TTGCGTATCT CGAAGGGGAC - 2220

2221 - GCAAGACAAG GACGGCACCA TTACCAGCCG CCGATCTGCA TAAGCATCCG CTATGCTGCA - 2280

2281 - CGCATGATCG GCATCGGTCG CACCAATGCC CATCAACTGA TTTCAAACCG AAGAAGTCGA - 2340

2341 - GTCCAAATCG ACGCGCATGA GCACCTCCAG CCCCCACGAT CTCATCAGGC AGCAGCGTGG - 2400

2401 - ATCGGCATTA GCTCGTCCTC TCATGATAAA ATCCTGCCTG GGACTATCTG GCTCCGGATC - 2460

2461 - ATCCCGACGA CTCTACAACA TTAGACGCAG CACACTATTA TCATGGGCGC ATCAACTCTC - 2520

2521 - GGGAGGTTTG GAGTCGTTTT TCTGCGGGTA TTGCAGGCCT TGGCATTTGG CGTCATTCCG - 2580

2581 - TTCTCAGGCA AAGAAATTTA TGACCAATAC TGAAAGTACA AGGAAAGTAA AGTCTATCAA - 2640

2641 - GGCTGATGTT GATGCCATGA AGCGCCAAAG CGTTCTAGAT GAATCCATCC AGCAATTCTT - 2700

2701 - TGATAACGGC TACGAGGCAA CATCCCTGGA ATCTATCGCA GATGCGCTGG GCGTCACCAA - 2760

2761 - GCAATTCATC TATTCACGCT TTAACAGCAA ATCCGAAATA CTGGTTTCGA TTTGCCGCTC - 2820

2821 - TGGAGCTAAG GCCGCAGAAA AAGCTGTTGA ACTGAGCGAG GAGATAGAAG GCAATGCGGC - 2880

2881 - TGTGCGACTG GCTTGTATCC TGCGTTTTTT TGTTCAATTG CAAATCGAAC ACCGCAGGGA - 2940

2941 - AGTTGCTATC TATTTTCGTG AATTCAAGAA TTTACCCGCC GACGAAGCCC ATGCAATTGA - 3000

3001 - TGCGTCCAAA CTGCGCTTTC ACCGGATGCT GTGCGCCGTC TTGAACGAGG GCAAGGCAGC - 3060

3061 - GGGACTGTTC GAATTCGATG ACACCTCTCT CGCTGCATCG GCCTTGGGCG GCATGGTTTC - 3120

3121 - TTGGCCATTT TTCTGGTTCC AGCCGGAAGG GCGGTGGGTT CCGACCCTGG TAGCCCATCA - 3180

3181 - ATTTGCGGCA CTTGCCCTCA AGACAGTTGG AGTTTCAGAC CCGTCGATTG TTGCCGCTGG - 3240

3241 - CTGAAAACTG CCCCCGTAAT CGAGCTTGAA GTAGACTCCG CCACAAGCTA GGAGAGCATC - 3300

3301 - CACCCCCCTG CAATCAACCA TATTGTTGAC TTTCTCACCC GGCATGCTGA CTATAGCCTA - 3360

3361 - CTGCCCCCTA CATGCTCACT TCATCACAGC GAACAGTTAA ACCAGACGGG GTTGATACAG - 3420

3421 - CCTGCCTTGA GCAAAAGCTT CAGGCGAAAA CAGTTCGTTA CAGAGAATCA CTCATGCGCA - 3480

3481 - AGTTCAGATC CTTCGCCTTC CAGCTGACGC TGGTAACAGT CACTGTGGGC TGCGGCATGA - 3540

3541 - ACACCATACC TGCAATCGCT GAGCCTGCCG GCAGGCAACA ACATCAAGTG CCCGGATTTT - 3600

3601 - ACCGCATGAA CCTGGGTGAG TTTGAAATCA CGGCGCTCTA TGACGGTTTT ATCAAGCTTG - 3660

3661 - ATCCGGCATG GCTCAGCGGC ATCAGTGCCG ACAACATTCA GAGCCTGCTG GCAAAAATGT - 3720

3721 - TCATCGATTC GAGCAAGGGC ATTCAAACCG CAGTGAACGG CTACCTGATC AATACCGGCG - 3780

```
3781 - AACACCTGGT GCTGGTCGAT GCAGGCTCAG CTCAGTGCTT CGGTTCGACG TTGGGGGTGA - 3840
3841 - TGCGCCGCAA CCTAGAAGCG TCCGGTTATC AGGTAGAACA AGTGGATAGC GTGCTGCTGA - 3900
3901 - CCCACCTGCA TCCGGATCAT GCCTGCGGTC TGGCCAATGC CGATGGCACG CCAACCTACC - 3960
3961 - CGAATGCCAG GGTCTACGTG CCGCGCCAGG AAGCCGAATT CTGGCTGGAT CAGGATATCG - 4020
4021 - CCGCCATGCC GGAACCCAGC CAAGCGTTTT TCCTGATGGC CAGGGCAGCA GTCGCACCCT - 4080
4081 - ATGCGCAGGG GCGCCTGCTG CGCTATGAGC CTGACGCCGC ATTGCTGCCG GGCGTGGAAA - 4140
4141 - GCGTGCCTAC CTACGGGCAT ACACCTGGTC ACTCAGCCTA CCTCTTTACC TCTGGCGATG - 4200
4201 - AACGCCTGAT GGTCTGGGGG GATCTGGTGC ATAACCATGC CATTCAGTTC GCCCGGCCGG - 4260
4261 - AAGTGGTTAT CGAGTTTGAC GCAGACTCTG CGCAAGCCAG GAGCTCACGG CAGAGCATGC - 4320
4321 - TGACGAATGC TGCAAAGGAG CATTTTTGGG TAGCGGGTGC ACATCTACCC TTTCCCGGGC - 4380
4381 - TTGGCCGTGT TCGCGCGACG GATGGCGCCT ACGCCTGGGT ACCCATTGAG TTTGGCCCAG - 4440
4441 - TTGGAGACCA CCCCTGAAAC AAACTCGGCA CGAACCCGCC TACTC                - 4485
```

Stop codon *camR* ←

Figure 3.10 The novel nucleotide sequence in the *Bam*HI insert extending from *camR* on the CAM plasmid. The stop codon (TAG) of *camR* is shown in bold .

3.1.4 Location of the pQR203 probe on the *Bam*HI insert

The specific region where the pQR203 probe bound on the *Bam*HI insert, was determined by using the Blast 2 sequences program. This alignment program is able to produce local alignment between two overlapping sequences. The parameters for the Blast 2 sequences were set as default. Figure 3.11 shows alignment between the pQR203 probe sequence and the 4485 bp DNA sequence. The 100% matched region between the pQR203 probe and the 4485 bp DNA sequence is at the nucleotide position 4018 to 4229 on the 4485 bp DNA sequence.

```
                                                  camR  -------------→  stop
BamHI 4553 caaaaagccaaactcgacattactaccttggtcgcgctcttcaatatcgaggaggagtag 4492
BamHI 4493 gcgggttcgtgccgagtttgtttcaggctgggtggtctccaactgggccaaactcaatgg 4432

           KpnI                                          SmaI
BamHI 4433 tacccaggcgtaggcgccatccgtcgcgcgaacacggccaagcccgggaaagggtagatg 4372

                                                 stop <-----ORF4   SacI
BamHI 4373 tgcacccgctacccaaaaatgctcctttgcagcattcgtcagcatgctctgccgtgagct 4312
BamHI 4313 cctggcttgcgcagagtctgcgtcaaactcgataaccacttccggccgggcgaactgaat 4253

                                     pQR203 primer
Probe                              1 ccagaccatcaggcgttcatcgccagaggtaaagag   36
                                     ||||||||||||||||||||||||||||||||||||
BamHI 4253 ggcatggttatgcaccagatcccccccagaccatcaggcgttcatcgccagaggtaaagag 4192

Probe   37 gtaggctgagtgaccaggtgtatgcccgtaggtaggcacgctttccacgcccggcagcaa   96
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
BamHI 4193 gtaggctgagtgaccaggtgtatgcccgtaggtaggcacgctttccacgcccggcagcaa 4132

Probe   97 tgcggcgtcaggctcatagcgcagcaggcgcccctgcgcatagggtgcgactgctgccct  156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
BamHI 4133 tgcggcgtcaggctcatagcgcagcaggcgcccctgcgcatagggtgcgactgctgccct 4072
                                                              pQR203 primer
Probe  157 ggccatcaggaaaaacgcttggctgggttccggcatggcggcgatatcctgatccagcca  216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
BamHI 4073 ggccatcaggaaaaacgcttggctgggttccggcatggcggcgatatcctgatccagcca 4012
                                                     EcoRV
Probe  217 gaattc 222
           ||||||
BamHI 4013 gaattc 4018
           EcoRI
```

Figure 3.11 The nucleotide sequence alignment between the pQR203 probe and 4485 bp-DNA sequence. The alignment also shows continuity of the novel nucleotide sequence from the end of the camR gene. The nucleotide sequence of the probe is in bold; restriction sites of *Eco*RI, *Eco*RV, *Sac*I, *Kpn*I and *Sma*I are underlined. The primer sequences used in amplification of the 222 bp-pQR203 probes are also indicated in bold italic (*Bam*HI: the 4485 bp *Bam*HI DNA sequence [complementary strand]; and Probe: 222 bp DNA fragment probe from pQR203).

3.2 Analysis of open reading frames (ORFs) on the 4485 bp DNA sequence

Two different approaches were used to determine potential open reading frames and coding regions on the 4485-bp DNA sequence. The first approach is the FramePlot2.3.2 program developed by Ishikawa and Hotta (1999). FramePlot is a sequence analysis tool that is able to predict coding regions and open reading frames in bacteria with a high G+C content. This program is considered useful in the analysis of coding regions in a high G+C content genome DNA such as *Streptomyces*, which prefers G or C (around 92%) in the third codon of its genetic codes (Nakamura *et al.*, 1997). FramePlot is able to calculate the G+C content in the third position of open reading frames and plot them parallel to the result of open reading frame analysis. The plot of G+C content in the third position of the genetic codes in each predicted ORF can be useful in distinguishing coding regions from non-coding regions in genome DNA. The second approach is the Open Reading Frame Finder program (ORF finder) from the National Centre for Biotechnology Information (NCBI). The ORF finder is an accessible web site based graphic analysis tool to identify all possible open reading frames from different DNA sources. The results from the ORF finder are given in an accurate start and stop codon of the possible gene and its translated amino acids.

In an analysis of ORFs by FramePlot, all parameters were set as default. Window size was 40 codons, which allowed the FramePlot to calculate the third-letter G+C content within a group of 40 codons. Step size was also set as default or 5 codons, which controlled the window size to move along the sequence by a set of 5 codons. The start codon to calculate the start site for ORFs was set as ATG, and the minimum of amino acids in an ORF was given as 20 amino acids.

The FramePlot of the 4485-bp DNA sequence is shown in Figure 3.12. In Figure 3.12, four potential open reading frames (*orf1*, *orf2*, *orf3* and *orf4*) were identified. All ORFs contain biased G+C content contributed at the third position in their genetic codes

compared to an average G+C content of 4485 bp DNA sequence (56.3%). This value is lower than the G+C content for *P. putida* genomic DNA (60.7-62.5%) (Mandel, 1966).

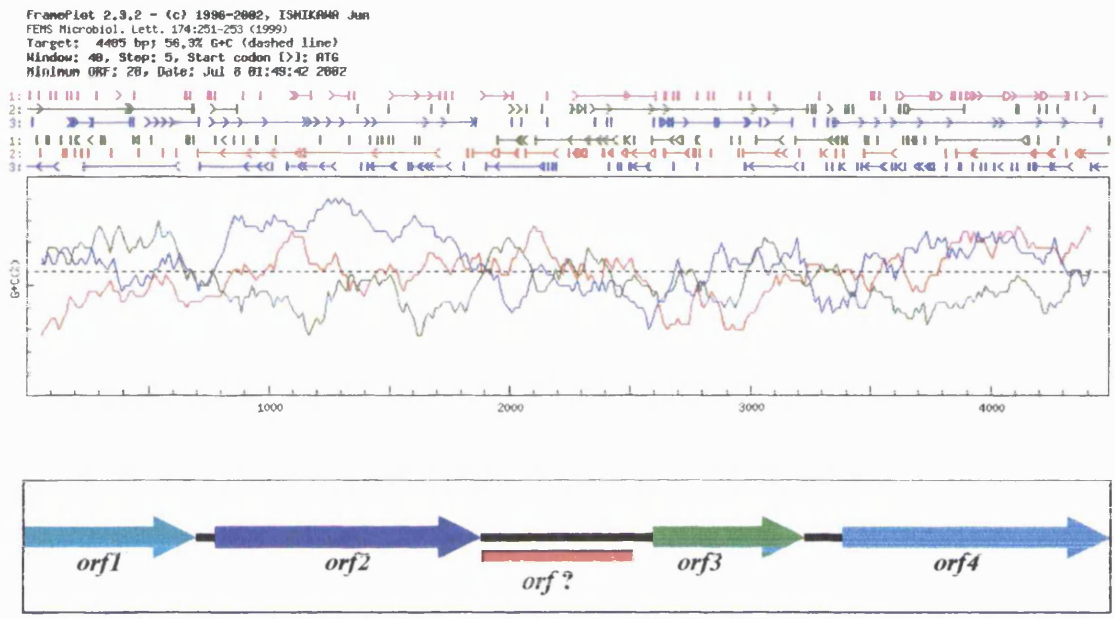
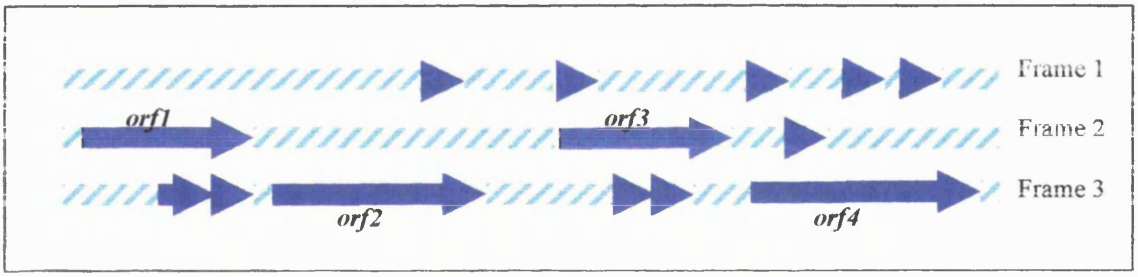


Figure 3.12 FramePlot analysis of the 4485 bp DNA sequence. Predicted open reading frames are shown in six frame translations (> indicates a start site [ATG codon] and ▮ indicates a stop site [TGA, TAG or TAA codon]). The G+C content of 4485 bp DNA sequence (56.3%) is represented as a dashed line. The coloured lines are the G+C content in the third position of the genetic codes of predicted open reading frames. The box below summarises the most likely open reading frames from this FramePlot.

To check the result from the first approach (FramePlot) in prediction of ORFs, an ORF finder was employed for reassurance. The result from the analysis of the 4485-bp DNA sequence by the ORF finder is shown in Figure 3.13. Four open reading frames were identified as potential coding regions. All four possible ORFs have a transcription direction from the left to right-hand side. The first ORF (*orf1*) is a DNA fragment length of 685 base pairs, starting at the nucleotide position 1 to 685 on the forward strand of

frame three. This *orf1* encodes for 225 amino acid residues. *orf2* is the longest ORF, which is a DNA fragment length of 1092 bp starting at the nucleotide position 774 to 1865 on the forward strand of frame three. The *orf2* is located downstream of *orf1*, and encodes a 363 amino acid residue protein. The third ORF is *orf3*, which is a DNA fragment length of 888 bp starting from the nucleotide position 2357 to 3244 on the forward strand of frame two. This *orf3* encodes a 295 amino acid residue protein and is located 492 nucleotides downstream of *orf2*. The last potential ORF is *orf4* located at the end of the 4485 bp DNA sequence. The *orf4* starts at the nucleotide position 3372 to 4457 on the forward strand of frame three. It is a DNA fragment length of 1086 bp, and encodes a protein of 361 amino acid residues (see Figure 3.13).

Forward strand



Reverse strand

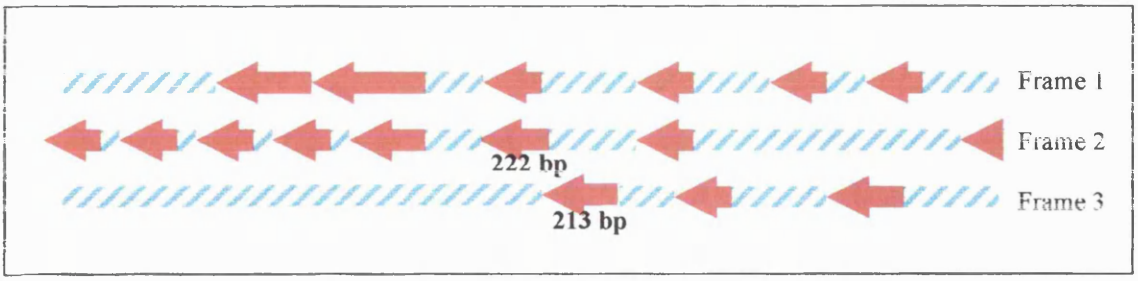


Figure 3.13 ORF finder analysis of 4485 bp DNA sequence. All potential ORFs are on the forward strand; *orf1* and *orf3* are in frame two and *orf2* and *orf4* are in frame three.

It is possible that the intercistronic region between *orf2* and *orf3* may contain a region coding for a small protein, which has to be considered. Both FramePlot and ORF finder analysis showed that there is a big gap between *orf2* and *orf3*. This intercistronic region is 492 base pairs in length, which could code a protein of about 163 amino acids. As a result of FramePlot and ORF finder analysis, it is found that two possible open reading frames are in this intercistronic region. Both open reading frames are transcribed divergently from *orf1-orf2-orf3-orf4* on the opposite strand of the 4485 bp DNA sequence. They are 222 bp and 213 bp region (see also Figure 3.13).

The first potential open reading frame is in the reverse strand of frame two, containing 222 nucleotide sequences or 73 amino acid residues. In bacteria, genes that encode for small proteins of around 70 amino acids. These small proteins are such as RNA-binding proteins and lipoproteins. The nucleotide sequence of the 222 bp DNA sequence and its deduced amino acid are shown in Figure 3.14.

```
                    <---- Start codon of orf3
1861 - TTCCTTGTACTTTCAGTATTGGTCATAAATTTCTTTGCCTGAGAACGGAATGACGCCAAA - 1920
1921 - TGCCAAGGCCTGCAATACCCGCAGAAAAACGACTCCAAACCTCCCGAGAGTTGATGCGCC - 1980
1981 - CATGATAATAGTGTGCTGCGTCTAATGTTGTAGAGTCGTCGGGATGATCCGGAGCCAGAT - 2040
2041 - AGTCCCAGGCAGGATTTTATCATGAGAGGACGAGCTAATGCCGATCCACGCTGCTGCCTG - 2100
2101 - ATGAGATCGTGGGGGCTGGAGGTGCTCATGCGCGTCGATTTGGACTCGACTTCTTCGGTT - 2160
2161 - TGAAATCAGTTGATGGGCATTGGTGCGACCGATGCCGATCATGCGTGCAGCATAGCGGAT - 2220
2221 - GCTTATGCAGATCGGCGGCTGGTAATGGTGCCGTCCTTGTCTTGCGTCCCCTTCGAGATA - 2280
2281 - CGCAATTGGATCGGTGATCCGGCGATTGATGGCAGATTTCTGTCAGCTAGTGCTGTGGCC - 2340

                                Start codon of 222 bp region--->
2341 - GCTGATCCTCAGCTGGCACGGGAATGTGCCTTCCGTGCGGTATAGGGCGGTGCGTTGGTT - 2400
                              M  C  L  P  C  G  I  G  R  C  V  G  S

2401 - CGAGGGGAGGGCTACGCAGGAGTAGAGACGCTCATGGACGGTTTCTCTGCGATGCTGGTG - 2460
     -   R  G  G  L  R  R  S  R  D  A  H  G  R  F  L  C  D  A  G  E

2461 - AACATGCTATATGTCAGAACCGCACAGCATTGCGACGAGGAGCACAGCACAGGCCACTGC - 2520
     -   H  A  I  C  Q  N  R  T  A  L  R  R  G  A  Q  H  R  P  L  P

2521 - CACCAGTTTCGGCTGAGGTGATCGCAATGCCCGATTGCAGAGAGAAATGGAGCCACGCCG - 2580
     -   P  V  S  A  E  V  I  A  M  P  D  C  R  E  K  W  S  H  A  A

2581 - CATAGCGGCGTGGTCCAACCTCTCCCGAATTGGTGTTCCCTCAGCCCATTCGAACCTTCG - 2640
     Stop codon of 222 bp region                  Stop codon of orf2 ←
```

Figure 3.14 The nucleotide sequence of the possible ORF of the 222 bp sequence between *orf2* and *orf3*. Its deduced amino acids are shown below the nucleotide sequence. The start and stop codon for the 222 bp DNA sequence and the stop codon of *orf2* and *orf3* are shown in bold.

A search of the protein data bank for the proteins that might be related to the deduced amino acid of the 222 bp DNA sequence found that this amino acid sequence is similar to a putative inositol polyphosphate 5-phosphatase [*Arabidopsis thaliana*] (GenBank accession number AC006533) and putative inositol 1,4,5-trisphosphate 5-phosphatase [*Arabidopsis thaliana*] GenBank accession number AC007153). The alignment of the deduced amino acid of 222-bp DNA sequence with the putative inositol polyphosphate 5-phosphatase showed that both sequences had 45 amino acids overlapping with 31% identity. The total length of the putative inositol polyphosphate 5-phosphatase and inositol 1,4,5-trisphosphate 5-phosphatase are 1144 and 1136 amino acids respectively.

The alignment of the deduced amino acid of 222 bp DNA sequence with the putative inositol 1,4,5-trisphosphate 5-phosphatase showed 45 amino acids overlapping between two sequences with 31% identity.The alignment of deduced amino acid of 222 bp DNA with the putative inositol 1,4,5-trisphosphate 5-phosphatase and putative inositol polyphosphate 5-phosphatase in *A. thaliana* is shown in Figure 3.15. This multiple alignment showed that the deduced amino acid of 222 bp region was partially overlapped with the two putative proteins. This suggested that the 222 bp DNA sequence is unlikely to encode for an active protein.

```
IPPP            --VPCGFGRAIGNKGGVGLRIRVYDRIMCFVNCHLAAHLEAVTRRNA------------- 45
ITPP            --VPCGFGRAIGNKGGVGLRIRVYDRIMCFVNCHLAAHLEAVNRRNA------------- 45
73-AA           MCLPCGIGRCVGSRGGLRRSRDAHGRFLCDAGEHAICQNRTALRRGAQHRPLPPVSAEVI 60
                  :***:**.:*.:**:     .:.*::* .. *  .: .:. **.*

IPPP            -------------
ITPP            -------------
73-AA           AMPDCREKWSHAA 73
```

Figure 3.15 Multiple alignment of the deduced amino acid of the 222 bp DNA sequence with the related protein (73-AA: deduced amino acid of 222 bp DNA sequence; ITPP: putative inositol 1,4,5-triphosphate 5-phosphate [*A. thaliana*]; and IPPP: putative inositol polyphosphate 5-phosphatase [*A. thaliana*]).

The second possible open reading frame has a DNA fragment length of 213 bp encoding 70 amino acid residues. Similar to the 222 bp DNA sequence, the 213 bp DNA sequence is on the opposite strand and translated divergently from *orf1234*. However, a BLAST query using the deduced amino acid of the 213 bp DNA sequence showed no similarity matches to known proteins in the GenBank protein database.

3.3 Southern hybridisation, DNA sequencing and ORF anlysis of *Kpn*I fragment

3.3.1 Southern hybridisation: identification of a *Kpn*I fragment using pQR277 as a probe

The data of the protein sequence of the N-terminal of 3,6-diketocamphane 1,6-monooxygenase (information supplied by Prof. J. Littlechild, University of Exeter) was used to design PCR primers based on these amino acid sequences and used to amplify a 0.7 kb fragment from *P. putida* NCIMB 10007 genomic DNA. This 0.7 kb DNA fragment was cloned in pCR 2.1-TOPO to form pQR277 (by Dr. John Ward). The fragment was used to hybridize to genomic DNA digests and recombinant DNA from transformants to detect the desired fragment. Preliminary results of the identification of desired fragments from bulk digestion of *P. putida* NCIMB 10007 genomic DNA showed that the 0.7 kb probe hybridised to a 3.5 kb *Bam*HI, 3.5 kb *Pst*I, 1.8 kb *Sac*I, 4 kb *Kpn*I and 8 kb *Eco*RI fragment from the genomic DNA (information from Dr. John Ward). DNA from preparative gels corresponding to the 3.5 kb *Bam*HI, 3.5 kb *Pst*I, 1.8 kb *Sac*I, 4 kb *Kpn*I and 8 kb *Eco*RI fragment were ligated into pUC18, to yield different recombinant plasmids. Transformation, screening, southern blotting and detection on the recombinant plasmids were performed. About 1,000 recombinant plasmids were screened, and one recombinant plasmid containing the desired fragment was identified as p4CK-27, pUC18 with 4 kb-*Kpn*I insert.

Figure 3.16 shows the gel electrophoresis of the p4CK-27 plasmid with the other recombinant plasmids digested with *Kpn*I. This gel was also subjected to further investigation in a southern hybridisation with the pQR277 probe. The result showed the probe bound specifically to the 4-kb *Kpn*I DNA fragment in p4CK-27. The southern hybridisation of the p4CK21-30 series with the pQR277 probe is shown in Figure 3.17. The binding of the pQR203 probe to the 4-kb DNA fragment in p4CK-27 is quite

specific. Within an hour, the precipitation of colorimetric substrate NBT/BCIP occurred strongly on the site of the 4 kb DNA fragment of p4CK-27-*Kpn*I.

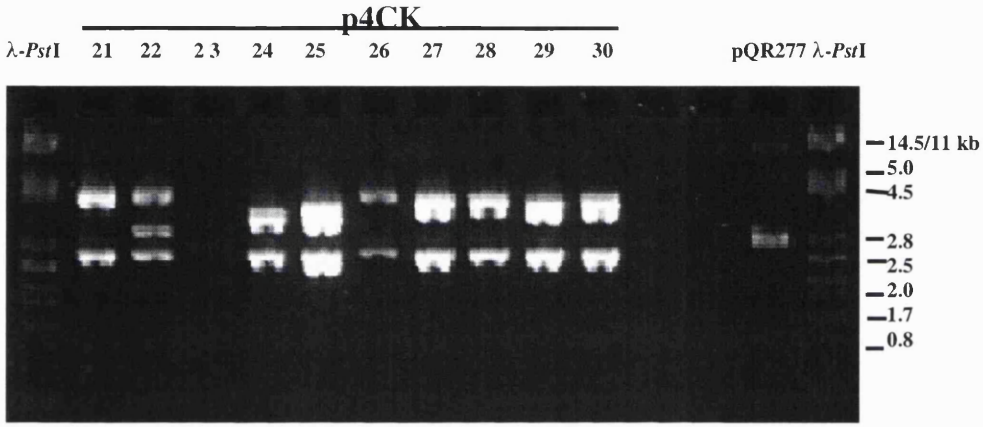


Figure 3.16 Agarose gel of p4CK-27 to p4CK-30 digested with *Kpn*I. pQR277 is the plasmid of pQR277 probe (λ-*Pst*I is used as a DNA marker). The DNA fragment lengths, in kilobases, are indicated on the right of the gel.

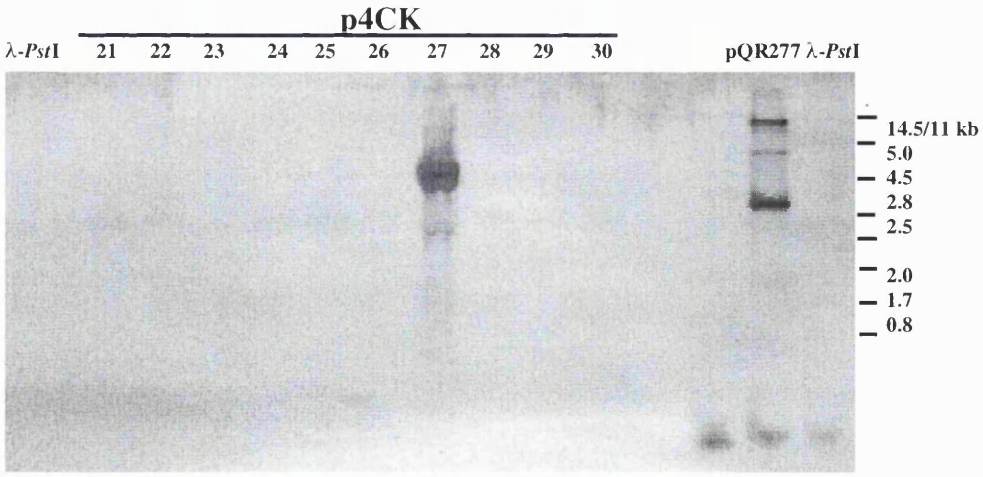


Figure 3.17 The southern hybridisation of the gel electrophoresis in Figure 3.16 by pQR277 probe. The binding of the pQR277 probe was on a 4 kb DNA fragment in p4CK-27 *Kpn*I and on the probe itself (in the pQR277 lane).

3.3.2 Restriction map of the 4 kb *Kpn*I DNA fragment

By employing different restriction endonucleases to digest the 4 kb *Kpn*I DNA fragment, the restriction map for the 4 kb *Kpn*I DNA fragment was established and is shown in Figure 3.18. The knowledge from this preliminary result of the restriction map of the 4 kb *Kpn*I DNA fragment was also used to select smaller DNA fragments to subclone into pUC18 vectors for further convenient sequencing. To put the recombinant plasmid of p4CK-27 (pUC18 with 4 kb *Kpn*I fragment) in our records, this recombinant plasmid was renamed pQR417.

3.3.3 DNA sequencing of the 4 kb *Kpn*I DNA fragment

The DNA fragments of 3.8 kb *Bam*HI, 0.6 kb *Eco*RI, 1.5 kb *Eco*RI, 1.5 kb *Eco*RI and 3.0 kb *Sac*I of the 4 kb DNA fragment were subcloned into pUC18 (see Figure 3.18). These recombinant plasmids were then sequenced using conventional M13-20 forward primer or M13-21 reverse primer.

The subcloning strategy helped to scale down DNA fragments to around the sequencing range (~500-600 bp for good sequence signals in automated DNA sequencing). In the first round of sequencing, DNA sequence data was obtained using conventional M13-21 and M13-20 primers, in some regions; however, the data were ambiguous. This problem can be eliminated by designing new primers based on the first round sequence data obtained.

Figure 3.19 shows the sequencing strategy used to determine the complete nucleotide sequence of the 4 kb *Kpn*I insert. Sixteen and ten DNA contigs were obtained from M13 primers and newly designed primers respectively.

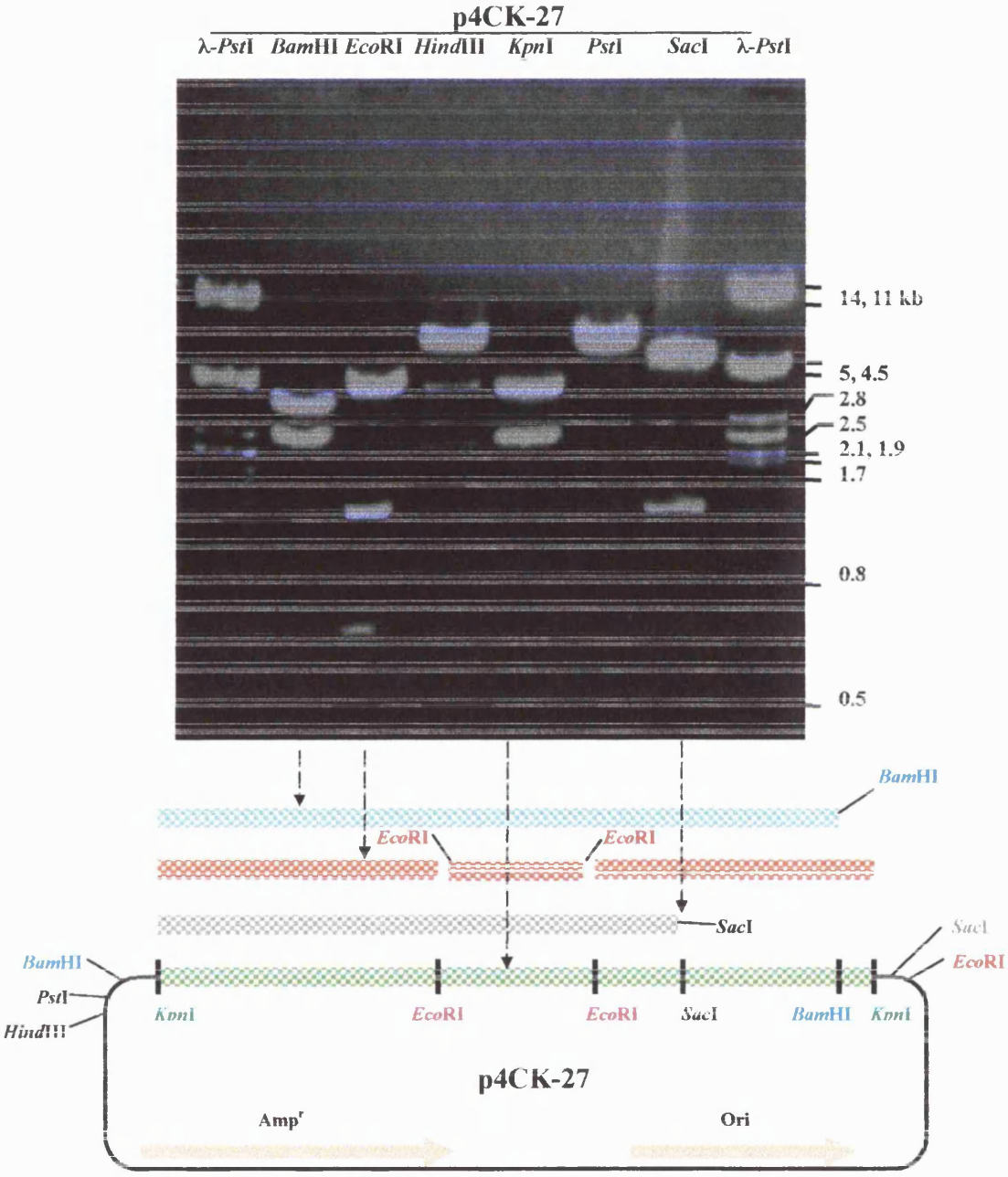


Figure 3.18 Agarose gel of the p4CK-27 digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I and *Sac*I. The digest shows the *Hin*dIII and *Pst*I sites on the 4-kb *Kpn*I DNA insert. From this gel, a restriction map of the *Kpn*I insert based on these digestions was determined (shown below the gel). THe restriction sites for *Hin*dIII, *Pst*I, *Bam*HI, *Sac*I and *Eco*RI on the pUC18-cloning vector are shown.

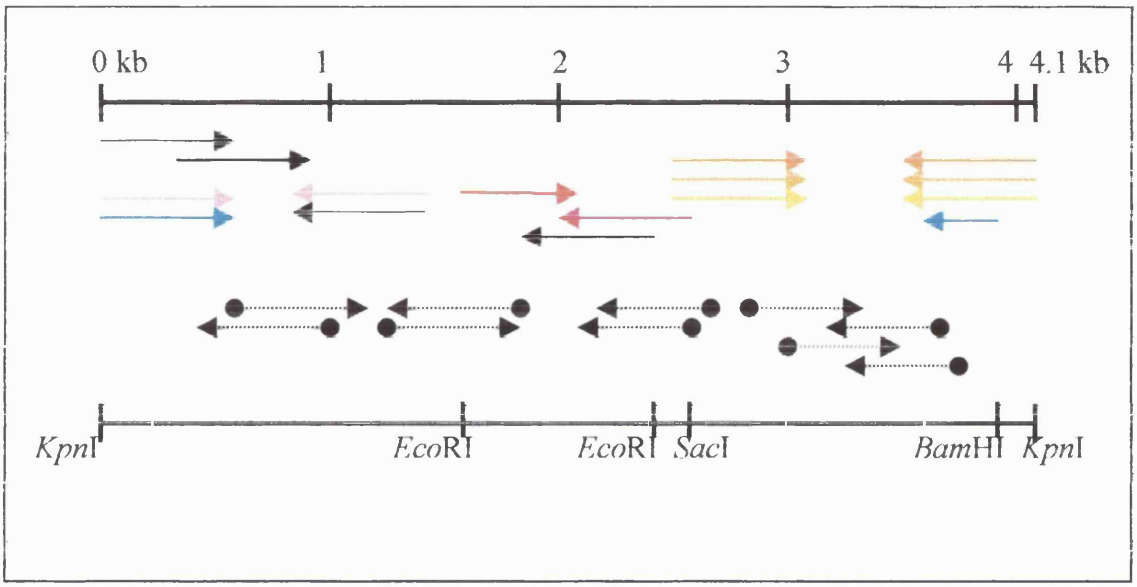


Figure 3.19 The sequencing strategy used to determine the complete nucleotide sequence of the 4 kb *Kpn*I insert of p4CK-27. Arrows show the direction of the DNA sequences obtained from conventional primers (M13-21 forward and M13-20 reverse primers) from the subcloning strategy. The lines beginning with a dot-arrow are the DNA sequences obtained from designed primers.

All DNA sequence data were also assembled by using CAP and BLAST2 sequences into one contig of the *Kpn*I insert. The DNA sequence length of the *Kpn*I insert was longer than expected. The nucleotide sequence of the 4 kb *Kpn*I insert is shown in Figure 3.20. This novel nucleotide sequence is 4,201 base pairs in length.

1 - GGTACCGCAG GCGTTCGCCC ATGAGCTGGC CTTCTCGAAG AAAGAAAACA TCAAGGTCGA - 60

61 - AGTACCTGGG GACGCCACCA CCTGGTGCAA ACCCGAGGTC GAGCTGACCA TCACCCGTCC - 120

121 - GGCCTGGGAC AACCAGGAGC TGCTATCGGG CTTGCTGACC AAGCTGCCGT TCGTGTTCGC - 180

181 - CAAGGACTGC TCGACCGCCA AGGTGAGCTG GAAAGCCGTG GACGCCAAAG GCAACCTCTA - 240

241 - CGCCAGCGGT TCGGGCAATG CCAGCAACCT GGGCCTGGTG ACCCTGGCCG CCGCGCCTGC - 300

301 - CGCTGTCGCT GCTGTTCCGG CACCCGCTCC GTCTGCGCCA GCCCCGGCAC CTGCACCGGC - 360

361 - AGAAGCTCCG GCGGCTGTTG CTCCAGCCCC AGCGCCTGAG CCAGCTCCTG CGGTAGTCGA - 420

421 - ATCGACACCT GCGAAAGTAG AAGCAGCCCC TGCTGCAGTG CCAGCTCCAG CCGTTGCCGC - 480

481 - AGAGCCCGCA CCAGCACCAG CAGCCGAAAC CCCCGCCGCA GCGCCTGCCG CTCCGCCCGT - 540

541 - CGCAGCCCCT GCCGCCGCAG TAGCCGCAGC CCCCACCTCG GATTTCGGTC GCGCCGTGGT - 600

601 - CCTGGAAAAC CGCAACCTGA TGCAGGTAAC GGACGGTTCT GGCTGCAAAT GGGTACTCAG - 660

661 - CCGCGAGATC ATCAGCAACG GTGACACCCT GTCGTTCGGC ACCACCCCGG CCATGCCGTG - 720

721 - CCCGGCGTCC GGTTTTGGCG AAGGCAACTT CGAAAAAATC AGCTGGAAGG CCGTGGGCAC - 780

781 - CTACCGTGGG GATAACTGGA GCCGCGTATA CGTGCACCCG AGCGGCCTGA TTTTCAACAA - 840

841 - GGTCTACGAG CCTGCGGTCA AGGACAAGGC CGTTTCCTAC CTCACGGCAG ACGCTGGCCA - 900

901 - GGCCACATTC CTGGTGGGCG AGATCCCCAG CCGGCAGATG AAGGTGTACC TGGCCTTCAC - 960

961 - CCGTGGCAGC TACGGCGTGC TCCGTCCGTT CGGCAGCGAC CCTTACTACG TGGCAGTAAC - 1020

1021 - CCCGGACGAG TCGTTCGCTC TGGACGCAGC CAAGTACAAG GAAGCTGCGC TGGAAATCTT - 1080

1081 - TGACCTGATC AAGACCACCT CGCCGACCAC CACCGACGTG GCCGACCTGC TCATCGTCAA - 1140

1141 - AGACATTTCG GCGATCACGA ACAACATGTG GGGCAACGAC GCGCAGAAGA TCACCCGCAA - 1200

1201 - CCGCATCGGC ATCAATCGCC AAGGCCTGTT CTTCGATGTG CGCGATGGCG CGAACTGGGG - 1260

1261 - CGGTTCAAGC GTGAGGAGCA GCGCGTGCGT GAGGCGCGCC GGCGTCAACA GGAACTGGGC - 1320

1321 - GGTTCAGCGT GAGGAGCAGC GCGTGCGTGA GGCGCGCCGG CGTCAACAGG AACTGGCCAG - 1380

1381 - CGTGCACACC CGCGTACTGG AGCGTTACCA GCAGTTGCAA GACGGCATGA GCGAGTTCAA - 1440

1441 - AGGGCGCGAA ACCGAAGCCC TGGCGCAAAT GGCCGGCATC AAGGTGCGCT TCGCTTCGCC - 1500

1501 - GCTGGAACAG CAAAACCCGG CAACTTCGGC CCGCGTGGTG CCGATGATGG TCCACGTCAC - 1560

1561 - CGGCAAGCAG GGCGATTTCT ACGCCATCGA CTTCCCGAGC AAGGGCCGCC TGGTGGCGGA - 1620

1621 - CGAGGAGTAC AGCGAAGGCT GGTACGTGGC GCAAGTGGCC AACGCCACGC CGTACTACCC - 1680

1681 - GCTGGACGAT GGCCGCGCGG TGCCCACATA CCGTGCCTAT AACGCGGGCG AGCCCCAGGC - 1740

1741 - ATGCAAACAG GACAAGTGTG CGGACCTCGT GTCGTTCGGC GCCGTGCTGG CCAAGGAATT - 1800

1801 - CCCTAACGCC GGTATCGATT TCAGCTGGAC CCCCGAGGTC TCCCAAAAGT ACGTGAACGA - 1860

1861 - CTGGAACAAC GCTTCCGCCA TGGTCCAGTG AGGCAAACAC CATGATCAAA CAGATGAACA - 1920

1921 - AGAAGCAGCT GCTGATTGGC GGTGTGGCCG CAGCGGTGTT GGTGCTCGGC GGTGGTTACG - 1980

1981 - CGACTGCTGT CCAGCTCGGG CAAGAAGGTT GCGATGGCCT ACTACGACGA CTTCAAGGAG - 2040

2041 - CGCTACTTCC TTGAGGACGT GCTGTCTGAA GGTGATATTT CCTACTCGGC GTTCTCTGGC - 2100

2101 - AACCTGACTG TCGCCGACCC CGAGATCCGC GTGGCGGCCG CGCAGACCAA CGGTGCCCAG - 2160

2161 - CAATTCATGC GTGGCCTGTC GGGCCTGATG GAGTCGGTTC GCGGTAGCTC GGCGGAGGAA - 2220

2221 - GGCCTGGCCG GTTGGGCCAA GTACCAACTC AACGCCAGCG GCAATGTTGC AGGCGTCTAC - 2280

2281 - CTGAAGGCCG ATGCGCTGAA GCTGGCGCAC GATGGTGACA GCAAGGACGG CACGATCCAC - 2340

2341 - ATCCAGCTGC TTGGGATGCA GATGGGCAAC CCTTTCGTTG CCAGCAAGGG CGCTGAAGTG - 2400

2401 - GTGCTGGTGT CTGATGTCAG CGACGAGATC CAGCCGCGTG CCGAACTGGC CGCCAATGGC - 2460

2461 - CGGGCCACCC AGTCGAACTT CGACTGGGGT AGTAACATGG TTGTGCGTCA GCCAATGACC - 2520

2521 - GGCGCGTTTT TGGTCAGTAG CACGGGGGAA TTCGGCACCA CGGTAGACGT GGACTTCACC - 2580

2581 - CTCAAGCGCT CCAGTGACGG CGAGGGCTCG ATGGCGTTCG TCGTGACCCA CCGTAACGAT - 2640

2641 - GGCTCCAAGG TGGGTGAGAT CGTGCGCAAG GCCGAGTTCC AGTCGCTGCC GGAGCTCGAC - 2700

2701 - GATGTCGAAA CCCAACTCAA GAGTGCGTTG AGCGCCATGC TGATCGGGGC CTACAGCACG - 2760

2761 - TACACCGGTC AGGCTGTACT GGCCGAGGCG GTGAGTGGTT TTGCACGCAA GGCCAAGGTG - 2820

2821 - GAGAACTACA GCGTGAGCTA CAGCGGCTTC AAACCGCTGA AAGAGGCTTA CAGCGACTAC - 2880

2881 - CAGCAGAATG TTCCGAAGGC GAAGTTTGCC GCCTTCTGTG AACAGGTCGG GTTGTCGATG - 2940

2941 - TGGAGCAGTG ATTTCGGTGC CAAGGGCAAG AACCACAGCG ACTCGGAATG TGCCATCGGG - 3000

3001 - CAGAAGCTGG TCGAGGATGG CAAGTTTGAA GAGCACTACA CCTTCAAGGA AGGCAAGAGC - 3060

3061 - CTGTTTGCGG CGTTGTTCGT GAGCAAGGCG TATGAGCTGG AAACCAATTG ACCGGTTGAT - 3120

3121 - CGCTGATAAA CAAAGCCCCC GGCAACGATG AGTTGCCGGG GGCTTTTTCA ATTCTGCGAA - 3180

3181 - TCAGAGGTTT CAGCGCATCG CCGCCAGCTT GATGCCGAAC CCTACCAGGC AAGCCCCGGC - 3240

3241 - CAGCCGCTCG AACATATTGG TAATGCGCGG GTTGGCGCGC ATGCGTTCGG CCAGCCGGTG - 3300

3301 - GGTCAGCACC ACGGCGATCA ACCCGTACAG GAACGTGATC ACCGCCACCG TTGCCGCCAT - 3360

3361 - GAAGCCGAAC GTCACCAACC CCTGATGCTT CACCGGGTCG ATGAATAATG GGAAGAACGC - 3420

3421 - CATGTAGAAC ATGATCGCCT TGGGGTTGAG CAGGGTGATC AGCATGGTCT GGCGCAGGTA - 3480

3481 - TTGGCCGTTG TCCATGCGGC TGGTGCGCGG CGCGTCACCC GGTTTGCTCA ACAGCATGCG - 3540

3541 - CAGGCCCAGG TAGGCGAGGT ACGCGGCGCC GGCCCATTGC ACGATGTGGA AGGCCGCAGG - 3600

3601 - GTAGGTGGCC AGCAAGGTGG CGACGCCAGC TACCGCTAGC CACAACAGGA CTTGGTCACC - 3660

3661 - GACGATCACC CCACAGGTCG CGGCCAGGCC TGCCTTGATG CCGCCCTTGC CGGTGGCAGT - 3720

3721 - GATCAGGGCA AAATTGCCCG GGCCCGGGAT GGCCAGAAGA ATGATGAAGG CGATGACGAA - 3780

```
3781 - TGCGCCGTAG TCGGTGACGC CGAGCATGGT GAACTCCTGT AGAGCAATGC AAAGGCAGAG - 3840
3841 - TCATGACTAT CGTCGCCAAC AGCGTGCTTT ACAACCTGTT GTTGCACATT CGCAATCCAT - 3900
3901 - CATTGCCAAG CACGACTTGG ATCCAATGCC TGTTTTCTGT ATCATCCCGC CGCCATCAGA - 3960
3961 - CGATGGTCGA AGATGATTCA TTTAAAGGAC TTTGTAATGA AAAAGCTGTT CAAGGCCACC - 4020
4021 - GTAGCCGTTG CTGTAGTTTC GGGTGTTGCC CTGCTGTCGG GTTGCACTGG CCAGGTTTAC - 4080
4081 - AACCAGCCGA AAAACTGCAC CTACGACTAC CTGTTCCACC CTTCGGTTTC CATCTCCAAG - 4140
4141 - ATCATCGGTG GCTGCGGCCC GATCGATAAA CTGCCTCAGC AGCAGTAATC TTGGCGGTAC - 4200
4201 - C
```

Figure 3.20 Nucleotide sequence of the 4 kb *Kpn*I insert in p4CK-27.

3.3.4 Binding of the pQR277 probe to the *Kpn*I fragment

The region of the pQR277 probe, which is homologous to the 4.2 kb DNA sequence, was determined by using the LALIGN program. This alignment program can find the best local alignment between two sequences by comparing the nucleotide sequence of each DNA fragment point by point. As shown in Figure 3.21, it reveals the alignment between the pQR277 probe and the 4201 bp *Kpn*I DNA. The best matched region between the pQR277 probe and the 4201 bp DNA sequence is at the nucleotide position 73 to 518 with 52.4% identity. This result is unlike the homology of the pQR203 probe to the 4485 bp DNA sequence in the previous experiment, as that was 100% match, and this will be discussed later.

```
         50        60         70        80        90       100
pQR277 GCCGGCAGCTGGC-CGAAGAAGACCTGCCGCTGAGCCGCAGCGAGCGTATGGAATTGCAG
       :::  :: ::::  : ::   ::  :   ::::: :  :: :   :::  ::  : : :
4201   GCCACCACCTGGTGCAAACCCGAGGTCGAGCTGACCATCACC---CGTCCGGCCTGGGAC
            80        90       100       110          120       130

         110       120        130       140       150           160
pQR277  AACCTGCTGAACAGCAA-CGGGCATGAGGCCGGAAATGCTGATGGCATTATC----GGCG
        :::: :  :: : :: : ::::: ::  : :  :  ::: : : :  ::  :    : :
4201    AACCAG--GAGCTGCTATCGGGCTTGCTGACCAAGCTGCCGTTCGTGTTCGCCAAGGACT
                 140       150       160       170       180

               170       180       190       200              210
pQR277  CCAATACCCGCAAGGCCATTCGCAATGCCCAGCAGGGGCTGGGGTG-------GCCGGCG
         :   :::  :::::  :   : :: :::  : : :     ::           ::: :::
4201    GCTCGACCGCCAAGGTGAGCTGGAAAGCCGTGGACGCCAAAGGCAACCTCTACGCCAGCG
       190       200       210       220       230       240

           220         230       240       250         260       270
pQR277  GATGGGT--ATCCGACCCACAAGCTGCTGGAAAGCCTG--CGCCAACAGTGAGATTGCCG
        : : ::   :: : : : ::  :   ::::  : ::::  ::::     ::    :: ::
4201    GTTCGGGCAATGCCAGCAACCTGGGCCTGGTGACCCTGGCCGCCGCGCCTGCCGCTGTCG
       250       260       270       280       290       300

               280       290        300       310       320
pQR277  GGGGGCGTTGCGCCCCCCGG-CAATCTCAAGCCTTGAACAGGTCCTGTTCCAGCACCAAT
          :   ::: :: : ::::  :  :::   :::     : :  ::::  :: :::  :
4201    CTGCT-GTTCCGGCACCCGCTCCGTCTGC-GCCAGCCCCGGCACCTGCACCGGCAGAAGC
       310        320       330        340       350       360

      330       340       350          360          370
pQR277  ACCTGCCGATCACTGTCCAGCCTCA---CCCGAGCCC---CCAGCGGCAAGC------AC
         :: :: : :     ::::::: ::   :: :::::    :: :::: :  :      ::
4201     TCCGGCGGCTGTTGCTCCAGCCCCAGCGCCTGAGCCAGCTCCTGCGGTAGTCGAATCGAC
         370       380       390       400       410       420

        380       390         400       410        420         430
pQR277  ACGTTCGGGTCACAATGC--CCACTGCGCCACCCCGCCAGCAC-AG--GTATCCCCAGAG
        :: : ::       : ::  :: ::::  ::    :::::: : ::  ::  :: :::::
4201    ACCTGCGAAAGTAGAAGCAGCCCCTGCTGCAG--TGCCAGCTCCAGCCGTTGCCGCAGAG
         430       440       450         460       470       480

              440       450       460       470
pQR277  GC-GCAAAGGTTTCCACCAACAACGGTGCAAGCGCCGCCG
         : :::   :    ::::: :: : :   :: : ::::::
4201    CCCGCACCAG----CACCAGCAGCCGA--AACCCCCGCCG
           490          500         510
```

Figure 3.21 Binding of the pQR277 probe to the 4201 bp DNA sequence from LALIGN analysis (: shows the identical nucleotides between two sequences; and gaps (-) are introduced into the sequences to obtain the best match between the two sequences). pQR277 is the 0.7 kb DNA fragment probe from pQR277, and 4201 is the 4201 bp DNA sequence.

3.4 Analysis of open reading frames on the 4201 bp DNA sequence

Pseudomonas are among bacterial species such as *Streptomyces* which have biased G+C content at the third position of each codon. This biased G+C content in the third position of *Pseudomonas* open reading frames enables the prediction of coding regions on *Pseudomonas* genes. By calculating the G+C content of the third position of 4201 bp *Kpn*I fragment, we can predict the potential open reading frames of coding regions.

Figure 3.22 is the result from the FramePlot analysis of the DNA sequence of the *Kpn*I fragment. The overall G+C content of the 4201-bp sequence is 61.3% (dashed line). The regions with the G+C content at the third position between 90-95% were predicted as coding regions and designated *orf5*, *orf6*, *orf7* and *orf8*. These four open reading frames are complete genes, and both start and stop codons were found on the genes.

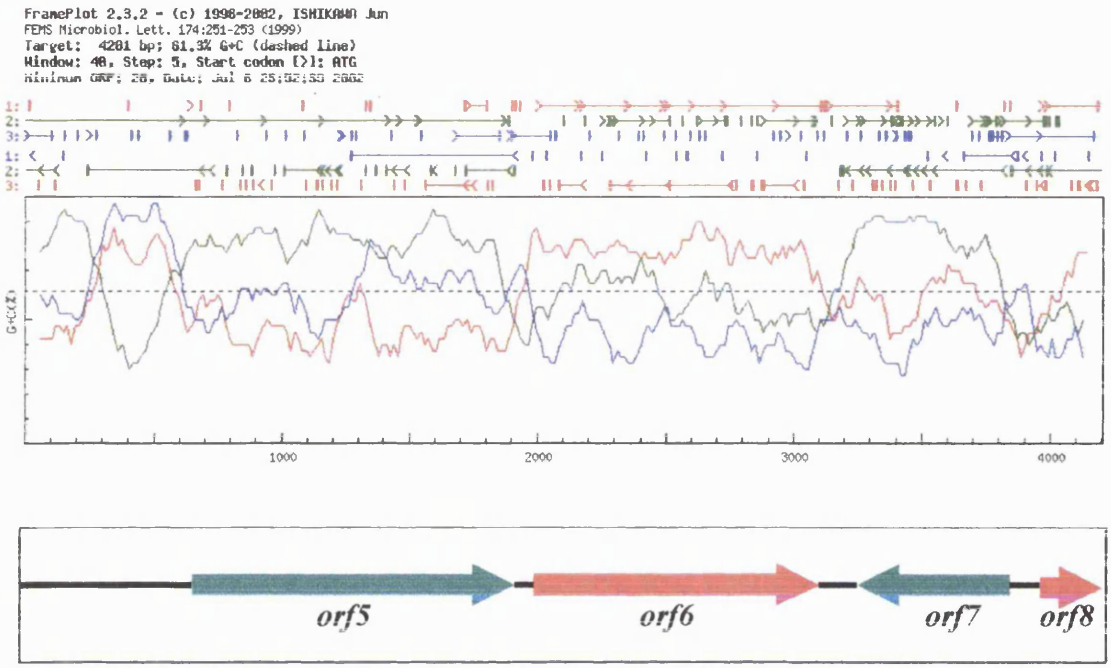
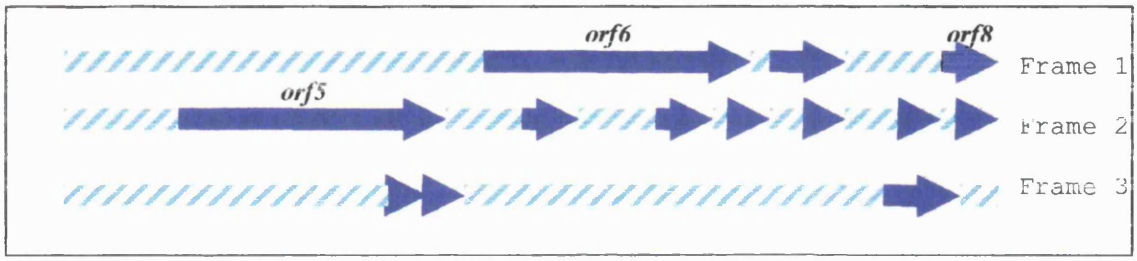


Figure 3.22 FramePlot analysis of the 4201 bp DNA sequence. Predicted open reading frames are shown in six frames (> indicates a start codon [ATG codon] and ▮ indicates a stop codon [TGA, TAG or TAA]). In the middle graph, the dashed line represents the average G+C content of the 4201 bp DNA sequence. The coloured lines are the G+C content in the third position of the genetic codes of predicted open reading frames. The box below summarises four potential ORFs from the FramePlot analysis.

The 4201 bp DNA sequence was, additionally, reassessed for ORFs in the ORF finder analysis. The analysis of the open reading frame by the ORF finder is shown in Figure 3.23. The analysis revealed four open reading frames (*orf5*, *orf6*, *orf7* and *orf8*) in the 4201-bp DNA sequence, confirm the result of the FramePlot analysis. From the ORF finder, the *orf5* region starts at the nucleotide sequence position 620 to 1891 and is 1272 nucleotides in length. The *orf6* region starts at the nucleotide position 2014 to 3111 and is 1098 nucleotides in length. The *orf7* region starts at the nucleotide position 3807 to 3190 and is 618 nucleotides in length (on the reverse strand transcribed in the opposite

direction with *orf5*, *orf6* and *orf8*). The *orf8* region starts at the nucleotide position 3973 to 4188 or 216 basepairs transcribed in the same direction as *orf5* and *orf6*. It is clear that only four potential open reading frames (*orf5*, *orf6*, *orf7* and *orf8*) are in this 4201 bp DNA sequence.

Forward strand



Reverse strand

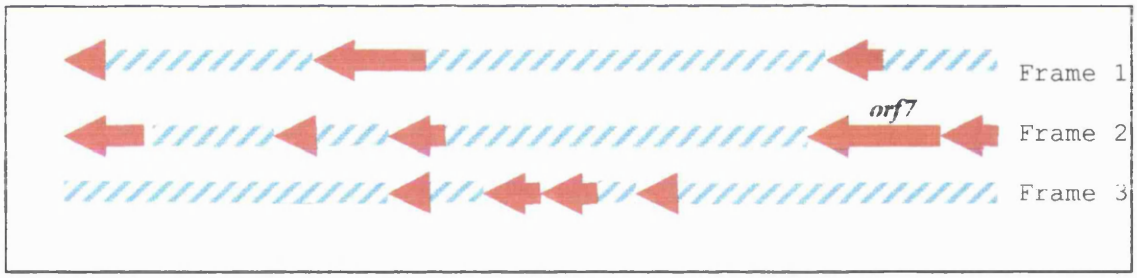


Figure 3.23 ORF finder analysis of the 4201 bp DNA sequence. Four potential open reading frames are indicated as *orf5*, *orf6*, *orf7* and *orf8*.

3.5 Summary

In a Southern hybridisation, using the pQR203 probe (a known sequence on the *cam* operon) to identify a DNA fragment further on the left-hand side of the *cam* operon, an *Bam*HI fragment of about 7.0 kb was identified. Restriction mapping of this *Bam*HI fragment verified that this fragment was located downstream of *camR*. The nucleotide sequence of 222 basepairs on the right end of the *Bam*HI fragment is identical to the nucleotide sequence of the pQR203 probe. The novel nucleotide sequence of 4485 bp of the 7.0 kb *Bam*HI fragment was determined. The open reading frame analysis of this 4485 bp DNA sequence revealed four open reading frames (*orf1*, *orf2*, *orf3* and *orf4*). However, there is a long intercistronic region between the orf2 and orf3, which is unusual for bacterial genomic DNA. *orf1*, *orf2*, *orf3* and *orf4* are nucleotide sequences of 685, 1092, 888 and 1086 basepairs respectively. These four ORFs are translated in the same direction, from left to right, and in the opposite direction to the *camR* gene (see Figure 3.24). Over all G+C content of the 4485 bp *Bam*HI fragment is 56.3%.

The Southern hybridisation of *P. putida* genomic DNA digested with *Kpn*I using the pQR277 probe derived from 3,6-diketocamphane 1,6-monooxygenase, showed that the probe bound to a 4201 bp *Kpn*I fragment. However, only 0.47 kb of the 0.7 kb pQR277 probe was complementary with the 4201 bp *Kpn*I fragment. The G+C content of the 4201-bp DNA fragment is 61.3%. The open reading frame analysis by the FramePlot and ORF finder of this 4201-bp *Kpn*I fragment showed that this fragment has four open reading frames which were designed as *orf5*, *orf6*, *orf7* and *orf8*. *orf5*, *orf6*, *orf7* and *orf8* are nucleotide sequences of 1272, 1098, 618 and 216 basepairs respectively. The *orf5*, *orf6* and *orf8* are translated in the same direction (left to right) compared to the *orf7*, which is translated in the opposite direction (see Figure 3.25). Further investigation of the 4485 bp DNA fragment and the 4201 bp *Kpn*I fragment will be described in the next chapter.

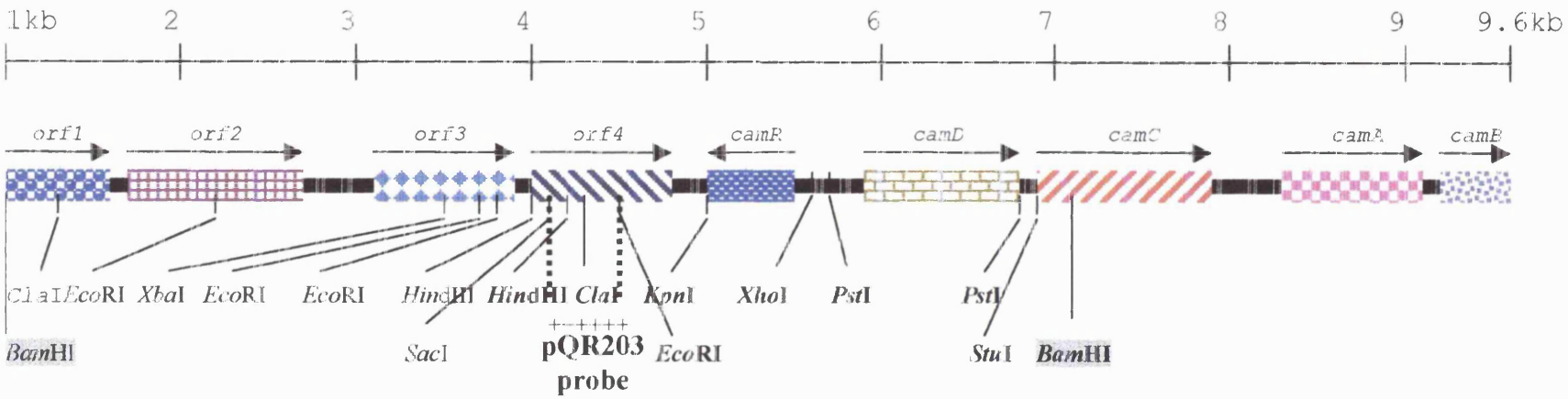


Figure 3.24 Genetic organisation of the genes (*orf1*, *orf2*, *orf3* and *orf4*) on the left-hand site of cam operon. The figure shows the direction of transcription in all four ORFs, and *camR*, *camD*, *camD*, *camC*, *camA* and *camC* (++++++ indicates the DNA region where the probe bound in the Southern hybridisation). The insert of pQR416 is the 7.0 kb *BamH*I fragment (the *BamH*I sites are highlighted in grey). The restriction sites at the end of the right hand side of the 7.0 kb BamHI fragment are at the same position as in previously published articles.

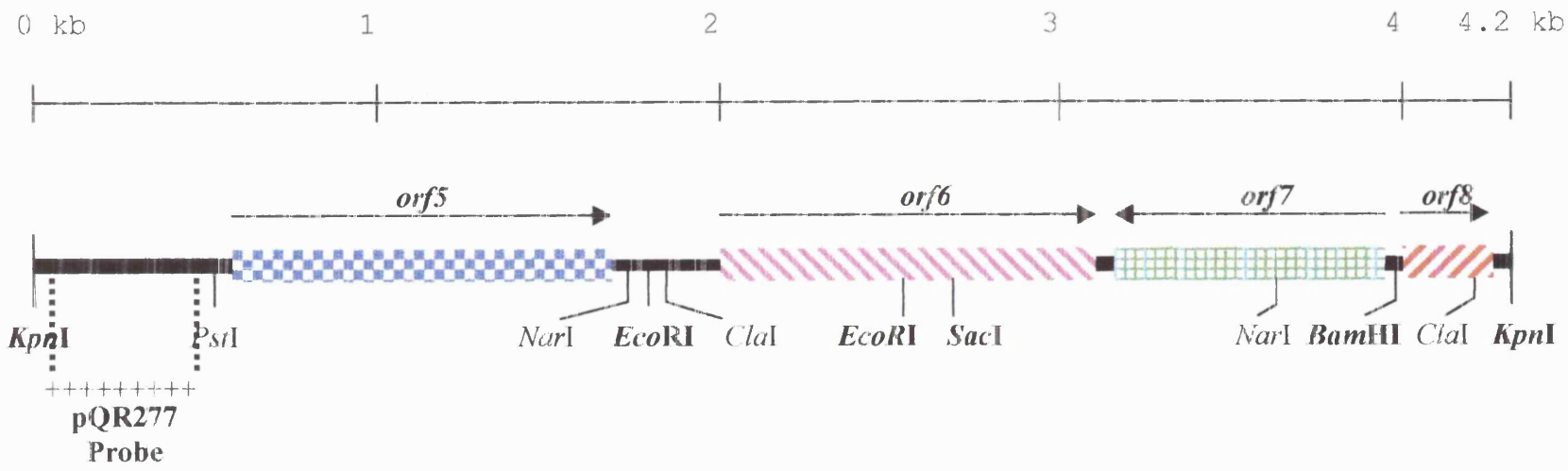


Figure 3.25 The genetic organisation of *orf5*, *orf6*, *orf7* and *orf8*. Restriction sites in bold were first obtained from our restriction mapping; the others were determined by computational analysis (++++++++ indicates the DNA region where pQR277 probe bound in the Southern hybridisation).

Chapter 4

Analysis of the open reading frames

4.1 Introduction

In the previous chapter, two novel DNA fragments were obtained, the *Bam*HI fragment containing *orf1*, *orf2*, *orf3* and *orf4* from the CAM plasmid and the *Kpn*I fragment, containing *orf5*, *orf6*, *orf7* and *orf8*, that is thought to come from the chromosome of *P. putida* NCIMB 10007. These eight open reading frames will be analysed in order to understand these sequences, and use this basic information in further studies. The analysis of eight open reading frames is divided into seven main aspects. They are 1) nucleotide sequence analysis, 2) G+C content and codon usage, 3) amino acid composition, 4) conserved domain search, 5) protein sequence homology search, 6) multiple alignment, and 7) protein secondary structure prediction.

In an analysis of DNA sequence, the start and stop codons and promoter sequences of genes provides key information when studying a gene. In addition, the G+C content and codon usage, which are usually found to vary in different organisms, are important considerations for adequate expression of an individual gene of interest. Therefore, codon usage pattern and amino acid composition of a gene of interest are important when considering an appropriate expression host. In an analysis of protein sequence, information of its conserved domains and similarity to other proteins are crucial to its characterisation and classification. In terms of information, protein secondary structure prediction is a common interpretation of protein structure and can give useful information. At this level, the analysis will provide an essential understanding, which is informative enabling further investigations.

4.2 Analysis of *orf1*

4.2.1 Nucleotide sequence analysis of *orf1*

The nucleotide sequence of *orf1* and its deduced amino acid are shown in Figure 4.2. The stop codon for *orf1* is TAA, however, the start codon of *orf1* is unknown, as it is likely that the *orf1* is an incomplete gene. The reasons are that the sequence of *orf1* showed no start codon proceeded by a ribosome-binding site. Also, it appeared that the *Bam*HI site cut the left-end of the *Bam*HI DNA fragment, interrupting this *orf1* before the real start. At the 3′ end of *orf1*, two inverted repeat sequences, which could lead to stem-loop configurations in transcription termination, were identified. Thses putative transcription termination sequences are located in the intercistronic region between *orf1* and *orf2*. These sequences are **GCGAAC**GCAC**GTTCGC** (28 nucleotides down stream of the *orf1*) **CACGTTCGC**TCTCC**GCGAACGTGC** (overlapping with the first inverted repeat sequence). These two inverted repeat has a free energy of – 5.7 and –12.8 kcal respectively (see Figure 4.1).

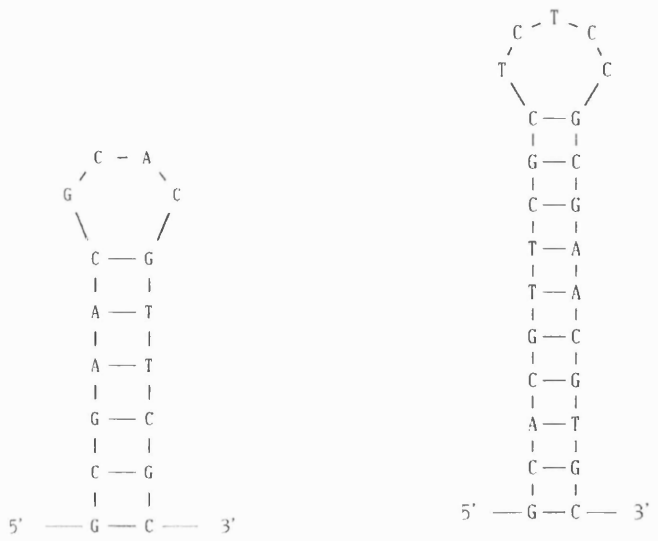


Figure 4.1 Putative stem-loop configurations in the intercistronic region between *orf1* and *orf2*.

```
1   - GGATCCCGTTGTCGCTGAAAAACTGATTCCCAAGGATCATCCCTTCGGTGCTAAGCGCGT - 60
 ?<--- Y  S  V  V  A  E  K  L  I  P  K  D  H  P  F  G  A  K  R  V

61  - GCCGATGGAAACCAATTATTACGAGACCTACAACCGCGATAACGTCCATCTGGTCGATAT - 120
       P  M  E  T  N  Y  Y  E  T  Y  N  R  D  N  V  H  L  V  D  I

121 - CCGTGAGGCACCGATTCAGGAGGTCACGCCGGAAGGGATCAAAACGGCTGACGCAGCCTA - 180
       R  E  A  P  I  Q  E  V  T  P  E  G  I  K  T  A  D  A  A  Y

181 - CGATCTTGATGTGATCATCTATGCCACGGGCTTTGATGCGGTCACTGGTTCACTCGACCG - 240
       D  L  D  V  I  I  Y  A  T  G  F  D  A  V  T  G  S  L  D  R

241 - GATCGACATCAGGGGCAAGGACAACGTCCGGCTGATCGATGCCTGGGCTGAAGGCCCAAG - 300
       I  D  I  R  G  K  D  N  V  R  L  I  D  A  W  A  E  G  P  S

301 - CACTTATCTCGGCCTTCAGGCTCGGGGCTTCCCGAACTTCTTCACCCTTGTCGGCCCGCA - 360
       T  Y  L  G  L  Q  A  R  G  F  P  N  F  F  T  L  V  G  P  H

361 - CAACGGCTCGACCTTTTGCAACGTCGGTGTATGTGGAGGATTGCAGGCGGAGTGGGTGCT - 420
       N  G  S  T  F  C  N  V  G  V  C  G  G  L  Q  A  E  W  V  L

421 - CCGAATGATCTCCTACATGAAGGATAACGGTTTCACCTATTCCGAACCGACCCAAGCAGC - 480
       R  M  I  S  Y  M  K  D  N  G  F  T  Y  S  E  P  T  Q  A  A

481 - AGAGAACCGGTGGACCGAGGAAGTCTATGCCGACTTCTCCCGCACTCTGCTTGCAGAGGC - 540
       E  N  R  W  T  E  E  V  Y  A  D  F  S  R  T  L  L  A  E  A

541 - CAATGCCTGGTGGGTCAAGACCACGACCAAACCGGATGGCTCGGTCGTGCGCCGCACGCT - 600
       N  A  W  W  V  K  T  T  T  K  P  D  G  S  V  V  R  R  T  L

601 - GGTGCATGTCAGTGGTGGACCGGAATACCGCAAGCGCTGCGAGCAGGTCGCTTATAATAA - 660
       V  H  V  S  G  G  P  E  Y  R  K  R  C  E  Q  V  A  Y  N  N
                                                   first inverted repeat
                                                   ------------->
661 - CTACAACGGATTTGAACTCGCCTAATAACCAGAATTGGCTACTTGCCTAGTGTGCGAACG - 720
       Y  N  G  F  E  L  A  * Stop codon of the orf1

                                                 start codon of the orf2
       <-------
721 - CACGTTCGCTCTCCGCGAACGTGCAACCAATAAGACCGATAAGAGGACACACTATGAAAT - 780
       ------->  <-------
       second inverted repeat                                M  K  C
```

Figure 4.2 DNA sequence of the *orf1* and its deduced amino acid sequence (in bold). The stop codon (TAA) of *orf1* and the start codon (ATG) of *orf2* (88 nucleotides downstream of *orf1*) are shown in bold. The inverted repeat sequences down stream of the *orf1* are indicated.

4.2.2 Sequence homology search of the Orf1 protein

Since the *orf1* is incomplete, further analysis has not been carried out. However, a sequence homology search using the 227 deduced amino acids of this incomplete *orf1* found this polypeptide to be highly related to proteins in the steroid and Baeyer-Villiger monooxygenase family.

The most highly related protein to the deduced amino acid of *orf1* in the recent GenBank database is a putative steroid monooxygenase from *Rhodococcus rhodochrous* (56% similarity). The known-steroid monooxygenase protein from *Rhodococcus rhod*ochrous was found to also have a high similarity level to the *orf1* deduced amino acid (56% similarity). Other known-Baeyer-Villiger monooxygenases found to have high identity and similarity to the deduced amino acid of *orf1* were cyclopentanone 1,2-monooxygenase from *Pseudomonas* sp., cyclohexanone monooxygenase from *Brevibacterium* sp. HCU, cyclohexanone 1,2-monooxygenase from *Acinetobacte*r sp. NCIMB 9871, cyclododecanone monooxygenase from *Rhodococcus* sp. and 4-hydroxyacetophenone monooxygenase from *Pesudomonas fluorescens* with 36%, 38%, 36%, 32% and 38% identity respectively (see Table 4.1). The deduced amino acid of *orf1* is; moreover, similar to a number of the putative proteins of the Baeyer-Villiger monooxygenase family in several bacteria.

Table 4.1 Proteins that have the high level of similarity to the deduced amino acid of *orf1*.

Protein	Microorganism	Identity	Similarity	Gap	Score (bits)	Expect value	Accession number
Putative steroid monooygenase	*Rhodococcus rhodochrous*	41%	56%	0%	186	2×10^{-46}	AJ418062
Steroid monooxygenase	*Rhodococcus rhodochrous*	43%	56%	3%	173	1×10^{-42}	AB010439
Baeyer-Villiger monooxygenase homologue	*Rhodococcus erthropolis*	39%	56%	1%	161	4×10^{-39}	AJ303350
Steroid monooxygenase	*Caulobacter crescentus*	39%	52%	3%	155	5×10^{-37}	AE005856
Cyclopentanone 1,2-monooygenase	*Pseudomonas* sp.	36%	54%	2%	152	4×10^{-36}	AB022102
Cyclohexanone monooxygenase	*Brevibacterium* sp.HCU	38%	55%	4%	151	4×10^{-36}	AF257214
Cyclohexanone 1,2-monooxygenase	*Acinetobacter* sp. NCIMB 9874	36%	54%	3%	140	7×10^{-33}	AB026668
Cyclododecanone monooxygenase	*Rhodococcus* sp.	32%	49%	1%	82.4	4×10^{-15}	AY052630
4-hydroxyaceto-pentanone monooxygenase	*Pseudomonas fluorescens*	38%	54%	6%	75.1	6×10^{-13}	AF355751
Probable flavin-containing monooxygenase	*Pseudomonas aeroginosa*	27%	44%	3%	72.8	3×10^{-12}	AE004582
Monooxygenase flavin-binding family	*Mycobacterium tuberculosis* CDC 1551	31%	46%	6%	68.9	4×10^{-11}	AE007131

4.2.3 Multiple alignment of the deduced amino acid of *orf1*

The multiple alignment of the deduced amino acid of *orf1* and three related proteins is shown in Figure 4.3. In the Figure, the Orf1 protein is aligned with the known proteins of steroid monooxygenase from *Rhodococcus rhodochrous* (Morii, *et al.*, 1999), cyclohexanone monooxygenase from *Brevibacterium* sp. HCU (Brzostowicz *et al.*, 2002) and cyclohexanone 1,2-monooxygenase from *Acinetobacter sp.* NCIMB 9871 (Chen *et al.*, 1988). This alignment also indicates an undiscovered polypeptide at the N-terminal sequence of Orf1 protein, which can be displayed as gaps (-) (see also Figure 4.3). The multiple alignment suggests that the *orf1* encodes a protein of around 550 amino acids derived from a gene of about 1650 nucleotides. The multiple alignment of the deduced amino acid of *orf1* with these steroid and cyclohexanone monooxygenases also shows the conserved domain of ATG-motif, FAD and NAD(P)H-binding domain, which is commonly found in flavoproteins (Vallon O., 2000). This ATG-motif is located at the amino acid position 68 to 70 on the deduced amino acid of *orf1* (see Figure 4.3). In addition, the dinucleotide binding-motif of GxGxxG has been found to be present (at the N-terminus) in FAD-containing flavoproteins. However, this motif would be consisting in the N-terminus of the deduced amino acid of *orf1*, which is not yet discovered.

```
CHMO-HCU -------------MPITQQLDHDAIVIGAGFSGLAILHHLR-EIGLDTQIVEATDGIGGT 46
CHMO     ----------------SQKMDFDAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGT 44
STMO     MNGQHPRSVVTAPDATTGTTSYDVVVVGAGIAGLYAIHRFR-SQGLTVRAFEAASGVGGV 59
Orf1     ------------------------------------------------------------

CHMO-HCU WWINRYPGVRTDSEFHYYSFSFSKEVRDEWTWTQRYPDGEEVCAYLNFIADRLDLRKDIQ 106
CHMO     WYWNRYPGALTDTETHLYCYSWDKELLQSLEIKKKYVQGPDVRKYLQQVAEKHDLKKSYQ 104
STMO     WYWNRYPGARCDVESIDYSYSFSPELEQEWNWSEKYATQPEILAYLEHVADRFDLRRDIR 119
Orf1     ------------------------------------------------------------

CHMO-HCU LNSRVNTARWNETEKYWDVIFEDGSSKRARFLISAMGALSQAIFPAIDGIDEFNGAKYHT 166
CHMO     FNTAVQSAHYNEADALWEVTTEYGDKYTARFLITALGLLSAPNLPNIKGINQFKGELHHT 164
STMO     FDTRVTSAVLDEEGLRWTVRTDRGDEVSARFLVVAAGPLSNANTPAFDGLDRFTGDIVHT 179
Orf1     ------------------------------------------------------------

CHMO-HCU AAWPADGVDFTGKKVGVIGVGASGIQIIPELAKLAGELFVFQRTPNYVVESNNDKVDAEW 226
CHMO     SRWPDD-VSFEGKRVGVIGTGSTGVQVITAVAPLAKHLTVFQRSAQYSVPIGNDPLSEED 223
STMO     ARWPHDGVDFTGKRVGVIGTGSSGIQSIPIIAEQAEQLFVFQRSANYSIPAGNVPLDDAT 239
ORF1     ------------------------------------------------------------
CHMO-HCU MQYVRDNYDEIFERASKHPFGVDMEYPTDSAVEVSEEERKRVFESKWEEGG-FHFANECF 285
CHMO     VKKIKDNYDKSLGWCMNSALAFALNESTVPAMSVSAEERKAVFEKAWQTGGGFRFMFETF 283
STMO     RAEQKANYAERRRLSRESGGGSPHRPHPKSALEVSEEERRAVYEERWKLGG--VLFSKAF 297
Orf1     ------------------------------------------------------------

CHMO-HCU TDLGTSPEASELASEFIRSKIREVVKDPATADLLCPKSYSFNGKRVPTGHGYYETFNRTN 345
CHMO     GDIATNMEANIEAQNFIKGKIAEIVKDPAIAQKLMPQ--DLYAKRPLCDSGYYNTFNRDN 341
STMO     PDQLTDPAANDTARAFWEEKIRAVVDDPAVAELLTPKDHAIGAKRIVTDSGYYETYNRDN 357
Orf1     --------------------------YSVVAEKLIPKDHPFGAKRVPMETNYYETYNRDN 34
                                   .. *: * *:.  : .**     .**:*:** *
                                          ATG-motif
CHMO-HCU VHLLDARGTPITRISSKGIVHGDTEY-ELDAIVFATGFDAMTGTLTNIDIVGRDGVILRD 404
CHMO     VRLEDVKANPIVEITENGVKLENGDFVELDMLICATGFDAVDGNYVRMDIQGKNGLAMKD 401
STMO     VELVDLRSTPIVGMDETGIVTTGAHY-DLDMIVLATGFDAMTGSLDKLEIVGRGGRTLKE 416
Orf1     VHLVDIREAPIQEVTPEGIKTADAAY-DLDVIIYATGFDAVTGSLDRIDIRGKDNVRLID 93
         *.* * :  **  :   *:   .  : :** :: ******: *.  .::* *:..  : :

CHMO-HCU KWAQDGLRTNIGLTVNGFPNFLMSLGPQTP--YSNLVVPIQLGAQWMQRFLKFIQERGIE 462
CHMO     YWKE-GPSSYMGVTVNNYPNMFMVLGPNGP--FTNLPPSIESQVEWISDTIQYTVENNVE 458
STMO     TWAA-GPRTYLGLGIDGFPNFFNLTGPGSPSVLANMVLHSELHVDWVADAIAYLDARGAA 475
Orf1     AWAE-GPSTYLGLQARGFPNFFTLVGPHNGSTFCNVGVCGGLQAEWVLRMISYMKDNGFT 152
          *   *  : :*:   .:**::   **   :   *:       .:*:   : :   ..

CHMO-HCU VFESSREAEEIWNAETIRGAESTVMSIEGPKAGAWFIGGNIP---GKSREYQVYMGGGQV 519
CHMO     SIEATKEAEEQWTQTCANIAEMTLF----PKAQSWIFGANIP---GKKNTVYFYLGGLKE 511
STMO     GIEGTPEAVADWVEECRNRAEASLLN----SANSWYLGANIP---GRPRVFMPFLGGFGV 528
Orf1     YSEPTQAAENRWTEEVYADFSRTLLA----EANAWWVKTTTKPDGSVVRRTLVHVSGGPE 208
           * :  *   *        . :::     .* :* .  .  ....  .    .:.*

CHMO-HCU YQDWCREAEESDYATFLNADSIDGEKVRESAGMK 553
CHMO     YRTCASNCKNHAYEGFD--IQLQRSDIKQPANA- 542
STMO     YREIITEVAESGYKGFA---ILEG---------- 549
Orf1     YRKRCEQVAYNNYNGFE----LA----------- 227
         *:    :     *  *     :
```

Figure 4.3 Multiple alignment of the Orf1 protein with the related proteins of steroid monooygenase [*Rhodococcus rhodochrous*] (STMO), cyclohexanone 1,2-monooxygenase [*Acinetobacter* sp. NCIMB 9871] (CHMO) and cyclohexanone monooxygenase [*Brevibacterium* sp. HCU] (CHMO-HCU). ATG-motif, FAD and NAD(P)H-binding domain, is idicated and shown in bold.

4.3 Analysis of *orf2*

4.3.1 Nucleotide sequence analysis of *orf2*

The complete *orf2* is 1092 nucleotides in length and encodes a protein of 363 amino acids. The start and stop codons for the *orf2* are ATG and TGA respectively. These start and stop codons correspond to that used most by other genes in the CAM plasmid. The start codon is proceeded by a ribosome binding sequence, which indicates the most likely true start codon for the *orf2*. The ribosome-binding sequence (Shine-Dalgano sequence) of GAGGA upstream of the start codon 6 basepairs, was identified. At the 3′ terminal of *orf2*, the inverted repeat sequence **CCACGCCGC**TAT**GCGGCGTGG** 27 nucleotides downstream of the *orf2* gene, was also identified (see Figure 4.4 and 4.5). The sequence is followed by a sequence of CTCCATTTCTCTC. This inverted repeated could form a stem-loop configuration, which is may be an important regulatory regulation mechanism for this gene. This hairpin structure has a free energy of –11.9 kcal (see Figure 4.4). Whether this inverted repeat for ORF2 is a real transcription termination, it has yet to be determined.

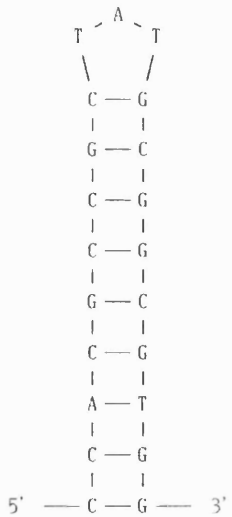


Figure 4.4 The inverted repeat sequence with the potential stem-loop configuration downstream of the *orf2*.

```
            orf1 -------->  stop
660  - CTACAACGGATTTGAACTCGCCTAATAACCAGAATTGGCTACTTGCCTAGTGTGCGAACG - 720
       Y  N  G  F  E  L  A  *
          Inverted repeat                          RBS       start orf2-->
721  - CACGTTCGCTCTCCGCGAACGTGCAACCAATAAGACCGATAAGAGGACACACTATGAAAT - 780
                                                              M  K  C

781  - GCGGATTTTTCCATACCCCATACAACTTGCCGACCCGTACCGCTCGGCAGATGTTCGACT - 840
       G  F  F  H  T  P  Y  N  L  P  T  R  T  A  R  Q  M  F  D  W

841  - GGTCCCTCAAGCTGGCGCAGGTTTGTGACGAGGCCGGTTTCGCCGACTTCATGATCGGCG - 900
       S  L  K  L  A  Q  V  C  D  E  A  G  F  A  D  F  M  I  G  E

901  - AGCATTCCACGCTGGCCTGGGAAAATATCCCCTGCCCGGAAATCATCATCGGCGCCGCAG - 960
       H  S  T  L  A  W  E  N  I  P  C  P  E  I  I  I  G  A  A  A

961  - CACCGCTGACCAAGAACATCCGCTTTGCACCGATGGCGCATTTGCTGCCTTACCACAACC - 1020
       P  L  T  K  N  I  R  F  A  P  M  A  H  L  L  P  Y  H  N  P

1021 - CGGCTACCCTGGCGATCCAGATCGGCTGGCTGTCGCAGATTCTCGAAGGCCGCTACTTCC - 1080
       A  T  L  A  I  Q  I  G  W  L  S  Q  I  L  E  G  R  Y  F  L

1081 - TCGGCGTGGCGCCGGGTGGCCACCATACCGATGCCATCCTGCATGGCTTCGAAGGCATTG - 1140
       G  V  A  P  G  G  H  H  T  D  A  I  L  H  G  F  E  G  I  G

1141 - GCCCGCTACAGGAGCAGATGTTCGAATCCCTGGAGCTGATGGAAAAAATCTGGGCCCGCG - 1200
       P  L  Q  E  Q  M  F  E  S  L  E  L  M  E  K  I  W  A  R  E

1201 - AGCCCTTCATGGAGAAAGGCAAGTTCTTCCAGGCTGGCTTCCCCGGCCCGGACACCATGC - 1260
       P  F  M  E  K  G  K  F  F  Q  A  G  F  P  G  P  D  T  M  P

1261 - CCGAGTACGATGTGGAGATCGCCGACAACAGCCCCTGGGGCGGACGCGAGTCGATGGAAG - 1320
       E  Y  D  V  E  I  A  D  N  S  P  W  G  G  R  E  S  M  E  V

1321 - TCGCGGTCACCGGCCTGACCAAGAATTCCTCGTCGCTGAAGTGGGCGGGTGAGCGCAACT - 1380
       A  V  T  G  L  T  K  N  S  S  S  L  K  W  A  G  E  R  N  Y

1381 - ACAGTCCGATCTCCTTCTTCGGCGGTCACGAAGTCATGCGCTCGCATTACGACACCTGGG - 1440
       S  P  I  S  F  F  G  G  H  E  V  M  R  S  H  Y  D  T  W  A

1441 - CGGCGGCTATGCAGTCGAAAGGCTTCACTCCCGAGCGTTCCCGCTTCCGTGTCACCCGTG - 1500
       A  A  M  Q  S  K  G  F  T  P  E  R  S  R  F  R  V  T  R  D

1501 - ACATCTTCATTGCCGACACCGATGCCGAAGCGAAGAAGCGTGCCAAGGCCAGTGGCCTGG - 1560
       I  F  I  A  D  T  D  A  E  A  K  K  R  A  K  A  S  G  L  G

1561 - GGAAAAGTTGGGAGCACTATCTGTTCCCGATCTACAAGAAGTTCAATCTGTTCCCCGGCA - 1620
       K  S  W  E  H  Y  L  F  P  I  Y  K  K  F  N  L  F  P  G  I

1621 - TCATCGCCGATGCCGGCCTCGACATCGATCCGAGCCAGGTGGACATGGATTTCCTCGCTG - 1680
       I  A  D  A  G  L  D  I  D  P  S  Q  V  D  M  D  F  L  A  E

1681 - AGCATGTCTGGCTTTGTGGCTCGCCGGAAACGGTGAAAGGCAAGATCGAGCGCATGATGG - 1740
       H  V  W  L  C  G  S  P  E  T  V  K  G  K  I  E  R  M  M  E

1741 - AGCGTAGCGGTGGCTGTGGGCAGATAGTCGTCTGCTCCCACGACAATATCGACAACCCGG - 1800
       R  S  G  G  C  G  Q  I  V  V  C  S  H  D  N  I  D  N  P  E

1801 - AACCTTATTTCGAATCGCTACAGCGCCTTGCCAGCGAAGTGTTACCGAAGGTTCGAATGG - 1860
       P  Y  F  E  S  L  Q  R  L  A  S  E  V  L  P  K  V  R  M  G
                                          Inverted repeat
1861 - GCTGAGGGAACACCAATTCGGGAGAGGTTGGACCACGCCGCTATGCGGCGTGGCTCCATT - 1920
       * stop

1921 - TCTCTCTGCAATCGGGCATTGCGATCACCTCAGCCGAAACTGGTGGCAGTGGCCTGTGCT - 1980

1981 - GTGCTCCTCGTCGCAATGCTGTGCGGTTCTGACATATAGCATGTTCACCAGCATCGCAGA - 2040
```

Figure 4.5 Nucleotide sequence of the *orf2* and its deduced amino acid. The deduced amino acid sequence of *orf2* is shown in bold below the DNA sequence. The start (ATG) and stop (TGA) codons are shown in bold, and the potential ribosome-binding site (RBS) for the *orf2* is shown in bold and underlined. Two inverted repeat sequences are also shown in bold and underlined.

4.3.2 G+C content and codon usage of the *orf2*

The G+C content of *orf2* is 59.25%, which is slightly lower to that of *P. putida* (60.7-62.5%) (Mundel, 1969). The G+C content of *orf2* is similar to that of *camA* (58.9%) and *camB* (57.1%), genes on the CAM plasmid, but much higher than that of OCT plasmid (the plasmid in the same Inc2 group with the CAM plasmid) with the G+C content of 45% (van Beilen *et al.*, 1992). The G+C contents in the first and second position of *orf2* are 57.69 % and 43.13 % respectively. In the third position, G+C composition is 76.92%, indicates the preferential of G and C at the third position of the *orf2* genetic codes (see Table 4.2).

Table 4.2 G+C content of the *orf2*.

Nucleotide	1		2		3		Total	
	Number	%	Number	%	Number	%	Number	%
A	88	24.18	106	29.12	31	8.52	225	20.60
C	87	23.90	85	23.35	161	44.23	333	30.49
G	123	33.79	72	19.78	119	32.69	314	28.75
T	66	18.13	101	27.75	53	14.56	220	20.15
A+C	175	48.08	191	52.47	192	52.75	558	51.10
A+G	211	57.97	178	48.90	150	41.21	539	49.36
A+T	154	42.31	207	56.87	84	23.08	445	40.75
G+C	210	57.69	157	43.13	280	76.92	647	59.25

The codon usage of *orf2* is shown in Table 4.3. The preferential usage codons are GGC (Gly), GAG (Glu), GAC (Asp), AAG (Lys), ATC (Ile), TAC (Tys), TTC (Phe), CGC (Asp), CAA (Glu), CTG (Leu) and CCG (Pro). Unused codons are GTA (Val), AGG (Arg), ACA (Thr), TCA (Ser), TCT (Ser) and CAA (Gln). The codons that have been used only once are ATA (Ile), ACT (Thr), TTA (Leu), CGG (Arg), CGA (Arg) and CCA (Pro). The triplet codons for Ser and Gln with A or T nucleotides in the third codon position were not found. This maybe because there is a strict bias towards G and C nucleotides in the third codon position.

Table 4.3 Codon usage of the *orf2*.

Amino acid	Codon	Number	/1000	Percentage	Amino acid	Codon	Number	/1000	Percentage
Gly	GGG	2	5.49	6.0%	Trp	TGG	9	24.73	100.0%
Gly	GGA	2	5.49	6.0%	End	TGA	1	2.75	100.0%
Gly	GGT	5	13.74	16.0%	Cys	TGT	3	8.24	50.0%
Gly	GGC	23	63.19	72.0%	Cys	TGC	3	8.24	50.0%
Glu	GAG	15	41.21	54.0%	End	TAG	0	0	0
Glu	GAA	13	35.71	46.0%	End	TAA	0	0	0
Asp	GAT	6	16.48	33.0%	Tyr	TAT	2	5.49	22.0%
Asp	GAC	12	32.97	67.0%	Tyr	TAC	7	19.23	72.0%
Val	GTG	5	13.74	36.0%	Leu	TTG	2	5.49	8.0%
Val	GTA	0	0	0	Leu	TTA	1	2.75	4.0%
Val	GTT	2	5.49	14.0%	Phe	TTT	2	5.49	9.0%
Val	GTC	7	19.23	50.0%	Phe	TTC	21	57.69	91.0%
Ala	GCG	9	24.73	29.0%	Ser	TCG	8	21.98	36.0%
Ala	GCA	3	8.24	10.0%	Ser	TCA	0	0	0
Ala	GCT	5	13.74	16.0%	Ser	TCT	0	0	0
Ala	GCC	14	38.46	45.0%	Ser	TCC	7	19.23	32.0%
Arg	AGG	0	0	0	Arg	CGG	1	2.75	6.0%
Arg	AGA	0	0	0	Arg	CGA	1	2.75	6.0%
Ser	AGT	3	8.24	14.0%	Arg	CGT	6	16.48	35.0%
Ser	AGC	4	10.99	18.0%	Arg	CGC	9	24.73	53.0%
Lys	AAG	12	32.97	67.0%	Gln	CAG	11	30.22	100.0%
Lys	AAA	6	16.48	33.0%	Gln	CAA	0	0	0
Asn	AAT	4	10.99	40.0%	His	CAT	7	19.23	58.0%
Asn	AAC	6	16.48	60.0%	His	CAC	5	13.74	42.0%
Met	ATG	15	41.21	100.0%	Leu	CTG	14	38.46	54.0%
Ile	ATA	1	2.75	4.0%	Leu	CTA	2	5.49	8.0%
Ile	ATT	3	8.24	13.0%	Leu	CTT	2	5.49	8.0%
Ile	ATC	19	52.2	83.0%	Leu	CTC	5	13.74	19.0%
Thr	ACG	2	5.49	13.0%	Pro	CCG	14	38.46	58.0%
Thr	ACA	0	0	0	Pro	CCA	1	2.75	4.0%
Thr	ACT	1	2.75	7.0%	Pro	CCT	2	5.49	8.0%
Thr	ACC	12	32.97	80.0%	Pro	CCC	7	19.23	29.0%

4.3.3 Amino acid composition of the Orf2 protein

Figure 4.6 shows the amino acid composition of Orf2 protein. The deduced amino acid sequence of *orf2* is Gly (G) and Ala (A) rich with 32 (G) and 31 (A) residues, out of the 363 predicted amino acid residues of the coding protein of *orf2*. Cys (C) is significantly less distribution in Orf2 protein with 6 residues or 1.7%, however, this distribution of cystine is normal for prokaryote proteins (1.0%) (Doolittle, 1986). The average amino acid compositions of proteins from *E. coli* (Doolittle, 1986) are as following (in percentage): 10.1 (A), 0.9 (C), 5.6 (D), 6.6 (E), 3.5 (F), 7.5 (G), 2.0 (H), 5.8 (I), 5.7 (K), 9.5 (L), 2.6 (M), 4.1 (N),4.1 (P), 4.3 (Q), 5.9 (R), 5.5 (S), 2.7 (T), 7.5 (V), 1.0 (W) and 2.7 (Y). A comparison of the amino acid composition of *orf2* to the average amino acid compositions of proteins from *E. coli* (Doolitle FR., 1986) demonstrated that both tend to have a similar distribution. The Orf2 protein however has a greater number of Trp (W) residues (2.5%) when compared to the (1.0%) composition of the proteins from *E. coli*. In addition, Pro (P), Phe (F), Met (M) and Trp (T) present in a greater number in the Orf2 protein (6.6%, 6.3%, 4.1% and 4.1% composition in the protein respectively) when compared to the *E. coli* proteins compositions for Pro, Phe, Met and Trp are 3.9, 3.5, 2.6 and 2.7 % respectively. The basic (K, R and H), acidic (D and E) and hydrophobic group (A, V, I, L, M, F, Y and W) amino acids for the Orf2 protein are 15.2%, 12.7% and 41.3% respectively. The Orf2 protein has a theoretical molecular weight and pI value of 40,704 and 5.58 respectively.

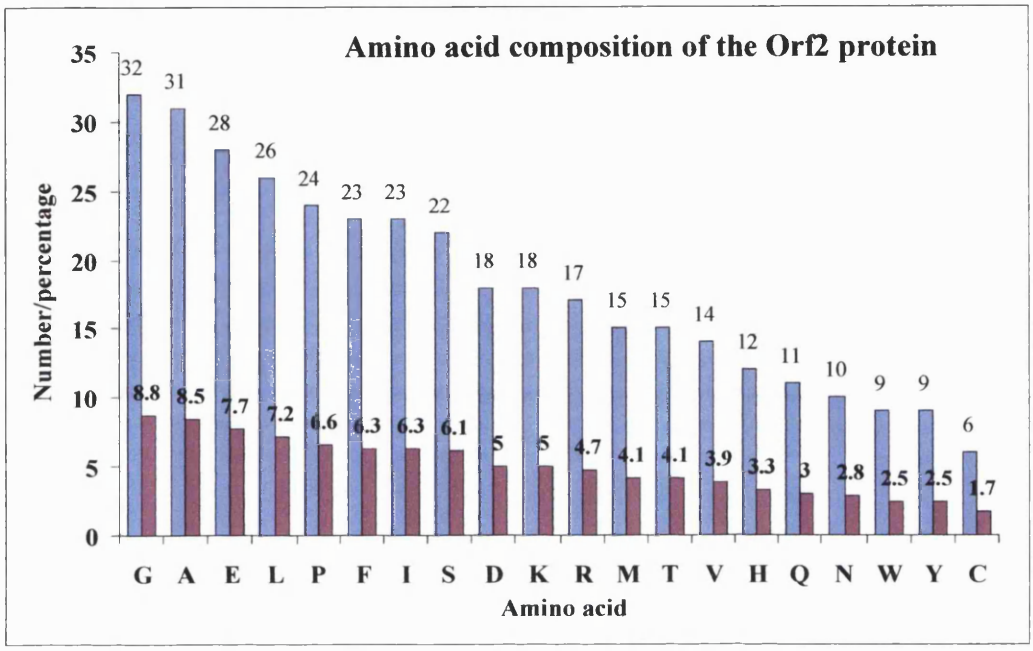


Figure 4.6 Amino acid composition of the Orf2 protein. X-axis represents the amino acids (A: Ala; C: Cys; D: Asp; E: Glu; F: Phe; G: Gly; H: His; I: Ile; K: Lys; L: Lue; M: Met; N: Asn; P: Pro; Q: Glu; R: Arg; S: Ser; T: Thr; V: Val; W: Trp and Y: Tyr). Y-axis represents a number of amino acids (blue columns) and the percentage of amino acid in molarity (red columns).

4.3.4 Conserved domain search of the Orf2 protein

The conserved domain of the FMN-binding bacterial luciferase, luciferase-like monooxygenase, was found in the amino acid sequence of Orf2 protein. This conserved domain is commonly found in luciferase proteins in marine bioluminous bacteria such as *Vibrio* species. The luciferase-like monooxygenase has 323 amino acid residues. There are 245 amino acid residues (75.9%) which overlap to 259 amino acid residues of the *orf2* deduced amino acid. A pairwise alignment of this overlapping amino acid

sequences is shown in Figure 4.7. The conserved domain sequence of the luciferase-like monooxygenase (residue 1 to 245) aligned to the Orf2 amino acid sequence (1-259). The finding that there is substantial homology of Orf2 protein to the luciferase-like monooxygenase suggests that the Orf2 protein is related to the proteins in the bacteria luciferase-like monooxygenase family.

```
CD-Length = 323 residues, only  75.9% aligned
Score = 91.2 bits (226), Expect = 5e-20

Orf2:  1    MKCG-FFHTPYNLPTRTARQMFDWSLKLAQVCDEAGFADFMIGEHSTLAWENIPCPEIII  59
Luc :  1    MKFGVFFLNFQR-PGQTSEEVIDNMVELAEAVDRLGFDTAWVAEHHFSPFGLVGAPLTAA  59

Orf2:  60   GAAAPLTKNIRFAPMAHLLPYHNPATLAIQIGWLSQILEGRYFLGVAPGGHHTDAILHGF  119
Luc :  60   AFLLGRTERIRVGTLGIVLPTHHPLRVAEEFGLLDQLSGGRFELGLGRGVYGFDFRFFG-  118

Orf2:  120  EGIGPLQEQMFESLELMEKIWAREPFMEKGKFFQAGFPGPDTMPEYDVEIADNSPWGGRE  179
Luc :  119  RDRDSQQALFEECYEILNDALRTGYCSADGDFYE--FPK----------ISVNPRPYQKP  166

Orf2:  180  SMEVAVTGLTKNSSSLKWAGERNYSPISFFG--GHEVMRSHYDTWAAAMQSKGFTPERSR  237
Luc :  167  HPPTWV--LATSPETVEWAAKYGL-PLVFKWIDSLAEREELLERYREAAAGHGNDISNVD  223

Orf2:  238  FRVTRDIFIADTDAEAKKRAKA  259
Luc :  224  HQLTLLVNVNEDGEKAKEEARP  245
```

Figure 4.7 Pairwise alignment of the conserved domain of luciferase-like monooxygenase and the conserved domain region on the Orf2 protein. Orf2: amino acid sequence of the Orf2 protein; and Luc: the conserved domain sequence of luciferase-like monooxygenase. The amino acid residues in red and blue indicate identities and similarities, respectively, between amino acid residues of the two sequences.

4.3.5 Sequence homology search of the Orf2 protein

The 363 amino acid sequence of Orf3 protein was used in a BLAST protein-protein sequence homology search. This coding sequence is highly related to the luciferase-related protein (*Mycobacterium tuberculosis* CDC 1551), a putative alkanal monooxygenase (*Streptomyces coelicolor*) and a limonene monooxygenase

(*Rhodococcus erythropolis*). The pair wise alignment of the luciferase related protein (*Mycobacterium tuberculosis* strain CDC 1551) with the deduced amino acid of *orf2* had the highest homology with 31% identity and 53% similarity. The pairwise alignment of the putative alkanal monooxygenase (*Streptomyces coelicolor*) with the deduced amino acid of ORF2 had 26% identity and 46% similarity. The pairwise alignment of limonene monooxygenase (*Rhodococcus erythropolis*) with the deduced amino acid of *orf2* had 25% identity and 44% similarity.

Moreover, there were several proteins, which produced a 22-26% identity alignment to the Orf2 protein. Example of these proteins are the probable peptide synthetase protein (*Ralstonia solanacearum*), the luciferase-alpha subunit (*Nostoc* sp. PCC 721) the conserved hypothetical protein (*Caulobacter crescentus*) the putative monooxygenase (*Streptomyces clavuligerus*), the alkanal monooxygenase-alpha *chain (Xenorhabdus luminescens*) and the luciferase-beta subunit (*Vibrio harveyi*) (see Table 4.4).

Table 4.4 Proteins that are highly related to the Orf2 protein.

Protein	Mocroorganism	Identity	Similarity	Gap	Score (bits)	Expect value	Accession number
Luciferase-related protein	*Mycobacterium tuberculosis* CDC1551	31%	53%	2%	197	1×10^{-49}	AE007053
Putative monooxygenase mmy0	*Streptomyces coelicolor*	26%	46%	9%	95.5	7×10^{-19}	AL590464
Limonene monooxygenase	*Rhodococcus erythropolis*	25%	44%	9%	92.0	9×10^{-18}	AJ272366
Probable peptide synthetase protein	*Ralstonia solanacearum*	24%	41%	11%	63.2	4×10^{-9}	AL646079
Luciferase alpha subunit	*Nostoc sp.* PCC 721	26%	42%	0%	58.2	1×10^{-7}	AP003589
Conserved hypothetical protein	*Caulobacter crescentus*	22%	39%	10%	54.3	2×10^{-6}	AE005869
Putative monooxygenase	*Streptomyces clavuligerus*	22%	38%	12%	53.9	2×10^{-6}	AF124929
Alkanal monooxygenase alpha chain	*Xenorhabdus luminescens*	24%	43%	0%	46.2	5×10^{-4}	C38448
Luciferase beta subunit	*Vibrio harveyi*	25%	44%	0%	44.3	0.002	X58791

Expect value: the number of hits that one might expect to see one match from the search with a similar score simply by chance.

Score: the quality of each pair-wise alignment assigned to each position.

Identity: identical or conserved residues in the alignment between 2 sequences.

Similarlity: score of non-identical residues between two sequences weighted by considering substitutions

Gap: non-overlapping alignment between two sequences

Interestingly, it was also found that the alignment of the N-terminal sequence of the *orf2* deduced amino acid sequence with that of 2,5-diketocamphane 1,2-monooxygenase, NADH-dependent Baeyer-Villiger monooxygenase (Kelly, 1997), had 80% identity (see Figure 4.8). With the exception of the amino acids in position 2, 3 and 7, all amino acids in the N-terminal sequence of 2,5-diketonecamphane 1,2-monooxygenase and the *orf2* deduced amino acid sequence are identical.

```
Orf2        1 - M K C G F F H T P Y N L P T R T A R Q M - 20
2,5-DKCMO   1 - M Q A G F F G T P Y D L P T R T A R Q M - 20
```

Figure 4.8 Alignment of the first 20 amino acid residues of Orf2 protein and that of 2,5-diketocamphane 1,2-monooxygenase. Amino acids in dark shadings indicate identities between the two sequences (Orf2: the N-terminal of the deuced amino acid of *orf2* and 2,5-DKCMO: the N-terminal sequence of 2,5-diketocamphane 1,2-monooxygenase).

With 80% identity in this alignment, it indicates that the Orf2 protein is highly related to 2,5-diketocamphane 1,2-monooxygenase, which also suggests that these two sequences are highly homologous.

4.3.6 Multiple sequence alignment of the Orf2 protein

To determine homology among protein sequences, multiple sequence alignments are commonly used to find the relationships between proteins that share similar sequence characteristics. CLUSTALW is the most widely multiple sequence alignment program

used at present. This multiple alignment method is previously described by Thomas and coworkers in 1994. Basically, the method of multiple sequence alignment is to arrange individual amino acid residues (identical or relative amino acids) from each protein sequences into the same position, similar to the way that protein sequences are assigned into a table. Proteins are assigned in the rows, and individual amino acids residues of proteins are separated by columns. These amino acid residues, which are identical or related, are arranged in the same column.

The deduced amino acid of *orf2* aligned with the luciferase-related protein (*Mycobacterium tuberculosis* CDC 1551), the limonene monooxygenase (*Rhodococcus erythropolis*) and the luciferase-alpha subunit (*Nostoc* sp. PCC 721) is shown in Figure 4.9.

The pairwise alignment of the *orf2* deduced amino acid sequence to the luciferase-related protein, limonene monooxygenase and alkanal monooxygenase had the pairwise score of 31, 21 and 13 bits respectively. These pairwise scores indicate that the Orf2 protein is more closely related to the luciferase-related protein, then the limonene monooxygenas and then the luciferase-alpha subunit.

```
Orf2    -------MKCGFFHTPYNLPTRTARQMFDWSLKLAQVCDEAGFADFMIGEHSTLAWENIP 53
Luc     -------MEIGIFLMPAHPPERTLYDATRWDLDVIELADQLGYVEAWVGEHFTVPWEPIC 53
LIMO    MTDDFGRVDFGAFLAPWHRADSDANFAIHQDLELVEHLDRLGFAEFWLGEHHSGGVEIVA 60
LucA    -------MKTGLFCN-YENHHQDSRRAIFEQVALVRQAEKLGFDEAWVTEHHFNEVNLSS 52
                :. * *  . .            .: : .  :. *: :  : **     :

Orf2    CPEIIIGAAAPLTKNIRFAPMAHLLPYHNPATLAIQIGWLSQILEGRYFLGVAPGGHHTD 113
Luc     APDLLLAQALLRTQQIKLAPGAHLLPYHHPVELAHRVAYFDHLAQGRFMLGVGASGIPGD 113
LIMO    SPEMFMAAAAQRTQRIKLGLGVVSLPYHHPFLVADRLVLLDHLSRGRMIFGAGPGQLADD 120
LucA    SILLLMAHLAGVTSTIKLGTAAVLLPFHNPIRVAEDIATLDNLCNGRLLFGVAKGGPFPQ 112
        . :: :.     *. *::.  .  **:*:*  :*  :  :.:: .** ::*.. .    :

Orf2    AILHGFEG-IGPLQEQMFESLELMEKIWAR-EPFMEKGKFFQAGFPGPDTMPEYDVEIAD 171
Luc     WALYDVDGKNGEHREMTREALEIMLRIWTEDEPWEHRGKYWNANG----IAPMFEGLMRR 169
LIMO    AKMLGIDP--IDSRRKMEEAFDVIHRLLAG-ETVTQKTDWFTCQDAYLHVAPYSN--IQK 175
LucA    HNKHFATP-MGESRAMMLEALALIQKLLYETDVSFN-GQYYQCDR-----LTVYPKPLQQ 165
                     :    *:: :: ::    :   .  .:: . .      .     :

Orf2    NSPWGGRESMEVAVTGLTKNSSSLKWAGERNYSPISFFGGHEVMRSHYDTWAAAMQSKGF 231
Luc     HIKPYQKPHPPIGVTGFSAGSETLKLAGERGYIPMSLDLNTEYVATHWDAVEEGALRSGR 229
LIMO    AVTATVSPTGPKLAGKYGSGILSLAATNP---------VGVEKLAEHWKIAEDIAAENGQ 226
LucA    EIPVYLATGDDVGIEFAAKHSFALMGGPP---------FALERLKKTVETYRALNSSGG- 215
                              :*      .  . :     * :    .         *

Orf2    TPERSRFRVTRDIFIADTDAEAKKRAKASGLGKSWEHYLFPIYKKFNLFPGIIADAGLDI 291
Luc     TPDRRDWRLVREVLVAETDEQAFRYAVDGTMGRAMREYVLPTFRMFGMTKFYKHNP--SV 287
LIMO    TVDRADWRLSGIMHVAETEEQARADVRHG--------LLYLMNYLSNITPGFAAAP---- 274
LucA    ----EKFVLARFFYVGKTYDEAVTEALPFIRYFSQKMTANSSQVMQNGSSGHKQFD---- 267
        : .   : :   . :..*  :*   .       :            .         . .

Orf2    DPSQVDMDFLAEHVWLCGSPETVKGKIERMMERSGGCGQIVVCSHDNIDNPEPYFESLQR 351
Luc     PDDEVTPEYLAENTFVVGSVQTVVDKLEATYDQVGGFGHLLILGFDYSDNPGPWKESLRL 347
LIMO    DVDSLIDGINDAGLAVIGTPEMAVTQIRRLQEKSGGFGKFLVLHGEWAS-TTAALHSFEL 333
LucA    RTNICFDEDYLIENSIIGDVATCRDKIKKFQDELD-LGTLALKPSSFAL--QKNQESLQR 324
          .            : *       ::.   :. .. * : :   .   ..    .*:.

Orf2    LASEVLPKVRMG------------------------------------------ 363
Luc     LAHEVMPRLNARLATKPATAVV------------------------------- 369
LIMO    IAQQVAPHFNGDLGPRLRGYNQTMNSNRSAADITQAAQEEAQKRFEAERAIRTN 387
LucA    YNQEVQNYV--------------------------------------------- 333
           :*   .
```

Figure 4.9 Multiple alignment of the deduced amino acid of *orf2* (Orf2) with luciferase related protein from *Mycobacterium tuberculosis* CDC 1551 (Luc), limonene monooxygenase from *Rhodococcus erythropolis* (LIMO) and luciferase-alpha subunit from *Nostoc* sp. PCC 721 (LucA). The conserved domain of luciferase-like monooxygenase is shown in bold.

* : identical or conserved residues in all sequences in the alignment

: : indicates conserved substitutes

. : indicates semi-conserved substitutions

- : gap, non-overlapping residue

The multiple alignment showed twenty-six amino acids are highly conserved among the luciferase-lelated protein (*Mycobacterium tuberculosis* CDC 1551), limonene monooxygenase (*Rhodococcus erythropolis*), luciferase-alpha subunit (*Nostoc* sp. PCC 271) and the Orf2 protein. These amino acids are G (4), F (6), G (35), E (43), H (44), T (65), I (68), L(77), P (78), H (80), P (82), A (86), G (98), R (99), G (103), E (130), L (195), E (213), G (230), T (249), A (253), G (309), G (329), S (348) and V (353).

4.3.7 Protein secondary structure prediction of Orf2 amino acid

In a prediction of three dimensional (3D) protein structure by computational analysis, prediction can only be made if a protein has more than 25% sequence identical to known 3D structures. For proteins that have no significant sequence identity to known 3D structures, the structure can be simplified. One-dimensional protein secondary structure prediction is a simplified method to describe a protein structure.

The assignments for protein secondary structure are usually based on DSSP (Dictionary of Secondary Structures of proteins); the database which contains the protein secondary structure of known 3D structures (Kabsch and Sander, 1983) and conformational preferences of amino acids (Williams et al., 1987; and Wilmot and Thronton, 1988). In addition, with the principle that a segment of amino acids forms a certain structure, the protein secondary structure can be predictable. However, at present, most protein secondary structure predictions can only achieve an accuracy of less than 80%.

Three different secondary structure prediction methods: DSC, PHD and Jpred, were used in this study. At the end of the predictions a consensus protein secondary structure based on the DSC, PHD and Jpred method is generated.

The DSC prediction method was described by King and Sternberg in 1996. This protein secondary structure prediction is assigned by important concepts in secondary structure prediction such as residue conformational propensities, hydrophobicity, the insertions and deletions in alignment, and the distance from the end of the sequence (King and Sternberg, 1996).

The PHD predicts one-dimentional protein structure by profile-based neural networks. This method is probably the most accurate protein secondary structure prediction, which offers an accuracy of 71.9% (Rost, 1996).

The Jpred is a consensus secondary structure prediction server. In this method, the protein is compared and aligned with the homologous protein sequences from a BLAST search using CLUSTALW. The Jpred method can also generate a consensus sequence from the combination of different prediction methods: DSC, PHD, PREDATOR and NNSSP, which can achieve an accuracy of 72.9% (Cuff *et a*l., 1998).

The protein secondary structure predictions of Orf2 protein is shown in Figure 4.10. The predicted secondary structure of Orf2 protein is also aligned with that of the known 3D structure of luciferase-beta subunit. This luciferase-beta-subunit *from Vibrio haveyi* (Fisher *et al.*, 1996) has the best identity (18%) to the Orf2 protein in the recent database.

The predicted protein secondary structure of Orf2 protein showed that the protein consists of twelve helices (α_1- α_{12}) and ten strands (β_1- β_{10}). Most of the structures of the predicted secondary structure of Orf2 protein appear to be coincide with that of the luciferase-beta-subunit. However, the alignment of the luciferase-beta-subunit has a wider gap introduced into its α_7b-helice when it is aligned with the predicted protein secondary structure of Orf2 protein.

```
LucbSec   CEEEEEECCC CCCCCCHHHH HHHHHHHHHH HHCCCCCEEE ECCCCCCCCC CCCCHHHHHH HHHHHCCCCE
Lucb      MKFGLFFLNF MNSKRSSDQV IEEMLDTAHY VDQLKFDTLA VYENHFSNNG VVGAPLTVAG FLLGMTKNAK
Orf2      MKCGFFHTPY NLPTRTARQM FDWSLKLAQV CDEAGFADFM IGEHSTLAWE NIPCPEIIIG AAAPLTKNIR
DSC       CCCCCCCCCC CCCCHHHHHH HHHHHHHHHH HCCCCCCEEE ECCCCCCCCC CCCCCCEEEE CCCCCHHHHH
PHD       CCCCEEECCC CCCHHHHHHH HHHHHHHCCC CHHHCCHHHH HHHHCHHHHH CCCCHHHHHH HHHHHHHHHC
Jpred     CCCCEEECCC CCCCCCCHHH HHHHHHHHHH HHHCCCCEEE ECCCCCCCCC CCCCHHHHHH HHHHCCCCEE
ConSeq    CCCCEEECCC CCCCHHHHHH HHHHHHHHHH HHHCCCCEEE ECCCCCCCCC CCCCHHHHHH HHHHCHHHHC
                                α1                  β1                  α2        α3

LucbSec   EEEEEEECCC CCHHHHHHHH HHHHHHHCCC EEEEEECCCC HHHHHHHCC. CCCHHHHHHH HHHHHHHHHH
Lucb      VASLNHVITT HHPVRVAEEA CLLDQMSEGR FAFGFSDCEK SADMRFFNR. PTDSQFQLFS ECHKIINDAF
Orf2      FAPMAHLLPY HNPATLAIQI GWLSQILEGR YFLGVAPGGH HTDAILHGFE GIGPLQEQMF ESLELMEKIW
DSC       CCCCCCCCCC CCCCCEEEEE EHHHHHCCCC EEEEECCCCC CCCEEEECCC CCCCCHHHHH HHHHHHHHHH
PHD       CCCCEEECCC CCCHHHHHHH HHHHHHHCCC CEEEEECCCC CCCEEEEECC CCCCCHHHHH HHHHHHHHHH
Jpred     ECCCCECCCC CCCHHHHHHH HHHHHHCCCC EEEEECCCCC CCCCCCCCCC CCCCHHHHHH HHHHHHHHHH
ConSeq    CCCCCECCCC CCCHHHHEHH HHHHHHCCCC EEEEECCCCC CCCEEEECCC CCCCCHHHHH HHHHHHHHHH
                      α4    α5          β2          β3                α6

LucbSec   HHCEECCCCC CCCC..CCEE CCCCCCCC.. ........CC CCEEEE..CC CHHHHHHHHH HCCCEEECCC
Lucb      TTGYCHPNND FYSF..PKIS VNPHAFTE.. ........GG PAQFVN..AT SKEVVEWAAK LGLPLVFRWD
Orf2      AREPFMEKGK FFQAGFPGPD TMPEYDVEIA DNSPWGGRES MEVAVTGLTK NSSSLKWAGE RNYSPISFFG
DSC       HCCHHHCCCC EEECCCCCCC CCCCEEEEEC CCCCCCCCCE EEEEEECCCC CCCCEEECCC CCCCCEEECC
PHD       CCCCCCCCCC EEEECCCCCC CCCCCCEEEE CCCCCCCCCC CEEEEEEEEE CCCCHHHHHH CCCCCCEECC
Jpred     CCCCCCCCCC CEECCCCCCC CCCCCCCCCC CCCCCCCCCC CCEEEECCCC CCHHHHHHHH CCCCCEECCC
ConSeq    CCCCCCCCCC EEECCCCCCC CCCCCCEEEE CCCCCCCCCC CEEEEECCCC CCCCHHHHHH CCCCCEEECC
                     β4          β5          β6          α7          β7

LucbSec   CCHHHHHHHH HHHHHHHHHH CCCCCCCCCE EEEEEEECCC HHHHHHHHHH H.....HHHH HHHHHCCCCH
Lucb      DSNAQRKEYA GLYHEVAQAH GVDVSQVRHK LTLLVNQNVD GEAARAEARV Y.....LEEF VRESYSNTDF
Orf2      GHEVMRSHYD TWAAAMQSKG FTPERSRFRV TRDIFIADTD AEAKKRAKAS GLGKSWEHYL FPIYKKFNLF
DSC       CCCEEECCCC CCHHCCCCCC CCCCCEEEEE CCEEEEECCH HHHHHEEEEC CCCCCCEEEC CCCCCCCCCC
PHD       CCCEEHHHHH HHHHHHHHHC CCCCCCCCCE EEEEEECCCH HHHHHHHHHC CCCCCCEEEE EEEECEECCC
Jpred     CHHHHHHHHH HHHHHHHHCC CCCCCCCEEE EEEEEECCCH HHHHHHHHHC CCCCCCECCC CCCCCCCCCC
ConSeq    CCCEEHHHHH HHHHHHHHCC CCCCCCCEEE EEEEEECCCH HHHHHHHHHC CCCCCCEEEC CCCCCCCCCC
                     α8          β8          α9          β9
α7b of the luciferase β-subunit
LucbSec   .......... .....HHHHH HHHHHCE.EE CHHHHHHHHH HHHHHHCCCE ..EEEECCCCCCHHHHHHHHH
Lucb      .......... .....EQKMG ELLSENA.IG TYEESTQAAR VAIECCGAAD ..LLMSFESM EDKAQQRAVI
Orf2      PGIIADAGLD IDPSQVDMDF LAEHVWLCGS PETVKGKIER MMERSGGCGQ IVVCSHDNID NPEPYFESLQ
DSC       CCCCCCCCCC CCCCCCCCHH HHHEEEECCC CCCCCCCHHH HHHCCCCCCE EEEEECCCCC CCCCCHHHHH
PHD       CCCCCCCCCC CCCCCCHHHH HHHHHHHCCC CHHHHHHHHH HHHHHCCCEE EEEEECCCCC CCHHHHHHHH
Jpred     CCCCCCCCCC CCHHHHHHHH HHCCCCECCC CHHHHHHHHH HHHHCCCCCE EEEECCCCCC CCCCHHHHHH
ConSeq    CCCCCCCCCC CCCCCCHHHH HHHHHHHCCC CHHHHHHHHH HHHHCCCCCE EEEEECCCCC CCCCHHHHHH
                                α10         α11         β10         α12
LucbSec   HHHHHHCC.. ...
Lucb      DVVNANIV.. ...
Orf2      RLASEVLPKV RMG
DSC       HHHHHCCCCC CCC
PHD       HHHHHHCCCC CCC
Jpred     HHHHHHCCCC CCC
ConSeq    HHHHHHCCCC CCC
```

Figure 4.10 Secondary structure predictions of the Orf2 protein. The predictions are made by the DSC, PHD and Jpred method. The figure below the predictions is the secondary structure prediction of the Orf2 protein given by the consensus sequence, combination of the predictions of the DSC, PHD and Jpred prediction method. (C: coil, E: Strand and H: helix). The α_7b-hilix of luciferase-beta-subunit is indicated.

4.4 Analysis of *orf3*

4.4.1 Nucleotide sequence analysis of *orf3*

The nucleotide sequence and deduced amino acid of *orf3* are shown in Figure 4.11. As the result of *orf3* analysis by the ORF finder discussed in the previous chapter, the start codon of *orf2* could possibly start at nucleotide sequence position 2357 (the first ATG), 2423 (the second ATG) or 2600 (the third ATG). However the third ATG codon is most likely true start codon. The third ATG codon is proceeded by putative ribosome binding site 5 base pairs upstream. This potential ribosome-binding site has a nucleotide sequence of AAAGA. This third ATG position will be considered the true start codon for the *orf3* from this point forward (see also Figure 4.11). The promoter sequences of TTAGAC (-35 region) and TATCAT (-10 region) upstream of third ATG codon were also identified. These promoter sequences are similar to that of *E. coli*, which has the promoter sequences of TTGACA (-35 region) and TATAAT (-10 region). Also, the putative sequence of operator (a short inverted repeat), CAACTCTCGGGAGGTTTG is identified. However, it is not yet known whether these putative promoter and operator sequences are a real promoter for the *orf3*.

```
        Stop codon of orf2
1861 - GCTGAGGGAACACCAATTCGGGAGAGGTTGGACCACGCCGCTATGCGGCGTGGCTCCATT - 1920
     -   *
1921 - TCTCTCTGCAATCGGGCATTGCGATCACCTCAGCCGAAACTGGTGGCAGTGGCCTGTGCT - 1980

1981 - GTGCTCCTCGTCGCAATGCTGTGCGGTTCTGACATATAGCATGTTCACCAGCATCGCAGA - 2040

2041 - GAAACCGTCCATGAGCGTCTCTACTCCTGCGTAGCCCTCCCCTCGAACCAACGCACCGCC - 2100

2101 - CTATACCGCACGGAAGGCACATTCCCGTGCCAGCTGAGGATCAGCGGCCACAGCACTAGC - 2160

2161 - TGACAGAAATCTGCCATCAATCGCCGGATCACCGATCCAATTGCGTATCTCGAAGGGGAC - 2220

2221 - GCAAGACAAGGACGGCACCATTACCAGCCGCCGATCTGCATAAGCATCCGCTATGCTGCA - 2280

2281 - CGCATGATCGGCATCGGTCGCACCAATGCCCATCAACTGATTTCAAACCGAAGAAGTCGA - 2340

                    First start codon of the orf3
2341 - GTCCAAATCGACGCGCATGAGCACCTCCAGCCCCCACGATCTCATCAGGCAGCAGCGTGG - 2400
                     M  S  T  S  S  P  H  D  L  I  R  Q  Q  R  G
                          Second start codon of the orf3
2401 - ATCGGCATTAGCTCGTCCTCTCATGATAAAATCCTGCCTGGGACTATCTGGCTCCGGATC - 2460
       S  A  L  A  R  P  L  M  I  K  S  C  L  G  L  S  G  S  G  S
                      -35 region   -10 region promoter
2461 - ATCCCGACGACTCTACAACATTTAGACGCAGCACACTATTATCATGGGCGCATCAACTCTC - 2520
       S  R  R  L  Y  N  I  R  R  S  T  L  L  S  W  A  H  Q  L  S
  short inverted repeat
2521 - GGGAGGTTTGGAGTCGTTTTTCTGCGGGTATTGCAGGCCTTGGCATTTGGCGTCATTCCG - 2580
       G  G  L  E  S  F  F  C  G  Y  C  R  P  W  H  L  A  S  F  R
                RBS     Third start codon of the orf3
2581 - TTCTCAGGCAAAGAAATTTATGACCAATACTGAAAGTACAAGGAAAGTAAAGTCTATCAA - 2640
       S  Q  A  K  K  F  M  T  N  T  E  S  T  R  K  V  K  S  I  K

2641 - GGCTGATGTTGATGCCATGAAGCGCCAAAGCGTTCTAGATGAATCCATCCAGCAATTCTT - 2700
       A  D  V  D  A  M  K  R  Q  S  V  L  D  E  S  I  Q  Q  F  F

2701 - TGATAACGGCTACGAGGCAACATCCCTGGAATCTATCGCAGATGCGCTGGGCGTCACCAA - 2760
       D  N  G  Y  E  A  T  S  L  E  S  I  A  D  A  L  G  V  T  K

2761 - GCAATTCATCTATTCACGCTTTAACAGCAAATCCGAAATACTGGTTTCGATTTGCCGCTC - 2820
       Q  F  I  Y  S  R  F  N  S  K  S  E  I  L  V  S  I  C  R  S

2821 - TGGAGCTAAGGCCGCAGAAAAAGCTGTTGAACTGAGCGAGGAGATAGAAGGCAATGCGGC - 2880
       G  A  K  A  A  E  K  A  V  E  L  S  E  E  I  E  G  N  A  A

2881 - TGTGCGACTGGCTTGTATCCTGCGTTTTTTTGTTCAATTGCAAATCGAACACCGCAGGGA - 2940
       V  R  L  A  C  I  L  R  F  F  V  Q  L  Q  I  E  H  R  R  E

2941 - AGTTGCTATCTATTTTCGTGAATTCAAGAATTTACCCGCCGACGAAGCCCATGCAATTGA - 3000
       V  A  I  Y  F  R  E  F  K  N  L  P  A  D  E  A  H  A  I  D

3001 - TGCGTCCAAACTGCGCTTTCACCGGATGCTGTGCGCCGTCTTGAACGAGGGCAAGGCAGC - 3060
       A  S  K  L  R  F  H  R  M  L  C  A  V  L  N  E  G  K  A  A

3061 - GGGACTGTTCGAATTCGATGACACCTCTCTCGCTGCATCGGCCTTGGGCGGCATGGTTTC - 3120
       G  L  F  E  F  D  D  T  S  L  A  A  S  A  L  G  G  M  V  S

3121 - TTGGCCATTTTTCTGGTTCCAGCCGGAAGGGCGGTGGGTTCCGACCCTGGTAGCCCATCA - 3180
       W  P  F  F  W  F  Q  P  E  G  R  W  V  P  T  L  V  A  H  Q

3181 - ATTTGCGGCACTTGCCCTCAAGACAGTTGGAGTTTCAGACCCGTCGATTGTTGCCGCTGG - 3240
       F  A  A  L  A  L  K  T  V  G  V  S  D  P  S  I  V  A  A  G

3241 - CTGAAAACTGCCCCCGTAATCGAGCTTGAAGTAGACTCCGCCACAAGCTAGGAGAGCATC - 3300
       *  Stop
3301 - CACCCCCCTGCAATCAACCATATTGTTGACTTTCTCACCCGGCATGCTGACTATAGCCTA - 3360
                 start codon of orf4
3361 - CTGCCCCCTACATGCTCACTTCATCACAGCGAACAGTTAAACCAGACGGGGTTGATACAG - 3420
```

Figure 4.11 DNA sequence of the *orf3* and its deduced amino acid. The deduced amino acid sequence is shown in bold below the DNA sequence. Potential promoter sequences (-35 and –10 region) are in bold and underlined. The short inverted repeat is shown and underlined. Also, three potential start codons and stop codon of *orf3* is also shown in bold. The potential ribosome-binding sequence of CAGG is shown in bold and underlined.

4.4.2 G+C content and codon usage of the *orf3*

With regard to G+C content, the *orf3* has a significantly lower G+C content (50.70%) than that of *P. putida* chromosomal (60.7-62.5%) (Mundel, 1966) and that of all known genes on the CAM plasmid. The G+C content of *camR, D, C, A* and *B* is 59.32%, 61.69%, 58.97%, 58.43% and 57.10 % respectively. In the third position of the *orf3* genetic code, it has a G+C content of 52.56%, which is very low. The G+C content in the first and second position of *orf3* genetic codes are 57.67 and 41.86 % respectively (see Table 4.5).

Table 4.5 G+C content of the *orf3*.

Nucleotide	1		2		3		Total	
	Number	%	Number	%	Number	%	Number	%
A	48	22.33	60	27.91	47	21.86	155	24.03
C	41	19.07	56	26.05	62	28.84	159	24.65
G	83	38.60	34	15.81	51	23.72	168	26.05
T	43	20.00	65	30.23	55	25.58	163	25.27
A+C	89	41.40	116	53.95	109	50.70	314	48.68
A+G	131	60.93	94	43.72	98	45.58	323	50.08
A+T	91	42.33	125	58.14	102	47.44	318	49.30
G+C	124	57.67	90	41.86	113	52.56	327	50.70

The preferentially used codons of *orf3* are GGC (Gly), GAA (Glu), GAT (Asp), GTT (Val), GCC (Val), AGG (Arg), AGC (Ser), AAG (Lys), AAC (Asn), ATG (Met), ATC (Ile), ACC (Thr), TGG (Trp), TGC (Cys), TAT (Tyr), TTG (Leu), TTT (Phe), TCT (Ser), CGC (Arg), CAA (Gln), CAT (His), CTG (Leu) and CCG (Pro) (see Table 4.6). Arginine shares equally the AAT and AAC codons in the *orf3*. Unused codons on this gene are AGA (Arg), ACG (Thr), GGT (Gly) and CCT (Pro). The codons that have been used only once are GGG (Gly), GTG (Val),CGA (Arg), AGT (Ser), ACT (Thr), TGT (Cys), CTT (Leu), CTA (Leu), TAC (Tyr), CCA (Pro) and CCC (Pro). Seven amino acids, namely, Asp, Val, Ala, Tyr, Phe, Ser and Gln have A or T nucleotide in the third codon position. This pattern may relate to the low G+C content in the third codon position of genetic codes for the *orf3*.

4.4.3 Amino acid composition of the Orf3 protein

Figure 4.12 shows the amino acid composition of Orf3 protein. The most used amino acids in the Orf3 protein were Ala (A), Lys (L) and Ser (S) with the amino acid composition of 13.6%, 8.4% and 8.4% respectively. The amino acid composition of Orf3 protein has a similar pattern as in *E. coli*; however, there are significant differences in Ser (S), Phe (F) and Cys (C). These three amino acids in the Orf3 protein have amino acid composition of 8.4%, 7.0% and 1.4%, respectively, which are 1.6 to 2.0 times more than the average amino acid composition of the proteins from *E. coli* (Doolittle FR., 1996). The basic, acidic and hydrophobic amino acids for the Orf3 protein are 13.1%, 12.6% and 40.8% respectively. The theoretical prediction for Orf3 molecular weight is 23,656 with a pI value of 5.86.

Table 4.6 Codon usage of the *orf3*.

Amino acid	Codon	Number	/1000	Percentage	Amino acid	Codon	Number	/1000	Percentage
Gly	GGG	1	4.65	9.0%	Trp	TGG	3	13.95	100.0%
Gly	GGA	3	13.95	27.0%	End	TGA	1	4.65	100.0%
Gly	GGT	0	0	0	Cys	TGT	1	4.65	33.0%
Gly	GGC	7	32.56	64.0%	Cys	TGC	2	9.3	67.0%
Glu	GAG	4	18.6	24.0%	End	TAG	0	0	0
Glu	GAA	13	60.47	76.0%	End	TAA	0	0	0
Asp	GAT	7	32.56	70.0%	Tyr	TAT	2	9.3	67.0%
Asp	GAC	3	13.95	30.0%	Tyr	TAC	1	4.65	33.0%
Val	GTG	1	4.65	6.0%	Leu	TTG	3	13.95	17.0%
Val	GTA	2	9.3	13.0%	Leu	TTA	1	4.65	6.0%
Val	GTT	11	51.16	69.0%	Phe	TTT	8	37.21	53.0%
Val	GTC	2	9.3	13.0%	Phe	TTC	7	32.56	47.0%
Ala	GCG	5	23.26	17.0%	Ser	TCG	3	13.95	17.0%
Ala	GCA	7	32.56	24.0%	Ser	TCA	2	9.3	11.0%
Ala	GCT	8	37.21	28.0%	Ser	TCT	5	23.26	28.0%
Ala	GCC	9	41.86	31.0%	Ser	TCC	4	18.6	22.0%
Arg	AGG	2	9.3	17.0%	Arg	CGG	2	9.3	17.0%
Arg	AGA	0	0	0%	Arg	CGA	1	4.65	8.0%
Ser	AGT	1	4.65	6.0%	Arg	CGT	2	9.3	17.0%
Ser	AGC	3	13.95	17.0%	Arg	CGC	5	23.26	42.0%
Lys	AAG	8	37.21	67.0%	Gln	CAG	2	9.3	25.0%
Lys	AAA	4	18.6	33.0%	Gln	CAA	6	27.91	75.0%
Asn	AAT	3	13.95	50.0%	His	CAT	2	9.3	50.0%
Asn	AAC	3	13.95	50.0%	His	CAC	2	9.3	50.0%
Met	ATG	4	18.6	100.0%	Leu	CTG	10	46.51	56.0%
Ile	ATA	2	9.3	17.0%	Leu	CTA	1	4.65	6.0%
Ile	ATT	3	13.95	25.0%	Leu	CTT	1	4.65	6.0%
Ile	ATC	7	32.56	58.0%	Leu	CTC	2	9.3	11.0%
Thr	ACG	0	0	0	Pro	CCG	3	13.5	60.0%
Thr	ACA	3	13.95	38.0%	Pro	CCA	1	4.65	20.0%
Thr	ACT	1	4.65	13.0%	Pro	CCT	0	0	0
Thr	ACC	4	18.6	50.0%	Pro	CCC	1	4.6	20.0%

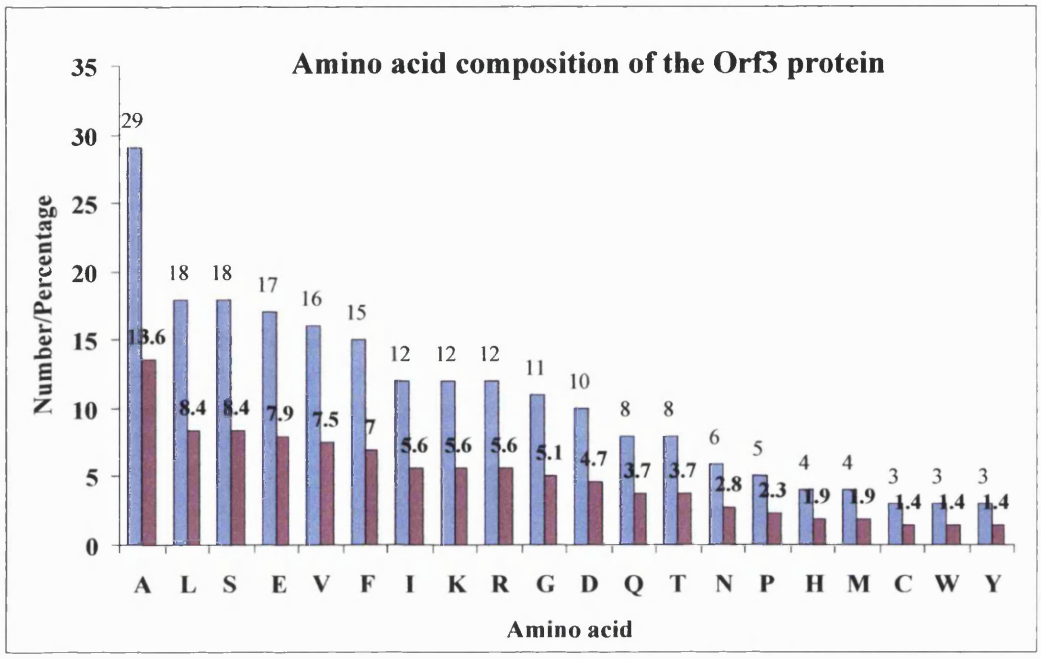


Figure 4.12 Amino acid composition of the Orf3 protein

4.4.4 Conserved domain search of the Orf3 protein

A conserved domain search for Orf3 found that the protein has the conserved domain of TetR, the parent member of the TetR family of regulatory proteins. This TetR-conserved domain consists of 47 amino acid residues. The amino aicd of tetR-conserved domain overlapped with the amino acid of Orf3 protein position 106-158. The alignment between the TetR-conserved domain and that in the Orf3 protein had the score of 46.6 bits and expect value of 2×10^{-6} (see Figure 4.13). This suggests the Orf3 protein is a member of the regulatory proteins in TetR family.

```
CD-Length = 47 residues, 100.0% aligned
Score = 45.0 bits (107), Expect = 2e-06

Orf3 :  25   VLDESIQQFFDNGYEATSLESIADALGVTKQFIYSRFNSKSEILVSI   71
TetR :  1    ILDAALELFAERGYDATTVREIAKEAGVSKGALYRHFPSKEELLAL   47
```

Figure 4.13 Pairwise alignment of the conserved domain of TetR and the conserved domain of the TetR on the Orf3 protein (Orf3: TetR-conserved domain on the deduced amino acid of Orf3 protein; and TetR: TetR conserved domain). The amino acids in red and blue indicate identities and similarities, respectively, between amino acids of the two sequences.

4.4.5 Sequence homology search of the Orf3 protein

The sequence homology search showed the Orf3 protein shared high identity and similarity to a number of regulatory proteins. The three most highly identical proteins to the *orf3* deduced amino acid are a putative transcriptional regulator from *Ralstonia salanacearum*, a transcription regulator from *Mesorhizobium loti* and a transcription regulator (TetR family) from *Deinococcus radiodurans*.

The pairwise alignment of the Orf3 protein and the putative transcriptional regulator (*Ralstonia solanacearum*) had 28% identity and 49% similarity. The alignment of the Orf3 protein to the second highest homologous protein (in the recent search), transcriptional regulator, TetR family (*Mesorhizobium loti*) had 28% identity and 46% similarity. The alignment between the transcriptional regulator, TetR family, from *Deinococcus radiodurans* and the deduced amino acid of *orf3* showed 28% identity and 50% similarity. However, there are a number of proteins that also produced significant alignments with the *orf3* deduced amino acid. Examples of these proteins are a putative regulator protein [*Streptomyces coelicolor* A (3)2], a transcriptional regulator (TetR/AcrR

family)[*Bacillus halodurans*] and other putative transcriptional regulators. Also, in the current search, the Orf3 protein is similar to several known-regulatory proteins. These proteins are a transcriptional repressor (CampR) from *Rhodococcus* sp. NCIMB 9784, a AmtR global repressor protein from *Corynebacterium glutamicum* and a gamma-butyrolactone binding protein (ScbR) from *Streptomyces colelicolor* (see also Table 4.7).

Table 4.7 Selected proteins that share homology to the *orf3* deduced amino acid.

Protein	Mocroorganism	Identity	Similarity	Gap	Score (bits)	Expect value	Accession number
Putative transcriptional regulatory	*Ralstonia solanacearum*	28%	49%	0	91.3	7×10^{-18}	AL646068
Transcriptional regulator	*Mesorhizobium loti*	28%	46%	0	77.0	2×10^{-13}	AP003012
Transcription regulator,TetR family	*Deinococcus rediodurans*	28%	50%	3%	73.2	2×10^{-12}	AE002068
Transcriptional regulator (TetR/AcrR family)	*Bacillus halodurans*	26%	49%	1%	68.6	6×10^{-11}	AP001517
Trascriptional repressor (campR)	*Rhodococcus* sp. NCIMB 9784	26%	47%	2%	61.6	5×10^{-9}	AF323755
AmtR protein	*Corynebacterium glutamicum*	28%	51%	0	46.6	2×10^{-4}	AJ133719
Gamma-butyrolactone binding protein (ScbR)	*Streptomyces coelicolor*	28%	41%	13%	44.7	6×10^{-4}	AL132824

4.4.6 Multiple alignment of the Orf3 protein

Multiple alignment of the deduced amino acid of *orf3* with the highly related proteins; the transcription regulator (*Mesorhizobium loti*) and the known proteins of campR repressor (*Rhodococcus* sp. NCIMB 9784) and AmtR global repressor (*Streptomyces coelicolor*) is shown in Figure 4.14. There is a helix-turn helix DNA binding motif found within the conserved domain of the *orf3* deduced amino acid. The helix-turn-helix prediction (http://pbil.univ-lyon1.fr/) showed 22 amino acid residues on the Orf3 protein had 100% probability of helix-turn-helix DNA binding motif. The protein sequence of this helix turn helix motif is TSLESIADALGVTKQFIYSRFN. The turn in the helix-turn-helix motif is predicted to be at the central G (Gly) residue, which is completely conserved in all three sequences. Also, the S (Ser) residue in the predicted first helix and Y (Tyr) residue in the predicted second helix of HTH-motif is highly conserved among these three sequences (see also Figure 4.14).

```
Orf3      MTNTESTRKVKSIKADVDAMKRQSVLDESIQQFFDNGYEATSLESIADALGVTKQFIYSR 60
PutLoti   MARTTGSDGER---------TEAAVREAAVNLIARYGYEAMSMRQLAAEVGVQAAALYRY 51
CampR     MPRQSRARASAPKKP----TKQERLMAAAVRLFSRQGYAGTSVRDLGEALGIQPGSVYAH 56
AmtRGlob  MAGAVGRPRRSAPRR-AGKNPREEILDASAELFTRQGFATTSTHQIADAVGIRQASLYYH 59
          *.         .  . .  .  :   : . :   *:   * ..:.  :*:    :*

Orf3      FNSKSEILVSICRSGAKAAEKAVELSEEIEGNAAVRLACILRFFVQLQIEHRREVAIYFR 120
PutLoti   FPTKEDLLFTLMREHMEGLRAAWEHVRPIDADPAEQLAAYVRNHIAFHIERRHSTHVSNM 111
CampR     IDSKHTMLVSLVES---GIDQYLDAVAELPGTPTEQLRHFVEAHVRVVAEDVNRARVVYH 113
AmtRGlob  FPSKTEIFLTLLKSTVEPSTVLAEDLSTLDAGPEMRLWAIVASEVRLLLSTKWNVGRLY- 118
          : :*  ::.:: .. .       :    : . .  :*   :   : .  .    .

Orf3      EFKNLPADEAHAIDASKLRFHRMLCAVLNEGKAAGLFEFDDTSLAAS-ALGGMVSWPFFW 179
PutLoti   ELRSLSPDRLTQILRMRTAYEKELRSILREGAEAGDFSIEDTGLTAM-ALIQMMTGVIVW 170
CampR     QWRHIQPPERTRIVAKRYAYEHRLRDIIDAGVAAGEFAPDLDRPTAVRAVLGMVNWCPEW 173
AmtRGlob  QLPIVGSEEFAEYHSQREALTNVFRDLAT--EIVGDDPRAELPFHITMSVIEMRRNDGKI 176
          :   : . .       :    . :  :   .  .*             ::  *

Orf3      FQPEGRWVPTLVAHQFAALALKTVGVSDPSIVAAG----------- 214
PutLoti   FRPGERLSVPEVTATYLSMTMRLVGARMDTYSAARPLDVQRTL--- 213
CampR     LPADGSEPAEAVAAGVAEIVLASVRVGHGAAHA------------- 206
AmtRGlob  PSPLSADSLPETAIMLADASLAVLGAPLPADRVEKTLELIKQADAK 222
            .        .:       :  : .   :  .
```

Figure 4.14 Multiple alignment of the deduced amino acid of *orf3* gene with putative transcriptional regulator from *Mesorhizobium loti* (Putloti), CampR repressor from *Rhodococcus* sp. NCIMB 9784 (CampR) and AmtR global repressor *from Streptomyces coelicolor* (AmtRGlob). The conserved domain of tetR on the *orf3* deduced amino acid is shown in bold. Potential helix-turn-helix motif is shown in bold and underlined.

Nine amino acids are completely conserved in the Orf3 protein, putative transcriptional regulator (*Mesorhizobium loti*), campR repressor (*Rhodoccus* sp. NCIMB 9784) and AtmR global repressor (*Streptomyces coelicolor*). These amino acids are M (1), G (38), S (43), G (52), Y (58), K (64), L (97), G (155) and M (172).

4.4.7 Protein secondary structure prediction of Orf3 amino acid

The predicted protein secondary structure of Orf3 protein is shown in Figure 4.15. The predicted protein secondary structure of Orf3 protein consists of eight helices (α_1-α_8) and two strands (β_1 and β_2). In the helix-turn-helix motif of Orf3 protein, this potential DNA binding domain is given a predicted structure as a helix and strand. In QacR repressor, the first three helices (α_1- α_3) of the protein form a DNA binding domain, in which α_2 and α_3 is a helix-turn-helix motif, DNA binding-domain (Schumacher *et al.*, 2002). Although, the polypeptide of TSLESIADALGVTKQFIYSRFN is predicted as a helix-turn-helix motif by the HTH prediction program, the polypeptide of FIYSR is predicted as a strand by protein secondary prediction programs.

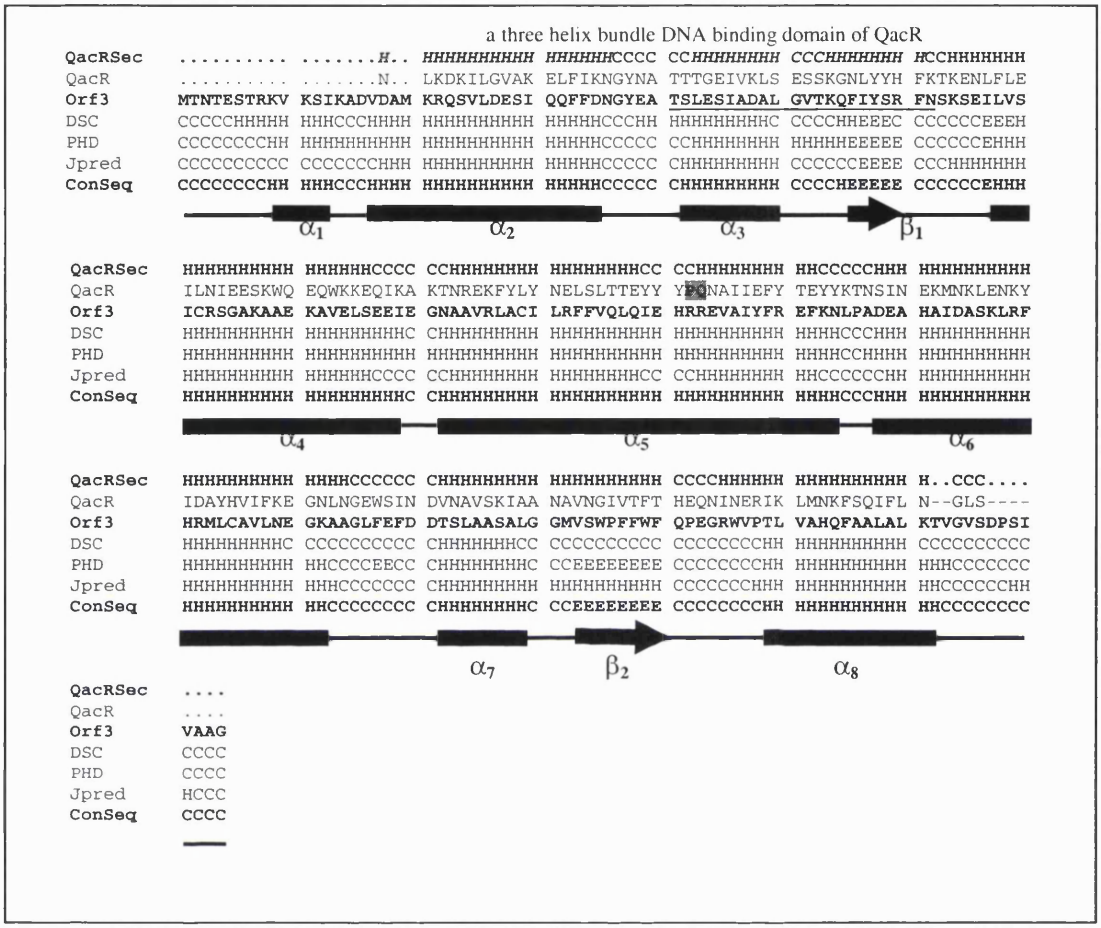


Figure 4.15 Secondary structure of the Orf3 protein. QacRSec: secondary structure of the QacR repressor from *Staphylococcus aureus*; QacR: amino acid of the QacR repressor from *Staphylococcus aureus*; Orf3: deduced amino acid of the *orf3* gene; DSC: predicted secondary structure of the Orf3 protein from the DSC prediction; PHD: predicted secondary structure of the Orf3 protein from the PHD prediction; Jpred: predicted secondary structure of the Orf3 protein from the Jpred prediction; and ConSeq: consensus sequence of the three secondary structure predictions. A grey shading is a deletion made in the alignment. The predicted helix-turn-helix in the Orf3 protein is underlined. The three helices that form a three helix bundle DNA binding domain in QacR repressor is shown in bold and italic.

4.5 Analysis of *orf4*

4.5.1 Nucleotide sequence analysis of *orf4*

The *orf4* gene is 984 basepairs in length and encodes for a protein of 327 amino acids. The start codon (ATG) and stop codon (TGA) of *orf4* are at the nucleotide sequence position 3474 and 4457, respectively, on the *Bam*HI fragment adjacent to the *camR* gene on the CAM plasmid. Potential promoter sequences were found at the 5′ flanking region of the *orf4* 120 basepairs from the start codon of *orf4*. The sequence of –10 region is TATAGC and the sequence of –35 region is TTCTCA. These promoter sequences are very similar to that of *E. coli* promoter (-10 region: TATAAT and –35 region: TTGACA). Nonetheless, a short inverted repeat sequence of TCAACCAT ATTGTTGA, probable operator sequence is identified 2 basepairs upstream of the putative promoter sequences. There is a potential ribosome-binding site at the 5′ flanking region of *orf4*, which has the sequence of GAGAA. A potential transcription termination or inverted repeat sequence has not been found at the 3′ flanking region of *orf4*. However, the effective termination for the *orf4* may not be necessary because *orf4* is transcribed in opposite direction with its neighbouring gene (*camR*) (see Figure 4.16).

```
 ----→ Stop codon of the orf3
3241 - CTGAAAACTGCCCCCGTAATCGAGCTTGAAGTAGACTCCGCCACAAGCTAGGAGAGCATC - 3300
     -  *      short inverted repeat   -35                  -10
3301 - CACCCCCCTGCAATCAACCATATTGTTGACTTTCTCACCCGGCATGCTGACTATAGCCTA - 3360

3361 - CTGCCCCCTACATGCTCACTTCATCACAGCGAACAGTTAAACCAGACGGGGTTGATACAG - 3420
                                                   start codon of the orf4
3421 - CCTGCCTTGAGCAAAAGCTTCAGGCGAAAACAGTTCGTTACAGAGAATCACTCATGCGCA - 3480
                                                            M  R  K

3481 - AGTTCAGATCCTTCGCCTTCCAGCTGACGCTGGTAACAGTCACTGTGGGCTGCGGCATGA - 3540
        F  R  S  F  A  F  Q  L  T  L  V  T  V  T  V  G  C  G  M  N

3541 - ACACCATACCTGCAATCGCTGAGCCTGCCGGCAGGCAACAACATCAAGTGCCCGGATTTT - 3600
        T  I  P  A  I  A  E  P  A  G  R  Q  Q  H  Q  V  P  G  F  Y

3601 - ACCGCATGAACCTGGGTGAGTTTGAAATCACGGCGCTCTATGACGGTTTTATCAAGCTTG - 3660
        R  M  N  L  G  E  F  E  I  T  A  L  Y  D  G  F  I  K  L  D

3661 - ATCCGGCATGGCTCAGCGGCATCAGTGCCGACAACATTCAGAGCCTGCTGGCAAAAATGT - 3720
        P  A  W  L  S  G  I  S  A  D  N  I  Q  S  L  L  A  K  M  F

3721 - TCATCGATTCGAGCAAGGGCATTCAAACCGCAGTGAACGGCTACCTGATCAATACCGGCG - 3780
        I  D  S  S  K  G  I  Q  T  A  V  N  G  Y  L  I  N  T  G  E

3781 - AACACCTGGTGCTGGTCGATGCAGGCTCAGCTCAGTGCTTCGGTTCGACGTTGGGGGTGA - 3840
        H  L  V  L  V  D  A  G  S  A  Q  C  F  G  S  T  L  G  V  M

3841 - TGCGCCGCAACCTAGAAGCGTCCGGTTATCAGGTAGAACAAGTGGATAGCGTGCTGCTGA - 3900
        R  R  N  L  E  A  S  G  Y  Q  V  E  Q  V  D  S  V  L  L  T

3901 - CCCACCTGCATCCGGATCATGCCTGCGGTCTGGCCAATGCCGATGGCACGCCAACCTACC - 3960
        H  L  H  P  D  H  A  C  G  L  A  N  A  D  G  T  P  T  Y  P

3961 - CGAATGCCAGGGTCTACGTGCCGCGCCAGGAAGCCGAATTCTGGCTGGATCAGGATATCG - 4020
        N  A  R  V  Y  V  P  R  Q  E  A  E  F  W  L  D  Q  D  I  A

4021 - CCGCCATGCCGGAACCCAGCCAAGCGTTTTTCCTGATGGCCAGGGCAGCAGTCGCACCCT - 4080
        A  M  P  E  P  S  Q  A  F  F  L  M  A  R  A  A  V  A  P  Y

4081 - ATGCGCAGGGGCGCCTGCTGCGCTATGAGCCTGACGCCGCATTGCTGCCGGGCGTGGAAA - 4140
        A  Q  G  R  L  L  R  Y  E  P  D  A  A  L  L  P  G  V  E  S

4141 - GCGTGCCTACCTACGGGCATACACCTGGTCACTCAGCCTACCTCTTTACCTCTGGCGATG - 4200
        V  P  T  Y  G  H  T  P  G  H  S  A  Y  L  F  T  S  G  D  E

4201 - AACGCCTGATGGTCTGGGGGGATCTGGTGCATAACCATGCCATTCAGTTCGCCCGGCCGG - 4260
        R  L  M  V  W  G  D  L  V  H  N  H  A  I  Q  F  A  R  P  E

4261 - AAGTGGTTATCGAGTTTGACGCAGACTCTGCGCAAGCCAGGAGCTCACGGCAGAGCATGC - 4320
        V  V  I  E  F  D  A  D  S  A  Q  A  R  S  S  R  Q  S  M  L

4321 - TGACGAATGCTGCAAAGGAGCATTTTTGGGTAGCGGGTGCACATCTACCCTTTCCCGGGC - 4380
        T  N  A  A  K  E  H  F  W  V  A  G  A  H  L  P  F  P  G  L

4381 - TTGGCCGTGTTCGCGCGACGGATGGCGCCTACGCCTGGGTACCCATTGAGTTTGGCCCAG - 4440
        G  R  V  R  A  T  D  G  A  Y  A  W  V  P  I  E  F  G  P  V
              Stop codon of the orf4
4441 - TTGGAGACCACCCCTGAAACAAACTCGGCACGAACCCGCCTACTC - 4485
        G  D  H  P  *                          * Stop codon of the camR
```

Figure 4.16 Nucleotide sequence of the *orf4* and its deduced amino acid. The start codon (ATG) and stop codon (TGA) of *orf4* are shown in bold. Potential promoter sequences (-35 and –10 region) and the ribosome binding site are shown in bold and underlined. The short inverted repeat, a probable operator, is indicated and underlined.

4.5.2 G+C content and codon usage of the *orf4*

The G+C content of *orf4* (58.64%) is similar to that of the *camC* (58.97%) and *camA* (58.43%), genes on CAM plasmid. The G+C content of *orf4* in the third position of its triplet codons is 65.55% compared to that of the first and second position, which are 64.94% and 45.43% respectively (see Table 4.8).

Table 4.8 G+C content of the *orf4*.

Nucleotide	1		2		3		Total	
	Number	%	Number	%	Number	%	Number	%
A	67	20.43	87	26.52	48	14.63	202	20.53
C	88	26.83	86	26.22	115	35.06	289	29.37
G	125	38.11	63	19.21	100	30.49	288	29.27
T	48	14.63	92	28.05	65	19.82	205	20.83
A+C	155	47.26	173	52.74	163	49.70	491	49.90
A+G	192	58.54	150	45.73	148	45.12	490	49.80
A+T	115	35.06	179	54.57	113	34.45	407	41.39
G+C	213	64.94	149	45.43	215	65.55	577	58.64

The codon usage of *orf4* is shown in Table 4.9. The prefered codons are GGC (Gly), GAA (Glu), GAT (Asp), GTG (Val), GCC (Ala), AGC (Ser), AAG (Lys), AAC (Asn), ATG (Met), ATC (Ile), ACC (Thr), TGG (Trp), TGC (Cys), TAC (Tyr), TTG (Leu), TTC (Phe), TCA (Ser), CGC (Arg), CAG (Gln), CAT (His) and CTG (Leu). However, for proline, *orf4* uses CCG and CCC codon equally. Unused codons are TGT (Cys), TTA (Leu) and CGA (Arg). The codons used once are AGA (Arg), CGT (Arg), AAA (Lys) and ACT (Thr). Although *orf4* tends to prefer G and C nucleotide in the third position of it genetic codes, an exception is found in Phe and Ser, which prefers to use A or T in the third codon position.

Table 4.9 Codon usage of the *orf4*.

Amino acid	Codon	Number	/1000	Percentage	Amino acid	Codon	Number	/1000	Percentage
Gly	GGG	5	15.24	18.0%	Trp	TGG	5	15.24	100.0%
Gly	GGA	2	6.1	7.0%	End	TGA	1	3.05	100.0%
Gly	GGT	7	21.34	25.0%	Cys	TGT	0	0	0%
Gly	GGC	14	42.68	50.0%	Cys	TGC	3	9.15	100.0%
Glu	GAG	6	18.29	38.0%	End	TAG	0	0	0%
Glu	GAA	10	30.49	63.0%	End	TAA	0	0	0%
Asp	GAT	11	33.54	65.0%	Tyr	TAT	4	12.2	36.0%
Asp	GAC	6	18.29	35.0%	Tyr	TAC	7	21.34	64.0%
Val	GTG	12	36.59	50.0%	Leu	TTG	2	6.1	7.0%
Val	GTA	4	12.2	17.0%	Leu	TTA	0	0	0%
Val	GTT	3	9.15	13.0%	Phe	TTT	9	27.44	53.0%
Val	GTC	5	15.24	21.0%	Phe	TTC	8	24.39	47.0%
Ala	GCG	7	21.34	18.0%	Ser	TCG	2	6.1	11.0%
Ala	GCA	12	36.59	30.0%	Ser	TCA	3	9.15	17.0%
Ala	GCT	3	9.15	8.0%	Ser	TCT	2	6.1	11.0%
Ala	GCC	18	54.88	45.0%	Ser	TCC	2	6.1	11.0%
Arg	AGG	4	12.2	24.0%	Arg	CGG	2	6.1	12.0%
Arg	AGA	1	3.05	6.0%	Arg	CGA	0	0	0%
Ser	AGT	1	3.05	6.0%	Arg	CGT	1	3.05	6.0%
Ser	AGC	8	24.39	44.0%	Arg	CGC	9	27.44	53.0%
Lys	AAG	4	12.2	80.0%	Gln	CAG	9	27.44	56.0%
Lys	AAA	1	3.05	20.0%	Gln	CAA	7	21.34	44.0%
Asn	AAT	4	12.2	40.0%	His	CAT	8	24.39	67.0%
Asn	AAC	6	18.29	60.0%	His	CAC	4	12.2	33.0%
Met	ATG	9	27.44	100.0%	Leu	CTG	20	60.98	69.0%
Ile	ATA	1	3.05	8.05	Leu	CTA	2	6.1	7.0%
Ile	ATT	4	12.2	31.0%	Leu	CTT	2	6.1	7.0%
Ile	ATC	8	24.39	62.0%	Leu	CTC	3	9.15	10.0%
Thr	ACG	6	18.29	38.05	Pro	CCG	7	21.34	33.0%
Thr	ACA	2	6.1	13.0%	Pro	CCA	2	6.1	10.0%
Thr	ACT	1	3.05	6.0%	Pro	CCT	5	15.24	24.0%
Thr	ACC	7	21.34	44.0%	Pro	CCC	7	21.34	33.0%

4.5.3 Amino acid composition of the Orf4 protein

The amino acid composition of Orf4 protein is shown in Figure 4.17. The three most highly used amino acids by the Orf4 protein are Ala (A), Leu (L) and Gly (G) with 12.2%, 8.9 % and 8.6% respectively. Most of the amino acids have similar distribution to the average amino acid compositions of proteins from *E. coli.* However, for Pro (P), Phe (F) and Thr (T), these amino acids are more often used (1.5 time plus) in the ORF4 protein compared to the proteins from *E. coli.* The amino acids in the basic (K, R and H), acidic (D and E) and hydrophobic group (A, V, I, L, M, F, Y and W) are 13.6%, 12.2% and 40.0% respectively. The Orf4 protein has a theoretical molecular weight of 35,688 and pI value of 5.49.

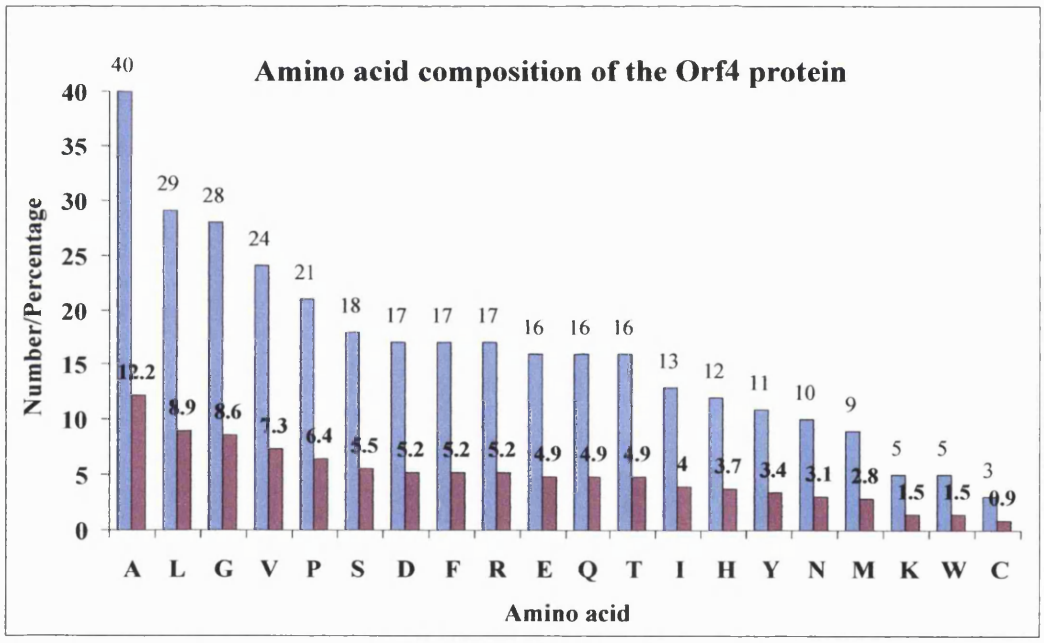


Figure 4.17 Amino acid composition of the Orf4 protein.

4.5.4 Conserved domain search of the Orf4 protein

The conserved domain of the metallo-β-lactamase family was found in the Orf4 protein. The pairwise alignment of the overlapping region between the conserved domain of metallo-β-lactamase and that in Orf4 protein had the score of 52.2 bits, and expect value of 3×10^{-8}. The metallo-β-lactamase consists of 180 residues. However, only 55 amino acids (position 6-140) overlap well with *orf4* amino acid position 92-252 (see also Figure 4.18). The result indicates that the Orf4 protein is highly related to the proteins in metallo-β-lactamase superfamily.

```
CD-Length = 180 residues, only  75.0% aligned
Score = 52.2 bits (124), Expect = 3e-08

Orf4 :  92   TAVNGYLINTGEHLVLVDAGSAQCFGSTLGVMRRNLEASGYQVEQVDSVLLTHLHPDHAC  151
Lact :  6    VDSNAYLVEDDDGAALIDTGT------TAPAAKALLRLLKDGGKKIDAIILTHAHADHIG  59

Orf4 :  152  GLANADGTPTYPNARVYVPRQEAEFWLDQDIAAMPEPSQAFFLMARAAVAPYAQGRLLRY  211
Lact :  60   GVPELLE--RTGAAEVYAGYAPDRLKELLKGE------------EDGPDEELKDGDELRV  105

Orf4 :  212  EPDAALLPGVESVPTYGHTPGHSAYLFTSGDERLMVWGDLV  252
Lact :  106  GDG----KELEVIHTPGHTPGSIVYYLP--EEKVLFTGDLL  140
```

Figure 4.18 Pairwise alignment of the conserved domain of metallo-β-lactamase superfamily and that in the Orf4 protein. Orf4: the conserved domain in the Orf4 protein; and Lact: the conserved domain in the metallo-β-lactamase family. The identical and similarities residues in both sequences are highlighted in red and blue respectively.

4.5.5 Sequence homology search of the Orf4 protein

The deduced amino acid sequence of *orf4* shares homology with several known and unknown proteins on the recent GenBank database. A BLAST search revealed that the deduced amino acid of *orf4* shared identity with known proteins of a methyl parathion degrading protein (*Plesiomonas* sp. DLL-1) and methyl parathion hydrolase (*Plesiomonas* sp. M6). Sequence alignment between the methyl parathion hydrolase and the deduced amino acid of *orf4* had 40% identity and 58% similarity. The alignment of the methyl parathion degrading protein with the Orf4 protein had 39% identity and 58% similarity. Moreover, the alignment of the Orf4 protein to the putative metallo-β-lactamase from *Agrobacterium tumefaciens* had 40% identity and 59% similarity. Also, the amino acid sequence of Orf4 had significant sequence similarity to a number hypothetical and known-proteins. Such as a hypothetical protein *from Ralstonia solanacearum*, hypothetical protein from *Sinorhizobium meliloti*, metallo-β-lactamase superfamily protein from *Agrobacteruim tumefaciens*, a putative β-lactamase from *Streptomyces coelicolor*, glyoxylase II protein from *Neisseria meningitidis*, putative arylsulfatase from *Pyrococcus furiosus* and teichoic acid phosphorylcholine from *Streptococcus pneumoniae* R6 (see Table 4.10).

Table 4.10 Selected proteins that highly share homology with the Orf4 protein.

Protein	Mocroorganism	Identity	Similarity	Gaps	Score (bits)	Expect value	Accession number
Conserved hypothetical protein	*Ralstonia solanacearum*	54%	66%	1%	323	5×10^{-87}	AL646068
Hypothetical protein	*Sinorhizobium meliloti*	41%	58%	2%	221	1×10^{-56}	AE007286
Methyl parathion hydrolase	*Plesiomonas* sp. M6	40%	58%	2%	216	1×10^{-55}	AF338729
Methyl parathion degrading protein	*Plesiomonas* sp.DLL-1	39%	58%	2%	241	1×10^{-54}	AY029773
Metallo-beta-lactamase	*Agrobacterium tumefaciens* str.(58)	40%	59%	2%	214	1×10^{-54}	AE008933
Metallo-betalactamase superfamily protein	*Agrobacterium tumefaciens* str.(58)	30%	47%	3%	115	9×10^{-25}	AE009030
Putative beta-lactamase	*Streptomyces coelicolor*	29%	44%	9%	83.2	4×10^{-15}	AL590942
Glyoxylase II protein	*Neisseria meningitidis*	27%	41%	16%	35.0	1.1	AE002555
Putative arylsulfatase	*Pyrococcus furiosusi*	30%	40%	15%	34.7	1.7	AE010239
Teichoic acid phosphoryl-choline	*Streptococcus pneumoniae* R6	41%	56%	13%	33.1	4.2	AE008458

4.5.6 Multiple alignment of the Orf4 protein

The multiple alignment of the Orf4 protein with methyl parathion hydrolase (*Plesiomonas* sp. M6), the methyl parathion degrading protein (*Plesiomonas* sp. DLL-1) and putative β-lactamase (*Agrobacterium tumefaciens* str. (58)) is shown in Figure 19. The highly conserved domain, zinc-binding motif, of metallo hydrolase and β-lactamase proteins was also identified at the amino acid position 144-149 on the Orf4 protein. This

amino acid sequence is HLHPDH which agrees with the consensus sequence of zinc-binding motif of HxHxDH (x is any amino acid) (Dong *et al.*, 1999; and Wang and Pabo, 1999).

```
MphM6          ----------MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRM 50
MpdDLL1        MPTHKDKEITMPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPEVRTSAPGYYRM 60
LactaB         ----------MPSRRSILK--SALTGLVAA-----PFVAPSLTFAKAPYAVVQAPGFYRL 43
Orf4           ----------MRKFRSFAFQLTLVTVTVGCG-----MNTIPAIAEPAGRQQHQVPGFYRM 45
                         *   . :    : ::  *..      :   .     *     ..**:**:

MphM6          LLGDFEITALSDGTVALPVDKR-LNQPAPKTQSALAKSFQKAP--LETSVTGYLVNTGSK 107
MpdDLL1        LLGDFEITALSDGTVALPVDKR-LNQPAPKTQSALAKSFQKAP--LETSVTGYLVNTGSK 117
LactaB         KIGSVEVTALSDGTIPLPLSKLYTNTSQQHAQSVLSDAFLPAS--VPTSVNAFLVNTGDR 101
Orf4           NLGEFEITALYDGFIKLDPAWL-SGISADNIQSLLAKMFIDSSKGIQ**TAVNGYLINTGEH** 104
                **..*:*** ** : *       . .  : ** *:. *  :...: *:*..:*:***.:

MphM6          LVLVDTGAAGLFGPTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGE-QLAFPN 166
MpdDLL1        LVLVDTGAAGLFGPTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGE-QLAFPN 176
LactaB         LVLIDAGTGTYIGPSLGKLAANIEASGYKVDDIDDVVLTHIHTDHSGGLVSNG-KRSFPN 160
Orf4           **LVLVDAGSAQCFGSTLGVMRRNLEASGYQVEQVDSVLLTHLHPDHACGLANADGTPTYPN** 164
               ***:*:*:.  :*.:** :  *::*:**: :::*.: :***:*.**  **    .  ::**

MphM6          AVVRADQKEADFWLSQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGI 226
MpdDLL1        AVVRADQKEADFWLSQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGI 236
LactaB         ATLRVNEREAKFWLSADNANAATG-IVKQQFAEADQCVTPYVKAEKFETFADNAAPVPGL 219
Orf4           **ARVYVPRQEAEFWLDQDIAAMPEP--SQAFFLMARAAVAPYAQG-RLLRYEPDAALLPGV** 221
               * : . .:**.***.      .     :  *  *  .: **.:. ::  :  ::  :**:

MphM6          KALASHGHTPGHTTYVVESQGQKLALLGDLILVAAVQFDDPSVTNQLDIDGKSAAVERKK 286
MpdDLL1        KALASHGHTPGHTTYVVESQGQKLALLGDLILVAAVQFDDPSVTTQLDSDSKSVAVERKK 296
LactaB         GSILYAGHTPGHSAITLESEGQKIVFWGDITHGDILQFDEPGVAIEFDIDQKAAVAARDI 279
Orf4           **ESVPTYGHTPGHSAYLFTSGDERLMVWGDLV**HNHAIQFARPEVVIEFDADSAQARSSRQS 281
                ::   ******::  . * .::: . **:     :**  * *. ::* *   .   *.

MphM6          AFADAAKGGYLIAASHLPFPGIGHIRAEGKGYRFVPVNYSVVNPK- 331
MpdDLL1        AFADAAKGGYLIAASHLSFPGIGHIRAEGKGYRFVPVNYSVVNPK- 341
LactaB         AFKQAVEGRYLVGGAHIAFPGIGHVRKDSTNYDWLPINYA------ 319
Orf4           MLTNAAKEHFWVAGAHLPFPGLGRVRATDGAYAWVPIEFGPVGDHP 327
                : :*.:  : :..:*:.***:*::*  .  * ::*:::.  .  .
```

Figure 4.19 Multiple alignment of the Orf4 protein (Orf4), methyl parathion degrading protein (*Plesiomonas* sp. DLL-1) (MpdDLL1), methyl parathion hydrolase (*Plesiomonas* sp. M6) (MphM6), and metallo-β-lactamase (*Agrobacterium tumefaciens* str. (58)) (LactaB). The conserved domain of proteins in the metallo-β-lactamse family on the *orf4* deduced amino acid is shown in bold. The zinc-binding motif is indicated in the square.

Eighty-five amino acids are highly conserved among the methyl parathion hydrolase (*Plesiomonas* sp. M6), methyl parathion degrading protein (*Plesiomonas* sp. DLL-1), metallo-β-lactamase (*Agrobacterium tumefaciens*) and the Orf4 protein. These amino acids are shown in the multiple alignment (Figure 4.19) indicated by asterisks.

4.5.7 Protein secondary structure prediction of Orf4 amino acid

A predicted secondary structure of Orf4 protein consists of thirteen helices (α_1-α_{13}) and twelve strands (β_1-β_{12}) (see Figure 4.20). The structure of L1 metlo-β-lactamase from *Stenotrophomonas maltophilia* (Ullah *et al.*, 1998) (16% identity to the Orf4 protein), can be compared with the Orf4 protein. The L1 metallo-beta-lactamase is α/β barrel structure (Ullah JH. et al., 1998). The α/β barrel structure has parallel β-strands arranged similar to the staves of a barrel surrounded by α-helices on the outside of the barrel. The potential Zn^{2+} binding-motif (HLHPDH sequence), probable central active site, of the Orf4 protein is situated in the loop region between the β_3-strand and α_8-helice of the protein. However, this loop region is much wider in related to that of the L1 metallo-beta-lactamase. There is a predicted structure where the predictions are in disagreement. This region is the β_3 strand, in which DSC, PHD and Jpred predicted differently as a helix, strand and loop region respectively.

```
SmlaSec   .......... .......... .......... .......... .......... ........CC CCCCCCCCCC
Smla      .......... .......... .......... .......... .......... ........EV PLPQLRAYTV
Orf4      MRKFRSFAFQ LTLVTVTVGC GMNTIPAIAE PAGRQQHQVP GFYRMNLGEF EITALYDGFI KLDPAWLSGI
DSC       HHHHHHHCCE EEEEEEEECC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC EEEHHHHCCC CCCCCCCCCC
PHD       HHHHHHHHEE EEEEEEEEEC CCCCCCCCCC CCCCCCCCCC CHHHCCCCCC CHHHHHCCCC CCCCHHHHCC
Jpred     CCCCCCCCCC EEEEEEEECC CCCCCCCCCC CCCCCCCCCC CHHHHCCCCH HHHHHHCCCC CCCCCEEEEE
ConSeq    HHHHHHHHCE EEEEEEEECC CCCCCCCCCC CCCCCCCCCC CHHHCCCCCC CHHHHHCCCC CCCCCHHHCC
          α1         β1                                α2         α3         α4

SmlaSec   CHHHHCCCCC EEEECCEEEC CCCCCCEEEE ECCEEEEECC CC....CCHH HHHHHHHHHH CCCHHHEEEE
Smla      DASWLQPMAP LQIADHTWQI GTDLTALLVQ TPDGAVLLDG GM....PQMA SHLLDNMKAR GVTPRDLRLI
Orf4      SADNIQSLLA KMFIDSSKGI QTAVNGYLIN TGEHLVLVDA GSAQCFGSTL GVMRRNLEAS GYQVEQVDSV
DSC       CCCHHHHHHH HHEECCCCCC EEEEEEEECC CCCCEEEECC CCCCCCCCCC CHHHHHHHCC CCCHHHHHHE
PHD       HHHHHHHHHH HHHHCCCHHH HHHHCCCCCC ECEEEEEECC CCCCCCCCCH HHHHHHHHHC CCCEEEECEE
Jpred     CCCCCHHHCC CCCCCCCCCC HHHHCCEEEE CCCCEEEEEC CCCCCHHHHH HHHHHHHHHH HCCCCCCEEE
ConSeq    CCCHHHHHHH HHHHCCCCCC HHHHCCCCCC CCCCEEEECC CCCCCCCCCH HHHHHHHHHC CCCEEEECEE
          α5         α6         β2         α7         β3

SmlaSec   ECCCCCCCCC CHHHHHHHH. ..CCCEEEEC .HHHHHHH.. .....HHCCC CCCCCCCCCC CCCCCCCEEC
Smla      LLSHAHADHA GPVAELKRR. ..TGAKVAAN .AESAVLL.. .....ARGGS DDLHFGDGIT YPPANADRIV
Orf4      LLTHLHPDHA CGLANADGTP TYPNARVYVP RQEAEFWLDQ DIAAMPEPSQ AFFLMARAAV APYAQGRLLR
DSC       EEECCCCCCC CCCCCCCCCC CCCCCEEECC CHHHHHHHHH HHHCCCCCHH HHHHHHHHHC CHHHHHHHHH
PHD       EEEECCCCCC CCCCCCCCCC CCCCHHHHCC HHHHHHHHHH HHHHCCCCCC CHHHEEECCC CCCCCCEEEC
Jpred     EECCCCCCCC CCCEECCCEE EECCCCCCCC CCCHHHHHHC CCCCCCCCCC CCEECCCCCC CCCCCCCCCC
ConSeq    EEECCCCCCC CCCCCCCCCC CCCCCHHHCC CHHHHHHHHH HHHCCCCCCC CHHHEEECCC CCCCCCEEEC
                                α8         α9         α10 β4     β5

SmlaSec   CCCEEEEECC EEEEEECCCC CCCCEEEEEE EECCEEEEC. CCCCCCCCCC CCCCCC.... ..HHHHHHHH
Smla      MDGEVITVGG IFTAHFMAGH TPGSTATWTD TRNGRIAYA. DSLSAPGYQL GNPRYP.... ..HLIEDYRR
Orf4      YEPDAALLPG VESVPTYGHT PGHSAYLFTS GDERLMVWGD LVHNHAIQFA RPEVVIEFDA DSAQARSSRQ
DSC       HCCCCCCCCC CCCCCCCCCC CCCEEEEECC CCCEEEECCC CCHHHHHHHC CCCEEEEECC CCCHHHHHHH
PHD       CCCCCCCCCC EEEEECCCCC CHHHEEEECC CCEEEEEECC HHHHHHHHHC CCEEEEEECC CHHHHCCCHH
Jpred     CCCCCCCCCC CEEEECCCCC CCCEEEEEEC CCCCEEECCC CCCCCCCCCC CCCCCCCCCC CHHHEEEEEH
ConSeq    CCCCCCCCCC CEEEECCCCC CCCEEEEECC CCCEEEECCC CHHHHHHHHC CCCEEEEECC CHHHHCCCHH
                     β6         β7         β8         α11        β9         α12

SmlaSec   HHHHHHHCCC CEEECCCCCC CHHHHC... .......... ........
Smla      SFATVRALPC DVLLTAGAKA LKETAG.... .......... ........
Orf4      SMLTNAAKEH FWVAGAHLPF PGLGRVRATD GAYAWVPIEF GPVGDHP
DSC       HHHHHHHHHH EEEECCCCCC CCCCEEEECC CCCEEEEEEC CCCCCCC
PHD       HHHHHHHHHH HHHHHCCCCC CCCCEEEECC CCCEEEEEEC CCCCCCC
Jpred     HHHHHHHCCC CEEECCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCC
ConSeq    HHHHHHHHHH HEEECCCCCC CCCCEEEECC CCCEEEEEEC CCCCCCC
          α13        β10        β11        β12
```

Figure 4.20 Predicted secondary structure of the Orf4 protein (H: helice; E: extended or strands; and C: coil or loop). The Zn^{2+} binding-motif, probable central active site, of the Orf4 protein is shown as a box. SmlaSec: the L1-metallo-beta-lactamase secondary structure; Smla: the L1-metallo-beta-lactamase sequence; Orf4: the Orf4 protein sequence; DSC, PHD and Jpred are the secondary structure of Orf4 protein predicted by the DSC, PHD and Jpred method respectively; and ConSec: the consensus sequence of the protein secondary structure of Orf4 protein, the combination of DSC, PHD and Jpred prediction.

4.6 Analysis of *orf5*, *orf6*, *orf7* and *orf8*

The analysis of the DNA sequence of *orf5*, *orf6*, *orf7* and *orf8* found on the 4201-bp *Kpn*I DNA fragment is shown in Figure 4.21. The *orf5* consists of 1272 nucleotides and encodes a protein of 423 amino acids. It has a start codon of ATG proceeded by the putative ribosome binding site of GGAAA, and the stop codon of TGA. The *orf6* is a gene of 1098 nucleotides, which encodes a protein of 365 amino acids. It has the start codon of ATG proceeded by the putative ribosome binding site of GAAG, 6 basepairs upstream of the start codon, and the stop codon of TGA. The *orf7* (618 nucleotides in length) , transcribed in diverted direction to the *orf6*, has the start codon of ATG located 6 base pairs after the putative ribosome binding site of GGAG, and the stop codon of TGA proceeded by the inverted repeat sequence of GCCCCGGCAACTCATCGTTGCCGGGGGC. This gene encodes a protein of 205 amino acids. The *orf8* consists of 216 nucleotides and encodes a protein of 71 amino acids. It has the start codon of ATG proceeded by the putative ribosome binding site of AGGA, and the stop codon of TAA. These four ORFs use ATG as an initiation codon, and seem to prefer TGA as a termination codon except the *orf8*, which used TAA instead.

```
  1 -GGTACCGCAGGCGTTCGCCCATGAGCTGGCCTTCTCGAAGAAAGAAAACATCAAGGTCGA- 60

 61 -AGTACCTGGGGACGCCACCACCTGGTGCAAACCCGAGGTCGAGCTGACCATCACCCGTCC- 120

121 -GGCCTGGGACAACCAGGAGCTGCTATCGGGCTTGCTGACCAAGCTGCCGTTCGTGTTCGC- 180

181 -CAAGGACTGCTCGACCGCCAAGGTGAGCTGGAAAGCCGTGGACGCCAAAGGCAACCTCTA- 240

241 -CGCCAGCGGTTCGGGCAATGCCAGCAACCTGGGCCTGGTGACCCTGGCCGCCGCGCCTGC- 300

301 -CGCTGTCGCTGCTGTTCCGGCACCCGCTCCGTCTGCGCCAGCCCCGGCACCTGCACCGGC- 360

361 -AGAAGCTCCGGCGGCTGTTGCTCCAGCCCCAGCGCCTGAGCCAGCTCCTGCGGTAGTCGA- 420

421 -ATCGACACCTGCGAAAGTAGAAGCAGCCCCTGCTGCAGTGCCAGCTCCAGCCGTTGCCGC- 480

481 -AGAGCCCGCACCAGCACCAGCAGCCGAAACCCCCGCCGCAGCGCCTGCCGCTCCGCCCGT- 540

541 -CGCAGCCCCTGCCGCCGCAGTAGCCGCAGCCCCCACCTCGGATTTCGGTCGCGCCGTGGT- 600
         RBS      Start codon of orf5---->
601 -CCTGGAAAACCGCAACCTGATGCAGGTAACGGACGGTTCTGGCTGCAAATGGGTACTCAG- 660
                         M  Q  V  T  D  G  S  G  C  K  W  V  L  S

661 -CCGCGAGATCATCAGCAACGGTGACACCCTGTCGTTCGGCACCACCCCGGCCATGCCGTG- 720
      R  E  I  I  S  N  G  D  T  L  S  F  G  T  T  P  A  M  P  C
```

```
 721 -CCCGGCGTCCGGTTTTGGCGAAGGCAACTTCGAAAAAATCAGCTGGAAGGCCGTGGGCAC- 780
       P  A  S  G  F  G  E  G  N  F  E  K  I  S  W  K  A  V  G  T

 781 -CTACCGTGGGGATAACTGGAGCCGCGTATACGTGCACCCGAGCGGCCTGATTTTCAACAA- 840
       Y  R  G  D  N  W  S  R  V  Y  V  H  P  S  G  L  I  F  N  K

 841 -GGTCTACGAGCCTGCGGTCAAGGACAAGGCCGTTTCCTACCTCACGGCAGACGCTGGCCA- 900
       V  Y  E  P  A  V  K  D  K  A  V  S  Y  L  T  A  D  A  G  Q

 901 -GGCCACATTCCTGGTGGGCGAGATCCCCAGCCGGCAGATGAAGGTGTACCTGGCCTTCAC- 960
       A  T  F  L  V  G  E  I  P  S  R  Q  M  K  V  Y  L  A  F  T

 961 -CCGTGGCAGCTACGGCGTGCTCCGTCCGTTCGGCAGCGACCCTTACTACGTGGCAGTAAC- 1020
       R  G  S  Y  G  V  L  R  P  F  G  S  D  P  Y  Y  V  A  V  T

1021 -CCCGGACGAGTCGTTCGCTCTGGACGCAGCCAAGTACAAGGAAGCTGCGCTGGAAATCTT- 1080
       P  D  E  S  F  A  L  D  A  A  K  Y  K  E  A  A  L  E  I  F

1081 -TGACCTGATCAAGACCACCTCGCCGACCACCACCGACGTGGCCGACCTGCTCATCGTCAA- 1140
       D  L  I  K  T  T  S  P  T  T  T  D  V  A  D  L  L  I  V  K

1141 -AGACATTTCGGCGATCACGAACAACATGTGGGGCAACGACGCGCAGAAGATCACCCGCAA- 1200
       D  I  S  A  I  T  N  N  M  W  G  N  D  A  Q  K  I  T  R  N

1201 -CCGCATCGGCATCAATCGCCAAGGCCTGTTCTTCGATGTGCGCGATGGCGCGAACTGGGG- 1260
       R  I  G  I  N  R  Q  G  L  F  F  D  V  R  D  G  A  N  W  G

1261 -CGGTTCAAGCGTGAGGAGCAGCGCGTGCGTGAGGCGCGCCGGCGTCAACAGGAACTGGGC- 1320
       G  S  S  V  R  S  S  A  C  V  R  R  A  G  V  N  R  N  W  A

1321 -GGTTCAGCGTGAGGAGCAGCGCGTGCGTGAGGCGCGCCGGCGTCAACAGGAACTGGCCAG- 1380
       V  Q  R  E  E  Q  R  V  R  E  A  R  R  R  Q  Q  E  L  A  S

1381 -CGTGCACACCCGCGTACTGGAGCGTTACCAGCAGTTGCAAGACGGCATGAGCGAGTTCAA- 1440
       V  H  T  R  V  L  E  R  Y  Q  Q  L  Q  D  G  M  S  E  F  K

1441 -AGGGCGCGAAACCGAAGCCCTGGCGCAAATGGCCGGCATCAAGGTGCGCTTCGCTTCGCC- 1500
       G  R  E  T  E  A  L  A  Q  M  A  G  I  K  V  R  F  A  S  P

1501 -GCTGGAACAGCAAAACCCGGCAACTTCGGCCCGCGTGGTGCCGATGATGGTCCACGTCAC- 1560
       L  E  Q  Q  N  P  A  T  S  A  R  V  V  P  M  M  V  H  V  T

1561 -CGGCAAGCAGGGCGATTTCTACGCCATCGACTTCCCGAGCAAGGGCCGCCTGGTGGCGGA- 1620
       G  K  Q  G  D  F  Y  A  I  D  F  P  S  K  G  R  L  V  A  D

1621 -CGAGGAGTACAGCGAAGGCTGGTACGTGGCGCAAGTGGCCAACGCCACGCCGTACTACCC- 1680
       E  E  Y  S  E  G  W  Y  V  A  Q  V  A  N  A  T  P  Y  Y  P

1681 -GCTGGACGATGGCCGCGCGGTGCCCACATACCGTGCCTATAACGCGGGCGAGCCCCAGGC- 1740
       L  D  D  G  R  A  V  P  T  Y  R  A  Y  N  A  G  E  P  Q  A

1741 -ATGCAAACAGGACAAGTGTGCGGACCTCGTGTCGTTCGGCGCCGTGCTGGCCAAGGAATT- 1800
       C  K  Q  D  K  C  A  D  L  V  S  F  G  A  V  L  A  K  E  F

1801 -CCCTAACGCCGGTATCGATTTCAGCTGGACCCCCGAGGTCTCCCAAAAGTACGTGAACGA- 1860
       P  N  A  G  I  D  F  S  W  T  P  E  V  S  Q  K  Y  V  N  D
                            -----> Stop codon of orf5
1861 -CTGGAACAACGCTTCCGCCATGGTCCAGTGAGGCAAACACCATGATCAAACAGATGAACA- 1920
       W  N  N  A  S  A  M  V  Q  *
```

```
1921 -AGAAGCAGCTGCTGATTGGCGGTGTGGCCGCAGCGGTGTTGGTGCTCGGCGGTGGTTACG- 1980
                              RBS     Start codon of orf6 ---->
1981 -CGACTGCTGTCCAGCTCGGGCAAGAAGGTTGCGATGGCCTACTACGACGACTTCAAGGAG- 2040
                                       M  A  Y  Y  D  D  F  K  E

2041 -CGCTACTTCCTTGAGGACGTGCTGTCTGAAGGTGATATTTCCTACTCGGCGTTCTCTGGC- 2100
      R  Y  F  L  E  D  V  L  S  E  G  D  I  S  Y  S  A  F  S  G

2101 -AACCTGACTGTCGCCGACCCCGAGATCCGCGTGGCGGCCGCGCAGACCAACGGTGCCCAG- 2160
      N  L  T  V  A  D  P  E  I  R  V  A  A  A  Q  T  N  G  A  Q

2161 -CAATTCATGCGTGGCCTGTCGGGCCTGATGGAGTCGGTTCGCGGTAGCTCGGCGGAGGAA- 2220
      Q  F  M  R  G  L  S  G  L  M  E  S  V  R  G  S  S  A  E  E

2221 -GGCCTGGCCGGTTGGGCCAAGTACCAACTCAACGCCAGCGGCAATGTTGCAGGCGTCTAC- 2280
      G  L  A  G  W  A  K  Y  Q  L  N  A  S  G  N  V  A  G  V  Y

2281 -CTGAAGGCCGATGCGCTGAAGCTGGCGCACGATGGTGACAGCAAGGACGGCACGATCCAC- 2340
      L  K  A  D  A  L  K  L  A  H  D  G  D  S  K  D  G  T  I  H

2341 -ATCCAGCTGCTTGGGATGCAGATGGGCAACCCTTTCGTTGCCAGCAAGGGCGCTGAAGTG- 2400
      I  Q  L  L  G  M  Q  M  G  N  P  F  V  A  S  K  G  A  E  V

2401 -GTGCTGGTGTCTGATGTCAGCGACGAGATCCAGCCGCGTGCCGAACTGGCCGCCAATGGC- 2460
      V  L  V  S  D  V  S  D  E  I  Q  P  R  A  E  L  A  A  N  G

2461 -CGGGCCACCCAGTCGAACTTCGACTGGGGTAGTAACATGGTTGTGCGTCAGCCAATGACC- 2520
      R  A  T  Q  S  N  F  D  W  G  S  N  M  V  V  R  Q  P  M  T

2521 -GGCGCGTTTTTGGTCAGTAGCACGGGGGAATTCGGCACCACGGTAGACGTGGACTTCACC- 2580
      G  A  F  L  V  S  S  T  G  E  F  G  T  T  V  D  V  D  F  T

2581 -CTCAAGCGCTCCAGTGACGGCGAGGGCTCGATGGCGTTCGTCGTGACCCACCGTAACGAT- 2640
      L  K  R  S  S  D  G  E  G  S  M  A  F  V  V  T  H  R  N  D

2641 -GGCTCCAAGGTGGGTGAGATCGTGCGCAAGGCCGAGTTCCAGTCGCTGCCGGAGCTCGAC- 2700
      G  S  K  V  G  E  I  V  R  K  A  E  F  Q  S  L  P  E  L  D

2701 -GATGTCGAAACCCAACTCAAGAGTGCGTTGAGCGCCATGCTGATCGGGGCCTACAGCACG- 2760
      D  V  E  T  Q  L  K  S  A  L  S  A  M  L  I  G  A  Y  S  T

2761 -TACACCGGTCAGGCTGTACTGGCCGAGGCGGTGAGTGGTTTTGCACGCAAGGCCAAGGTG- 2820
      Y  T  G  Q  A  V  L  A  E  A  V  S  G  F  A  R  K  A  K  V

2821 -GAGAACTACAGCGTGAGCTACAGCGGCTTCAAACCGCTGAAAGAGGCTTACAGCGACTAC- 2880
      E  N  Y  S  V  S  Y  S  G  F  K  P  L  K  E  A  Y  S  D  Y

2881 -CAGCAGAATGTTCCGAAGGCGAAGTTTGCCGCCTTCTGTGAACAGGTCGGGTTGTCGATG- 2940
      Q  Q  N  V  P  K  A  K  F  A  A  F  C  E  Q  V  G  L  S  M

2941 -TGGAGCAGTGATTTCGGTGCCAAGGGCAAGAACCACAGCGACTCGGAATGTGCCATCGGG- 3000
      W  S  S  D  F  G  A  K  G  K  N  H  S  D  S  E  C  A  I  G

3001 -CAGAAGCTGGTCGAGGATGGCAAGTTTGAAGAGCACTACACCTTCAAGGAAGGCAAGAGC- 3060
      Q  K  L  V  E  D  G  K  F  E  E  H  Y  T  F  K  E  G  K  S
                                                 ----->Stop codon of orf6
3061 -CTGTTTGCGGCGTTGTTCGTGAGCAAGGCGTATGAGCTGGAAACCAATTGACCGGTTGAT- 3120
      L  F  A  A  L  F  V  S  K  A  Y  E  L  E  T  N  *
```

```
1140 -GACAAACGCCGCAACAAGCACTCGTTCCGCATACTCGACCTTTGGTTAACTGGCCAACTA- 3120

3121 -CGCTGATAAACAAAGCCCCCGGCAACGATGAGTTGCCGGGGGCTTTTTCAATTCTGCGAA- 3180
      GCGACTATTTGTTTCGGGGGCCGTTGCTACTCAACGGCCCCCGAAAAAGTTAAGACGCTT
                          Inverted repeat
           Stop codon of orf7 <-----
3181 -TCAGAGGTTTCAGCGCATCGCCGCCAGCTTGATGCCGAACCCTACCAGGCAAGCCCCGGC- 3240
      AGTCTCCAAAGTCGCGTAGCGGCGGTCGAACTACGGCTTGGGATGGTCCGTTCGGGGCCG
                 *  R  M  A  A  L  K  I  G  F  G  V  L  C  A  G  A

3241 -CAGCCGCTCGAACATATTGGTAATGCGCGGGTTGGCGCGCATGCGTTCGGCCAGCCGGTG- 3300
      GTCGGCGAGCTTGTATAACCATTACGCGCCCAACCGCGCGTACGCAAGCCGGTCGGCCAC
        L R  E  F  M  N  T  I  R  P  N  A  R  M  R  E  A  L  R  H

3301 -GGTCAGCACCACGGCGATCAACCCGTACAGGAACGTGATCACCGCCACCGTTGCCGCCAT- 3360
      CCAGTCGTGGTGCCGCTAGTTGGGCATGTCCTTGCACTAGTGGCGGTGGCAACGGCGGTA
       T  L  V  V  A  I  L  G  Y  L  F  T  I  V  A  V  T  A  A  M

3361 -GAAGCCGAACGTCACCAACCCCTGATGCTTCACCGGGTCGATGAATAATGGGAAGAACGC- 3420
      CTTCGGCTTGCAGTGGTTGGGGACTACGAAGTGGCCCAGCTACTTATTACCCTTCTTGCG
       F  G  F  T  V  L  G  Q  H  K  V  P  D  I  F  L  P  F  F  A

3421 -CATGTAGAACATGATCGCCTTGGGGTTGAGCAGGGTGATCAGCATGGTCTGGCGCAGGTA- 3480
      GTACATCTTGTACTAGCGGAACCCCAACTCGTCCCACTAGTCGTACCAGACCGCGTCCAT
       M  Y  F  M  I  A  K  P  N  L  L  T  I  L  M  T  Q  R  L  Y

3481 -TTGGCCGTTGTCCATGCGGCTGGTGCGCGGCGCGTCACCCGGTTTGCTCAACAGCATGCG- 3540
      AACCGGCAACAGGTACGCCGACCACGCGCCGCGCAGTGGGCCAAACGAGTTGTCGTACGC
       Q  G  N  D  M  R  S  T  R  P  A  D  G  P  K  S  L  L  M  R

3541 -CAGGCCCAGGTAGGCGAGGTACGCGGCGCCGGCCCATTGCACGATGTGGAAGGCCGCAGG- 3600
      GTCCGGGTCCATCCGCTCCATGCGCCGCGGCCGGGTAACGTGCTACACCTTCCGGCGTCC
       L  G  L  Y  A  L  Y  A  A  G  A  W  Q  V  I  H  F  A  A  P

3601 -GTAGGTGGCCAGCAAGGTGGCGACGCCAGCTACCGCTAGCCACAACAGGACTTGGTCACC- 3660
      CATCCACCGGTCGTTCCACCGCTGCGGTCGATGGCGATCGGTGTTGTCCTGAACCAGTGG
       Y  T  A  L  L  T  A  V  G  A  V  A  L  W  L  L  V  Q  D  G

3661 -GACGATCACCCCACAGGTCGCGGCCAGGCCTGCCTTGATGCCGCCCTTGCCGGTGGCAGT- 3720
      CTGCTAGTGGGGTGTCCAGCGCCGGTCCGGACGGAACTACGGCGGGAACGGCCACCGTCA
       V  I  V  G  C  T  A  A  L  G  A  K  I  G  G  K  G  T  A  T

3721 -GATCAGGGCAAAATTGCCCGGGCCCGGGATGGCCAGAAGAATGATGAAGGCGATGACGAA- 3780
      CTAGTCCCGTTTTAACGGGCCCGGGCCCTACCGGTCTTCTTACTACTTCCGCTACTGCTT
       I  L  A  F  N  G  P  G  P  I  A  L  L  I  I  F  A  I  V  F
                 <-----Start codon of orf7
3781 -TGCGCCGTAGTCGGTGACGCCGAGCATGGTGAACTCCTGTAGAGCAATGCAAAGGCAGAG- 3840
      ACGCGGCATCAGCCACTGCGGCTCGTACCACTTGAGGACATCTCGTTACGTTTCCGTCTC
       A  G  Y  D  T  V  G  L  M        RBS

3841 -TCATGACTATCGTCGCCAACAGCGTGCTTTACAACCTGTTGTTGCACATTCGCAATCCAT- 3900
                        -35                      -10
3901 -CATTGCCAAGCACGACTTGGATCCAATGCCTGTTTTCTGTATCATCCCGCCGCCATCAGA- 3960
                              RBS        Start codon of orf8 ---->
3961 -CGATGGTCGAAGATGATTCATTTAAAGGACTTTGTAATGAAAAAGCTGTTCAAGGCCACC- 4020
                                          M  K  K  L  F  K  A  T

4021 -GTAGCCGTTGCTGTAGTTTCGGGTGTTGCCCTGCTGTCGGGTTGCACTGGCCAGGTTTAC- 4080
      V  A  V  A  V  V  S  G  V  A  L  L  S  G  C  T  G  Q  V  Y

4081 -AACCAGCCGAAAAACTGCACCTACGACTACCTGTTCCACCCTTCGGTTTCCATCTCCAAG- 4140
      N  Q  P  K  N  C  T  Y  D  Y  L  F  H  P  S  V  S  I  S  K
```

```
                                        ----->Stop codon of orf8
4141 -ATCATCGGTGGCTGCGGCCCGATCGATAAACTGCCTCAGCAGCAGTAATCTTGGCGGTAC- 4200
      I  I  G  G  C  G  P  I  D  K  L  P  Q  Q  Q  *

4201 -C
```

Figure 4.21 DNA sequence of the *orf5*, *orf6*, *orf7* and *orf8* and their deduced amino acids. The start and stop codons, putative ribosome binding-sites (RBSs), probable -35 and –10 regions (promoter sequences) and inverted repeat sequence are indicated.

The sequence homology search using the deduced proteins of *orf5*, *orf6*, *orf7* and *orf8* showed no similarity to 3,6-diketocamphane monooxygenase or the proteins in Baeyer-Villiger monooxygenase family. It is possible that we have failed in identification of the *Kpn*I DNA fragment when using the pQR277 probe, which resulted in picking up a DNA fragment that has highly homologous sequence to the probe (52.4% identity).

A result of the sequence homology search of the deduced amino acids of *orf5*, *orf6*, *orf7* and *orf8* is shown in Table 4.11. The BLAST search revealed the deduced amino acid of *orf5* had no significant identity to any proteins on the current database. The Orf5 protein has only partial overlapping sequences (40-60 residues) to a proteins of transcription factor EC from Homo sapiens and transporter ATP-binding protein from *Fusobacterium nucleatum* supsp. nucleatum ATCC 25586. The ORF6 also has a partial overlapping sequence to a putative protein from *Mesorhizobium loti* and pyruvate phosphate dikinase from *Clostidium symbiosis*. However, for the Orf7 and Orf8 protein, significant identities were found. The protein-protein BLAST search showed the amino acid sequence of *orf7* shared identity with several putative, membrane and amino acid transporter proteins. The examples of these proteins are a hypothetical protein from *E. coli*, probable transporter transmembrane protein from *Ralsotonia solanacearum* and Lys-type translocator from *Bacillus antracis* str. A 2012, with 33%, 31% and 27% identity respectively. The amino acid of Orf8 had significant sequence similarity to the conserved hypothetical protein *from Salmonella senterica* subsp. enterica serora typhi and

hypothetical protein from *E. coli* O157: H7, with 64% and 56% identity respectively (see also Table 4.11).

Table 4.11 Selected proteins that share homology with the deduced amino acid of *orf5*, *orf6*, *orf7* and *orf8*.

Protein	Microorganism	Identity	Similarity	Gap	Score (bits)	Expect value	Accession number
Orf5							
Transcription factor EC	*Homo sapiens*	42%	69%	0	34.7	2.1	D43945
ABC transporter ATPbinding protein	*Fusobacterium Nucleatum subsp. Nucleatum ATCC 25586*	32%	58%	0	34.4	2.9	AE010513
Type 2 ribosome-inactivating protein cinnamomum precursor	*Cinnamonmum camphora*	24%	40%	12%	33.9	3.4	AY039802
Putative monosaccharide transporter	*Petunia hybrida*	28%	43%	0	33.5	4.8	AF061106
Orf6							
Unknown protein	*Mesorhizobium loti*	30%	50%	14%	33.1	5.2	AP003001
Pyruvate phosphate dikinase	*Clostidium symbiosis*	23%	43%	13%	32.7	7.5	M60920
Orf7							
Hypothetical protein	*Escericia coli*	33%	57%	1%	113.0	1×10^{-24}	AE000274
Probable transporter transmembrane protein	*Ralsotonia solanacearum*	31%	48%	3%	87.0	1×10^{-16}	AL646057
LysE type translocator	*Bacillus anthracis str. A2012*	27%	45%	4%	79.0	3×10^{-14}	NC_003995
Orf8							
Conserved hypothetical protein	*Salmonella enterica* supsp. enterica serovar typhi	64%	80%	2%	66.6	7×10^{-11}	AC627281
Hypothetical protein	*Escericia coli* O157:H7	56%	70%	5%	65.5%	1×10^{-10}	AP002565

4.7 Summary

The DNA sequence analysis of *orf1*, *orf2*, *orf3* and *orf4* are summarized in Table 4.12. The ribosome binding sites, upstream region, of *orf2*, *orf3* and *orf4* also shown in Table 4.13.

Table 4.12 Summary of the analysis of the DNA and protein of *orf1*, *orf2*, *orf3* and *orf4*.

ORF	Nucleotide	Start Codon	Stop Codon	RBS	-35 -10 promoter region
orf1	684	Unknown	TAA	Unknown	Unknown
orf2	1092	ATG	TGA	GAGGA	None
orf3	645	ATG	TGA	AAAGA	TTAGAC (-35) TATTAT (-10)
orf4	984	ATG	TGA	GAGAA	TTCTCA (-35) TATAGCC (-10)

ORF	Inverted repeat	G+C Content	Deduced amino acid	Protein Mw	pI value
orf1	**GCACGTTCGC** TCTCC **GCGAACGTGC** **GCGAAC** GCAC **GTTCGC**	Unknown	>227	Unknown	Unknown
orf2	**CCACGCCGC** TAT **GCGGCGTGG**	59.25%	363	40,704	5.58
orf3	CAACTCTC GG GAGGTTTG	50.70%	214	23,656	5.86
orf4	TCAACCAT ATTGTTGA	58.64%	327	35,688	5.49

Table 4.13 Ribosome binding sites upstream of the *orf2*, *orf3* and *orf4*. The putative ribosome binding sites and start codons are shown in bold and underlined.

Region	Location	Sequence
orf2	763-776	**GAGGA**CACACT**ATG**
orf3	2590-2612	**AAAGA**AATTT**ATG**
orf4	3463-3474	**GAGAA**TCACTC**ATG**

The *orf1* gene that we obtained is incomplete, however, we were able to show that the deduced amino acid of *orf1* is highly related to steroid monooxygenase from *Rhododcoccus rhodochlorous* and Baeyer-Villger monoxygenase from *Acinetobacter* sp. NCIMB 9871. Moreover, the deduced amino acid of *orf1* contains the ATG-motif (AlaThrGly), FAD and NAD(P)H-binding domain, which is highly conserved among these monooxygenases.

The Orf2 protein has the conserved domain of FMN-binding bacteria luciferase, the luciferase-like monooxygenase. The sequence homology search found the deduced amino acid of *orf2* is homologous to the limonene monooxygenase from *Rhodococcus erythropolis* and several luciferase related proteins. Interestingly, the N-terminal sequence of the deduced amino acid of *orf2* is significantly homologous to that of 2,5-diketocamphane 1,2-monooxygenase, enzyme in the third step of camphor catabolic pathway. The predicted protein secondary structure of Orf2 protein consists of twelve helices (α_1- α_{12}) and ten strands (β_1-β_{10}), which coincide with the protein secondary structure of the luciferase-beta-subunit from *V. haveyi* whose 3D structure is known.

The Orf3 protein is identified as a regulatory protein. It has the conserved domain of tetR, bacterial regulatory protein in the tetR family. In this domain, it also contains a helix-turn-helix DNA binding motif. The sequence homology search showed that the deduced amino acid of *orf3* shared high identity and similarity to a number of regulatory proteins in different bacteria. The predicted secondary structure of Orf3 protein revealed that the protein consists predominantly of eight α-helices (α_1-α_8) and two β-strands (β_1 and β_2). The probable helix-turn-helix DNA binding motif in the Orf3 protein has a similar topology as that of QacR repressor protein from *S. aureus*.

The deduced amino acid of *orf4* has the conserved domain of proteins in the metallo-β-lactamase family. The Orf4 protein also shares significant homology with the known-protein of methyl parathion hydrolase from *Plesiomonas* sp. M6 and methyl parathion degrading protein from *Plesiomonas* sp. DLL-1. Also, the Orf4 protein showed similarity to a number of putative beta-lactamases and hydrolases. In the Orf4 amino acid sequence, the zinc-binding motif was identified. This motif is the amino acid sequence of HLHPDH agreed with that of HxHxDH (X is any amino acids), the consensus sequence of zinc-binding motif. The predicted protein secondary structure of Orf4 protein is similar to α/β barrel structure. The Orf4 protein consists of thirteen helices (α_1-α_{13}) and twelve strands (β_1-β_{12}). The potential zinc-binding motif of Orf4 protein is situated in the loop region between a β_3-strand and an α_8-helix. This conformation is characteristic of the active site of all proteins with the α/β barrel structure.

The deduced amino acid of *orf5*, *orf6*, *orf7* and *orf8* showed identities to several membrane proteins in current database. However, their N-terminal sequences had no similarity to that of 3,6-diketocamphane 1,6-monooxygenase. The deduced amino acids of these ORFs also had no similarity to the proteins in Baeyer-Villiger monooxygenase family as well.

Chapter 5

Expression and characterisation of Orf2 (monooxygenase) and Orf4 (hydrolase) protein

5.1 Introduction

In this chapter, further investigations to identify and characterise the novel gene that we are obtained are described. Two genes, *orf2* and *orf4*, were chosen for further investigations because the deduced amino acids of *orf2* and *orf4* had significant matches to proteins in current protein database.

As described in the previous chapter, the deduced amino acid of *orf2* is related to the known protein limonene monooxygenase from *R. erythropolis* and the deduced amino acid of *orf4* has 39-40% identity to the parathion hydrolase from *Plesiomonas* sp.. The results implied that the Orf2 protein might have functional similarity in an oxygenating reaction resembling that of limonene monooxygenase, and the ORF4 protein may perform a similar function to the parathion hydrolase. Moreover, it is probably that these two novel proteins may be involved in a oxygenation and hydrolytic reaction on the intermediates in the camphor metabolic pathway in *P. putida* NCIMB 10007.

In this chapter, the *orf2* and *orf4* were cloned into protein expression vectors, and these recombinant vectors were transformed into *E. coli*. The activities of the proteins encoded by *orf2* and *orf4* were then investigated.

5.2 Characterisation of Orf2 monooxygenase

5.2.1 Overexpression of the Orf2 monooxygenase in *E. coli* BL21(DE3)pLysS

To express the Orf2 monooxygenase gene, the *orf2* was subcloned into the expression vector pET21a (Figure 5.1). The pET21a expression vector carrying the *orf2* was designated pQR422. This recombinant plasmid was then transformed into *E. coli* BL21(DE3)pLysS. In the pET21a, the Orf2 protein was expressed under the control of a strong T7 promoter. Upon induction with IPTG, mRNAs from the *orf2* will be produced in a large amount by T7 RNA polymerase from the host bearing the T7 RNA polymerase, that is *E. coli* BL21(DE3)pLysS. Subsequently, the mRNAs will be translated to the Orf2 protein in large amounts.

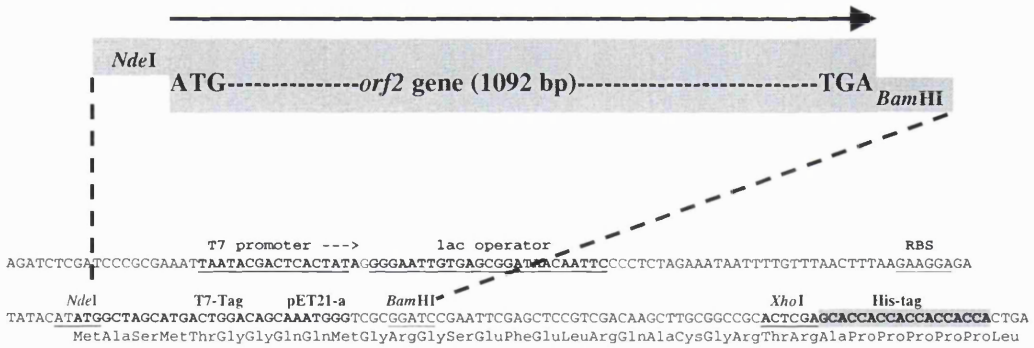


Figure 5.1 Cloning of the *orf2* into pET21a. The *orf2* with an additional *Nde*I and *Bam*HI site at its 5′ and 3′ end, respectively, was amplified by PCR and subcloned into a TOPO cloning vector. This recombinant vector was then digested with *Nde*I and *Bam*HI, and the *orf2* fragment was then cloned into the expression region of pET21a at the *Nde*I and *Bam*HI site.

Figure 5.2, after induction with IPTG, the Orf2 monooxygenase from *E. coli* Bl21(DE3)pLysS harbouring pQR422 was produced in substantial amounts in relation to the *E. coli* BL21(DE3)pLysS harbouring pQR422 without induction and to the control *E. coli* BL21(DE3)pLysS harbouring pET21a. A 40 kDa protein of the Orf2 monooxygenase can be prominently seen on the SDS-PAGE gel. An ability of the *orf2* to express in a reasonable level in the *E. coli* BL21(DE3)pLysS also supports that the codon usage pattern of *orf2* and that of *E. coli* genes are very similar, as the codon usage and amino acid composition of *orf4* compered to that of *E. coli* were shown in previous chapter (section 4.3.2 and 4.3.3)

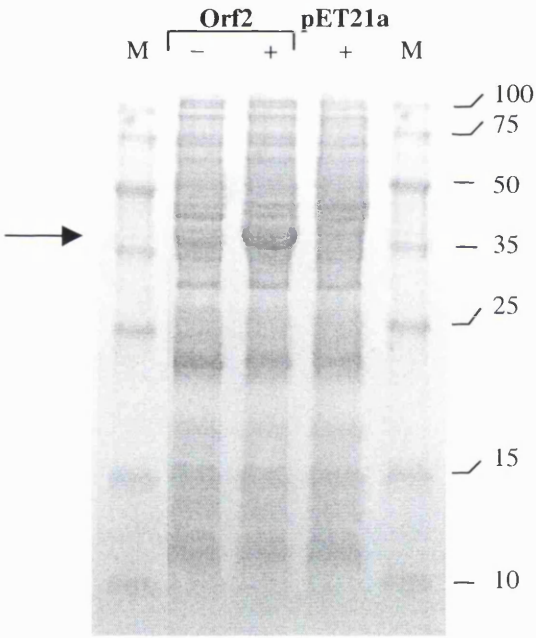


Figure 5.2 Overexpression of native Orf2 monooxygenase in *E. coli* BL21(DE3)pLysS. Pellete cells were boiled in 2×SDS-PAGE sample buffer and loaded onto a 12% SDS-polyacrylamide gel. The gel was stained by using Coomassie Brilliant Blue (as described in materials and methods). Orf2: *E. coli* BL21(DE3)pLysS harbouring pQR422 cells were induced with IPTG (+) or not induced (-); and pET21a: *E. coli* BL(DE3)pLysS harbouring pET21a induced with IPTG. The arrow indicates a protein band corresponding in size to an approximately 40 kDa protein of the Orf2 monooxygenase.

5.2.2 Overexpression of the His-tagged Orf2 monooxygenase in *E. coli* BL21(DE3)pLysS

For purification purposes, the *orf2* was amplified without the stop codon of TGA and cloned in frame with the DNA sequence of His in pET21a. This recombinant vector was used to produce His-tagged Orf2 monooxygenase (the Orf2 monooxygenase containing six histidine residues at its C-terminus). Thus, the His-tagged Orf2 monooxygenase can be exploited in protein purification by affinity chromatography. The recombinant vector pET21a-*orf2* with the histidine sequence was designated pQR423. This pQR423 was transformed into *E. coli* BL21(DE3)pLysS cells, the host for gene expression of a gene cloned in the pET vectors. After induction with IPTG, the fusion protein of Orf2 monooxygenase with six-histidines was produced in large amounts (Figure 5.3).

The overexpression of His-tagged Orf2 monooxygenase was monitored by SDS-PAGE (Figure 3). Upon induction with IPTG, the expression level of His-tagged Orf2 monooxygenase in *E. coli* BL21(DE3)pLysS was increased substantially compared to the expression level of non-induced cells. A prominent protein band (of 40 kDa) of the His-tagged Orf2 monooxygenase is clearly seen on the SDS-PAGE gel (Figure 5.3).

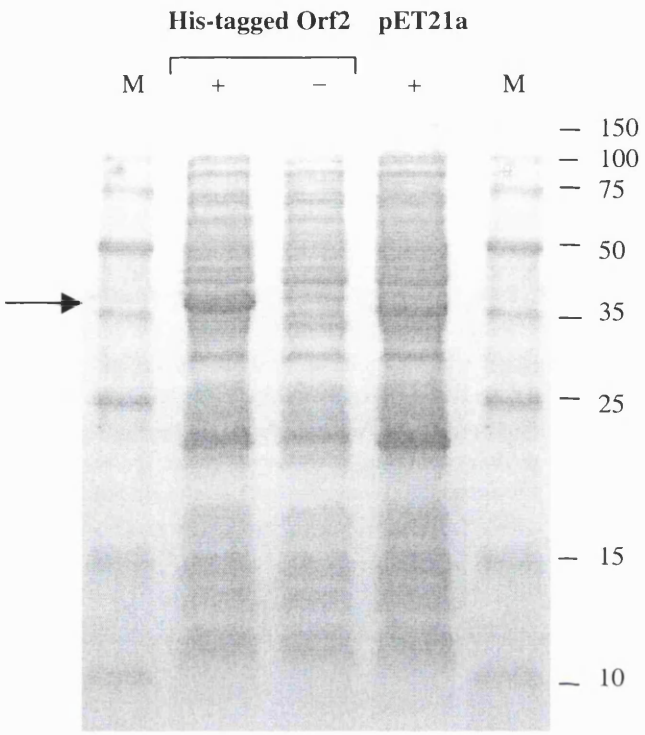


Figure 5.3 Overexpression of the His-tagged Orf2 monooxygenase in *E. coli* BL21(DE3)pLysS. *E. coli* BL21(DE3)pLysS harbouring pQR423 cells were induced with (+) and without (-) IPTG. The arrow indicates the protein band of His-tagged Orf2 monooxygenase (His-tagged Orf2, +). No protein band of the same molecular weight as the His-tagged Orf2 monooxygenase about 40 kDa can be seen in *E. coli* BL21(DE3)pLysS harbouring pET21a cells induced with IPTG.

5.2.3 Purification of the His-tagged Orf2 monooxygenase

A 200-ml culture of the *E. coli* BL21(DE3)pLysS harbouring pQR422 induced with IPTG was cultured until the late exponential phrase. Briefly, *E. coli* BL21(DE3)pLysS harbouring pQR422 was inoculated into 1 ml nutrient broth supplemented with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol and cultured at 37°C with 200 rpm shaking for 16-18 hours. This inoculumn was then put into 250 ml

nutrient broth supplemented with the same antibiotics as the previous culture and incubated at 37°C with 200 rpm shaking until the optical density (OD_{600}) of culture reached 0.6-0.8, then added IPTG to a final concentration of 1 mM and carried out further culture until the OD_{600} reached 4.0-5.0. The cells were harvested, sonicated, and the soluble fraction containing His-tagged Orf2 monooxygenase was purified using His-Bind columns (Novagen), Ni^{2+}-charged His-Bind resin columns (see section 2.30 in materials and methods).

After eluting the His-tagged Orf2 monooxygenase from the His-Bind columns, a protein yield from the final step of His-Bind purification was estimated as 0.18 mg/ml (from 200 ml culture), or approximately 9.0 mg of protein/ litre of culture. A single band of the His-tagged Orf2 monooxygenase is seen on SDS-PAGE (Figure 4) indicating that the His-tagged Orf2 monooxygenase was purified to homogeneity.

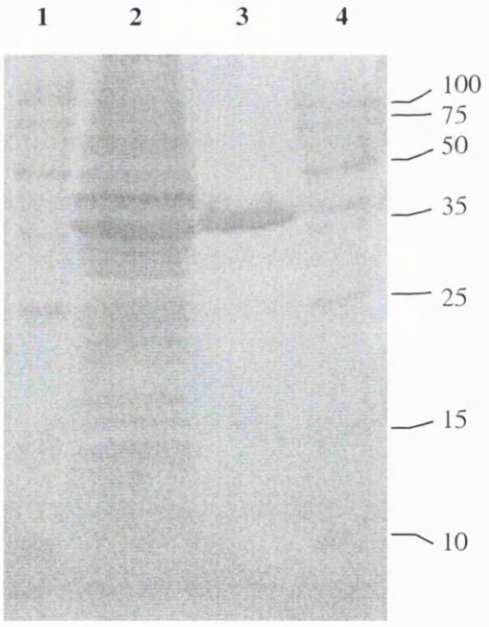


Figure 5.4 SDS-PAGE of the purified His-tagged Orf2 monooxygenase from *E. coli* BL21(DE3)pLysS harbouring pQR422. Lane1 and 4: marker proteins; Lane 2: *E. coli* BL21(DE3)pLysS harbouring pQR423 induced with IPTG; and Lane 3: the purified His-taged Orf2 monooxygenase (3.75 μg).

5.2.4 Characteristic spectra of the His-tagged Orf2 monooxygenase

The purified His-tagged Orf2 monooxygenase was added to 0.1 mM potassium phosphate buffer pH 7.2. A scanning mode in the spectrophotometer was selected to produce an absorption spectrum of the protein. The absorption spectrum of His-tagged Orf2 protein showed maxima at 221 nm (Figure 5.5).

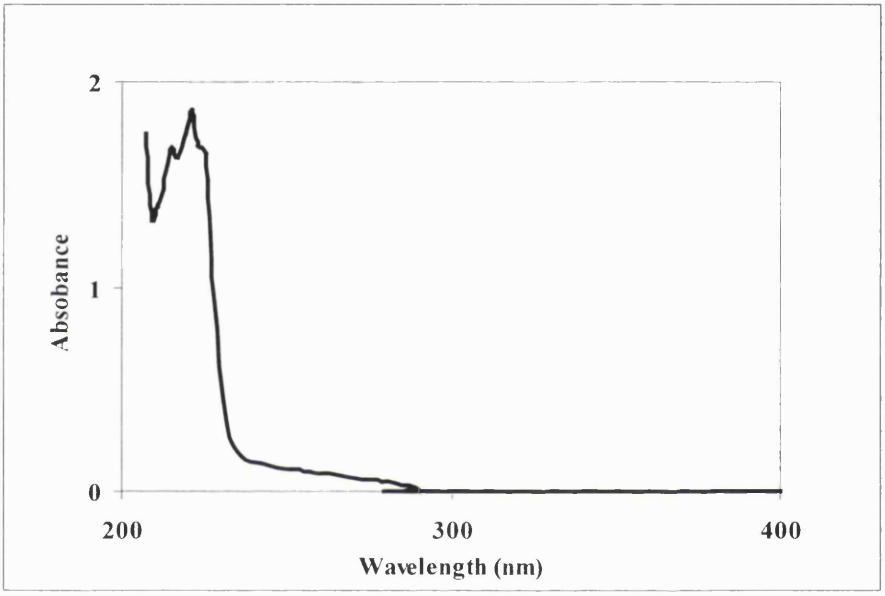


Figure 5.5 Absorption spectrum of His-tagged Orf2 monooxygenase (16.6 μg/ml) in 0.1 mM potassium phosphate buffer pH 7.2.

Although all protein can be measured by absorbance at 220 nm (peptide bonds (O=C–N–) give light absorbance at 220 nm), it has to be free of other substances, which have an absorbance in the far UV, in the solution. However, spectrophotometric analysis of the amino acid of tryptophan, tyrosine and phenylalanine in the side chains of proteins is more useful (Yang *et al.*, 1985). The light absorbance of tryptophan, tyrosine and phenylalanine is between 250-300 nm. This absorbance is generally used to detect and

quantify proteins. However, some proteins, which have structures predominated by α-helical, exhibit maxim at 220 nm, as helical peptides exposure (Surette and Stock, 1996; and Horn *et al.*, 1993). As the His-tagged Orf2 monooxygenase showed maxima at 220 nm, this reflectes that the protein is predominated by α-helical structures. The predominant of α-helices in the Orf2 monooxygenase component was also shown by the protein secondary prediction, as in the previous chapter section 4.3.7, which showed that the Orf2 monooxygenase consists of 12 α-helices or approximately 54% of its structural elements.

5.2.5 Characteristics of the Orf2 monooxygenase

To identify a probable substrate for the Orf2 monooygenase, the protein was tested with different substrates and coenzymes. (+)-limonene, (-)-limonene, (+)-α-pinene, (-)-α-pinene, cyclohexanone and cyclopentanone were tested as a substrate for the ORF2 monooxygenase. In the reaction, NADH or NADPH was also added as a coenzyme. In theory, if the Orf2 monooxygenase can catalyse the oxidation of the tested substrates and the coenzymes, the reaction can be monitored by reading the oxidation of NADH or NADPH (light absorption at 340 nm) as is oxidised to NAD^+ or $NADP^+$. However, the result showed that non of the above compounds served as a substrate for the Orf2 monooxygenase in the reaction with NADH or NADPH.

5.3 Characterisation of the Orf4 hydrolase

5.3.1 Overexpression of the Orf4 hydrolase in *E. coli* BL21(DE3)pLysS and BL21(DE3)CodonPlus-RP

To overexpress a 35.7 kDa protein of Orf4 hydrolase, the *orf4* was cloned into the expression vector of pET-21a, under the control of a strong T7 promoter. The expression

vector with the *orf4* insert was designated pQR424. The pQR424 was then introduced into a conventional *E. coli* BL21(DE3)pLysS host. However, the expression level of the Orf4 protein in the conventional *E. coli* BL21(DE3)pLysS host was very low, as can be seen on the SDS-PAGE in Figure 5.6 The level of the Orf4 hydrolase upon an induction with IPTG was about the same as that without an induction. The result supports a difference in the codon usage pattern of *E. coli* genes and *orf4*, as shown in previous chapter section 4.53 and 4.5.4. Moreover, because forced high-level expression of heterologous proteins can deplete the pool of rare tRNAs and stall translation, this is thought to be the reason why the expression of Orf4 hydrolase was very low.

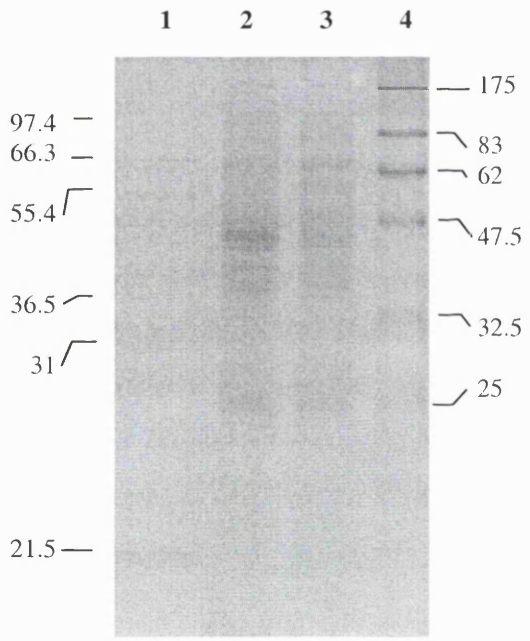


Figure 5.6 SDS-PAGE of the Orf4 hydrolase expression in *E. coli* BL21(DE3)pLysS. Lane 1 and 4: marker proteins; Lane 2: *E. coli* BL21(DE3)pLysS harbouring pQR424 without induction with IPTG; and Lane 3: *E. coli* BL21(DE3)pLysS harbouring pQR424 upon an induction with IPTG.

To achieve a high-level expression of the Orf4 hydrolase, a different host containing extra properties was used. This host was *E. coli* BL21(DE3) CodonPlus-RP (Stratagene). The *E. coli* BL21(DE3) CodonPlus-RP cells contain extra copies of the *argU* and *proL* tRNA genes. The *argU* and *proL* gene in the BL21(DE3)CodonPlus–RP encode the tRNAs that recognise the codon of AGA and AGG (arginine), and the codon of CCC (proline). In *E. coli* genes, the number of AGA and AGG codons are 3-4 codons per 1,000 codons. However, in the *orf4*, the number of AGA and AGG codons are 3 and 12 codons per 1,000 codons respectively.

With the extra tRNA genes in *E. coli* BL21(DE3)CodonPlus-RP, the protein production of Orf4 hydrolase was improved. Figure 5.7 shows the expression of Orf4 hydrolase in *E. coli* BL21(DE3)CodonPlus-RP upon an induction with IPTG. The expression level of Orf4 hydrolase in *E. coli* BL21 (DE3)CodonPlus-RP is better than in a conventional *E. coli* BL21(DE3)pLysS. A protein band of about 36 kDa can be prominently seen in the SDS-PAGE of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 induced with IPTG (Figure 5.7).

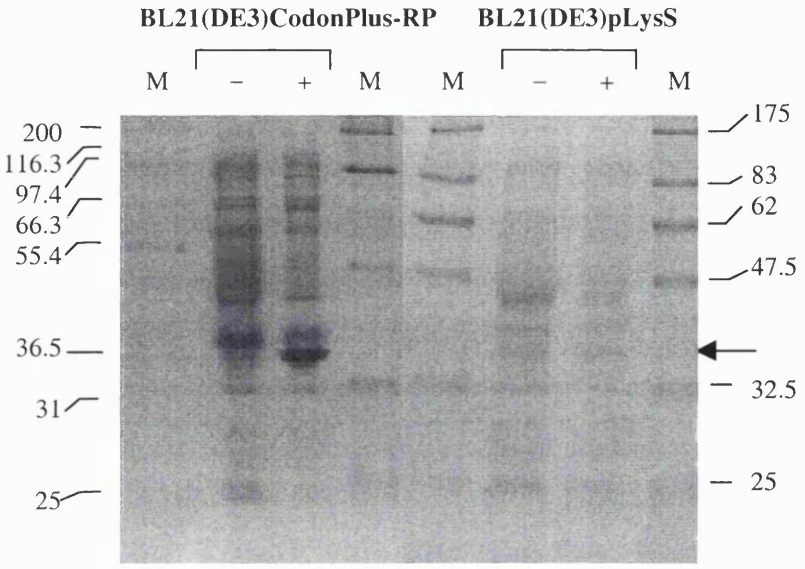


Figure 5.7 The expression of Orf4 hydrolase in *E. coli* BL21(DE3)CodonPlus-RP. M: marker proteins; BL21(DE3)pLysS: *E. coli* BL21(DE3)pLysS harbouring pQR424 with (+) and without (-) IPTG; and BL21(DE3)CodonPlus-RP *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 with (+) and without (-) IPTG. The arrow indicates the protein band of Orf4 hydrolase at 35.7 kDa (BL21(DE3)CodonPlus-RP, +).

5.3.2 Overexpression and purification of the His-tagged Orf4 hydrolase in *E.coli* BL21(DE3)CodonPlus-RP

For purification purposes, the *orf4* was cloned in frame with the C-terminal his-tagged sequences in pET21a. This recombinant vector was designated pQR425. The pQR425 was then transformed into *E. coli* BL21(DE3)CodonPlus-RP, as used as the expression host for the previous protein expression of Orf4 hydrolase. The SDS-PAGE of His-tagged Orf4 hydrolase expression in *E. coli* BL21(DE3)CodonPlus-RP is shown in Figure 5.8. However, when the His-tagged Orf4 hydrolase was overexpressed in *E. coli* BL21(DE3)CodonPlus-RP, the expected protein band of about 36 kDa could not be seen, instead, a protein band about 27 kDa was observed.

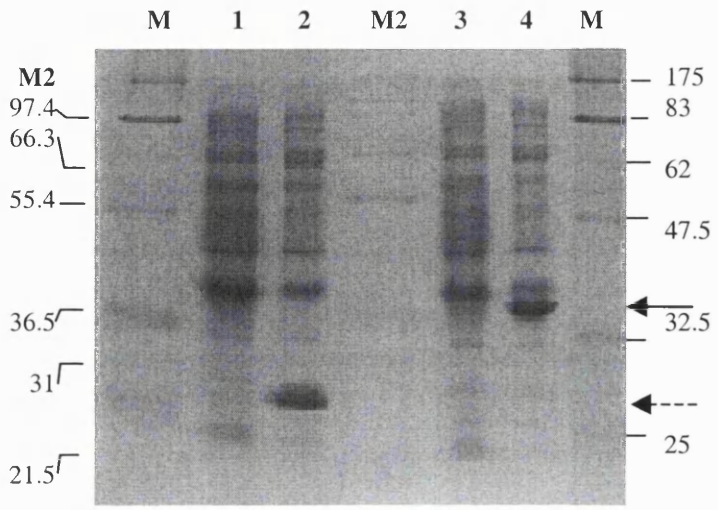


Figure 5.8 SDS-PAGE of the His-tagged Orf4 hydrolase expression in *E. coli* BL21(DE3)CodonPlus-RP. Lane1and 2: *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR425 without and with an induction with IPTG; Lane 3 and 4: *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 without and with an inducition with IPTG; and M and M2: marker proteins. The Orf4 hydrolase is indicated by the arrow (lane 4). The His-tagged Orf4 hydrolase is indicated by the dashed line arrow (lane 4).

A 200-ml culture of the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR425 induced with IPTG in late exponential phrase was harvested, and the cells were disrupted by sonication. Both soluble and membrane fraction of these cells were separated and subjected to His-tagged purification procedures using His-bind columns. Proteins eluted from the His-Bind columns, from soluble and membrane fractions, as well as different protein fractions from steps in the His-tagged purification procedures were monitored by SDS-PAGE (in Figure 5.9). However, the expected protein band of purified His-tagged Orf4 hydrolase on the SDS-PAGE gel was not observed.

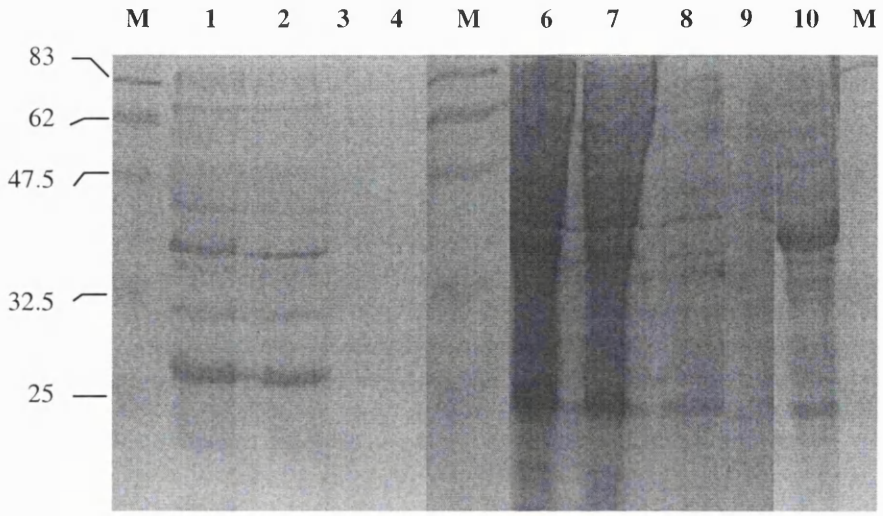


Figure 5.9 Analysis of the His-tagged Orf4 hydrolase by SDS-PAGE gel. M: marker proteins; Lane 1-4: inclusion body fraction purification (1: the supernatant; 2: the flow through of the supernatant; 3: the flow through after the washing; and 4: the eluted his-tagged Orf4 hydrolase); Lane 6-9: the soluble fraction purification (6: the supernatant; 7: the flow through of the supernatant; 8: the flow through after the washing; and 9 the eluted His-tagged Orf4 hydrolase); and Lane 10: *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR425 induced with IPTG.

Although the over expression of the His-tagged Orf4 protein was successful in *E. coli* BL21(DE3)CodonPlus-RP, we failed to purify the His-tagged Orf4 hydrolase. Unfortunately, it has been found that there were silent mutations during PCR mutation of the recombinant vector pQR421 (pCR2.1-TOPO-*orf4* with its stop codon deleted). First silent mutation is on the nucleotide position 69 (A→G); and the second mutation is on the nucleotide position 743 (G→A) of *orf4*. The first mutation altered a threonine (T24) of Orf4 hydrolase to an alanine (A); and the second mutation altered a tryptophan (W248) of Orf4 hydrolase to a stop codon (TAG), see Figure 5.10. As the result of the second silent mutation, a protein of about 27.1 kDa was generated. This explains the protein bands

observed in the overexpression of His-tagged Orf4 hydrolase (Figure 5.8 and 5.9), which have an identical molecular mass as the protein from the second mutation on *orf4*.

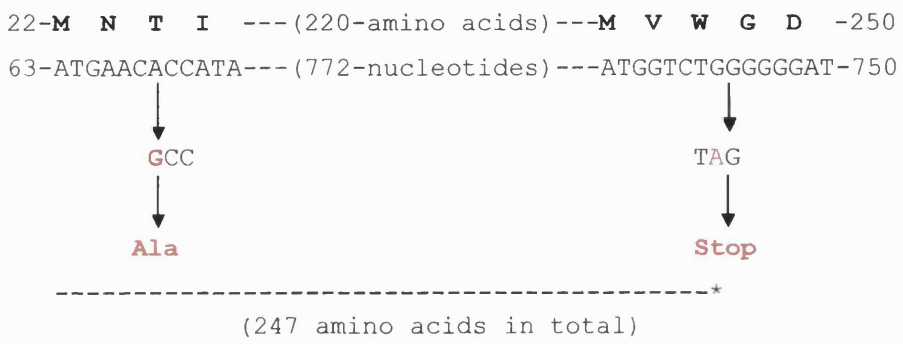


Figure 5.10 Two silent mutations during the PCR amplification of pQR421 (pCR2.1-TOPO-*orf4* with its stop codon deleted). The second mutation caused a shortened Orf4 hydrolase, which generated a protein about 27.1 kDa.

5.3.3 Localisation of the Orf4 hydrolase in *E. coli*

To determine the localisation of Orf4 hydrolase in *E. coli*, the soluble and membrane fractions of *E. coli* BL21(DE3)CodonPus-RP harbouring pQR424 (pET21a-*orf4*) were pelleted by centrifugation. The soluble fraction (consisting of the cytoplasm and periplasm) and membrane fraction was monitored by SDS-PAGE. A band of about 36-kDa protein corresponding to the Orf4 hydrolase can be seen clearly in the membrane fraction of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 (Figure 5.11). This result suggests that the Orf4 hydrolase is localised in the *E. coli* membrane.

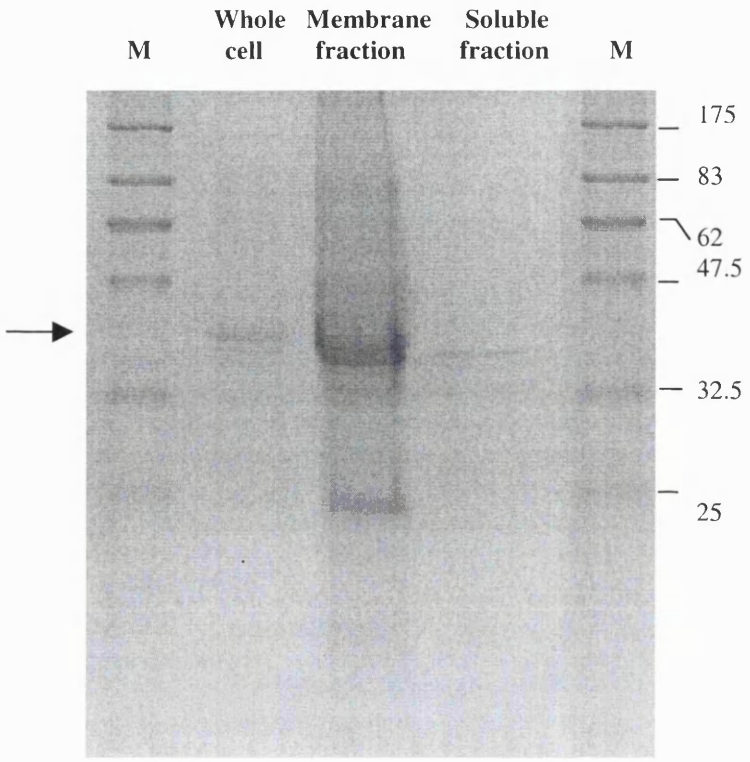


Figure 5.11 SDS-PAGE of the membrane and soluble fraction of the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 (pET21a-*orf4*). The arrow indicates the protein band of Orf4 hydrolase at about 36 kDa.

In gram negative bacteria, several proteins are localised to the periplasm or outer membrane. To localise these proteins to the periplasm or outer membrane, the proteins must have a recognition sequence for translocation and modification. The recognition sequence can be predicted by computational program SignalP, which is available at www.cbs.dtu.dk/services/SignalP. The prediction is made based on the basic characteristics of the distinct regions of signal peptides. There are three distinct regions on the signal peptides: 1) a positively charged N-terminus (n-region), 2) a central hydrophobic (h-region); and 3) a hydrophilic C-terminus (c-region) (von Heijne, 1988).

A prediction of the signal-peptidase cleavage-site on the Orf4 hydrolase is shown in Figure 5.12.

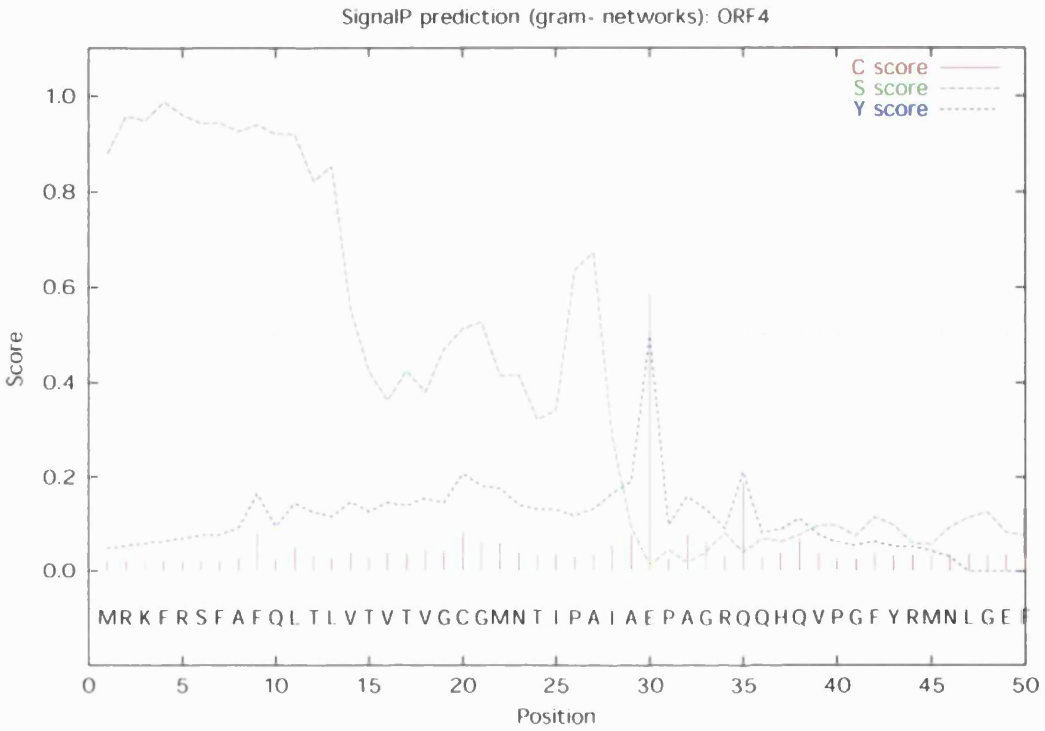


Figure 5.12 Signal peptidase cleavage-site on the Orf4 hydrolase predicted by SignalP. C-Score (raw cleavage site score): the score from the networks, which is trained to recognise cleavage sites, and it is high immediately after the cleavage site. S-score (signal peptide score): the score from the networks, which is trained to recognise signal peptide, and it is high before the cleavage site. Y-score (combined cleavage site score): the score formed by the combination of the C-score and S-score, it has a high score where the C-score is high and the S-score changes from a high to a low value.

The most likely cleavage-site on the Orf4 hydrolase is between amino acid positions 29 and 30: AIA-EP. To confirm a localisation of the Orf4 hydrolase, we also used the PSORT program at http://psort.nibb.ac.jp. This computational program is used to predict a subcellular localisation of the protein based on information of protein sorting

(Nakai and Hortan, 1999). According to the prediction of PSORT, it is likely that the Orf4 hydrolase is localised to periplasmic space.

If the Orf4 hydrolase is cleaved by the signal peptidase, as predicted by SignalP, the molecular mass of the protein after the modification should be about 32 kDa. However, as can be seen in the previous SDS-PAGE gels of the Orf4 hydrolase (previous Figure 5.7), the Orf4 hydrolase is not cleaved by the signal peptidase in the *E. coli.*

5.3.4 Whole cell activities of the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 toward paraoxon, parathion and methylparathion

As a result of the deduced amino acid of *orf4* showing homology to the methyl parathion and methyl parathion degrading protein from *Plesiomonas* sp., the activities of ORF4 hydrolase toward parathion, methyl parathion and the related compound, paraoxon, were tested.

At the start, the soluble and membrane fraction of the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 were pelleted by centrifugation, and both were used to test activity towards paraoxon. The result showed that both the soluble fraction, consisting of the cytoplasm and periplasm, and membrane fraction of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 were capable of hydrolysing the paraoxon. However, more paraoxon was converted by the membrane fraction than the soluble fraction. This confirmed that the Orf4 hydrolase is associated in *E. coli* membrane. (Figure 5.13).

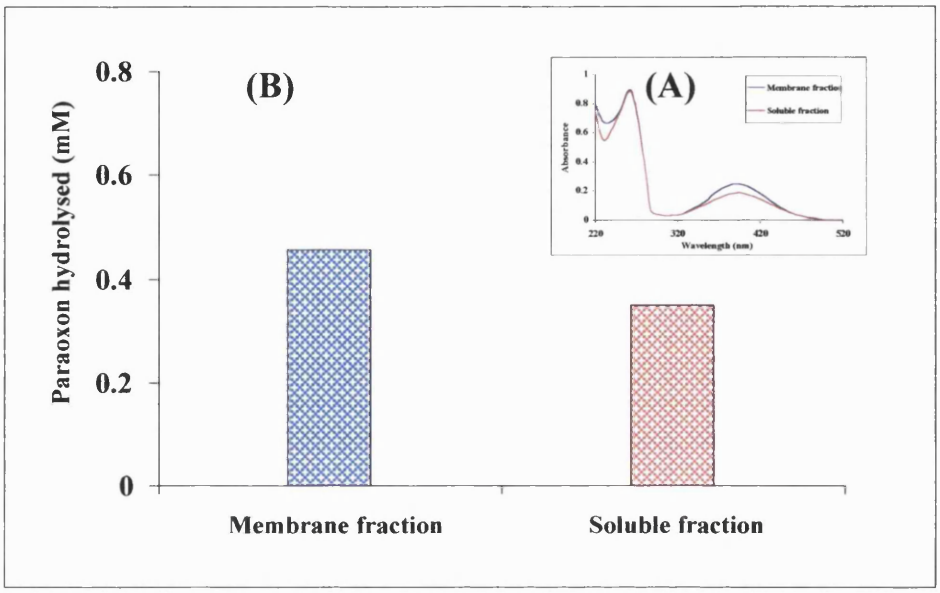


Figure 5.13 Hydrolysis of paraoxon by the membrane and soluble fraction from *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. The membrane and soluble fraction from the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 induced with IPTG were separated by centrifugation of the disrupted recombinant *E. coli* cells. The procedure was carried out as described in materials and methods section 2.30 except the culture volume was 20 ml and *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 was used instead. Each fraction was washed and resuspended in 1 ml 0.2 M Tris-Cl buffer pH 7.9. The reaction was initiated by the addition of paraoxon to the final concentration of 2 mM. The reaction tube was incubated at 37°C with 200 rpm shaking for 15 hours. After that, the cells were spun down, and the supernatant was sampled to determine the absorbance at 400 nm. The samples form the membrane fraction and soluble fraction tube had absorbance at 400 nm of 0.242 and 0.184 (**A**). These absorbance values were converted to the amount of paraoxon hydrolysed, which were 0.4564 and 0.3468 mM respectively (**B**). For the control i.e. *E.coli* harbouring pET21a suspended in Tris-Cl buffer, pH 7.9 and shaken with 2 mM paraoxon, small amount of paraoxon (0.06 mM) was hydrolysed.

In further investigation in the determination of Orf4 hydrolase towards paraoxon, parathion and methyl parathion, the activities were determined from the hydrolysis of the compounds with the cell suspension of *E. coli* harbouring pET21a-*orf4*. The whole cells of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 induced with IPTG were suspended in 1 ml of 0.2 M Tris-Cl buffer, pH 7.9. To initiate the reaction, the tested substrate was added to the final concentration of 2 mM. The reaction tube was incubated at 37°C with 200 rpm shaking. In parallel, a cell suspension of *E. coli* BL21 (DE3)pLysS cells harbouring pET21a was incubated with the tested substrate.

At time 0, 1,2,3,4.5,7.5,16 and 20 hours the mixture from the reaction tube was sampled, the cells were spun down, and the supernatant was read by spectrophotometer for absorbance at 400 nm. This is the peak absorption wavelength of *p*-nitrophenol, the product from the hydrolysis of parathion, methylparathion and paraoxon by the hydrolase enzyme. As can be seen in Figure 5.14, one mole of paraoxon, parathion or methyl parathion is catalysed by the hydrolase enzyme to one mole of *p*-nitrophenol and phosphoric acid compounds. Because of this, by measuring the amount of the *p*-nitrophenol formation in the reaction tube, we will be able to calculate the amount of paraoxon, parathion or methyl parathion hydrolysed accurately.

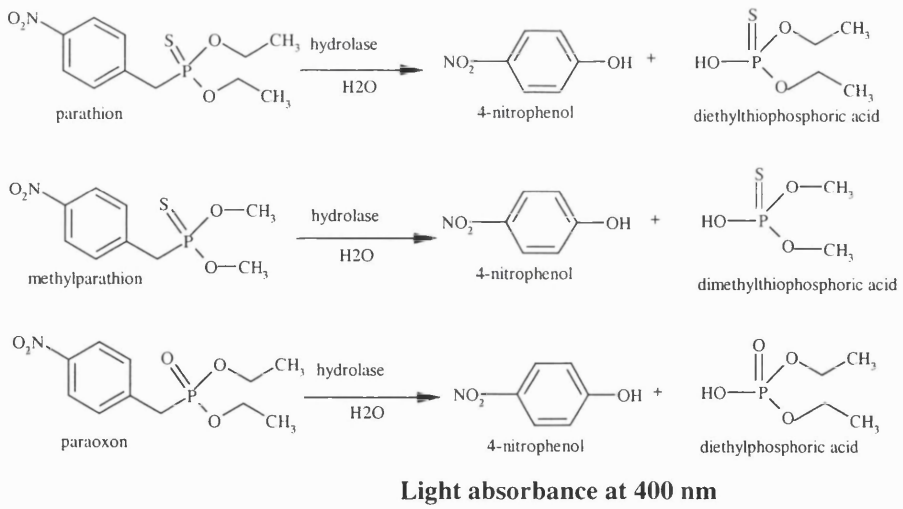


Figure 5.14 Hydrolysis reaction of parathion, methyl parathion and paraoxon by hydrolase enzyme.

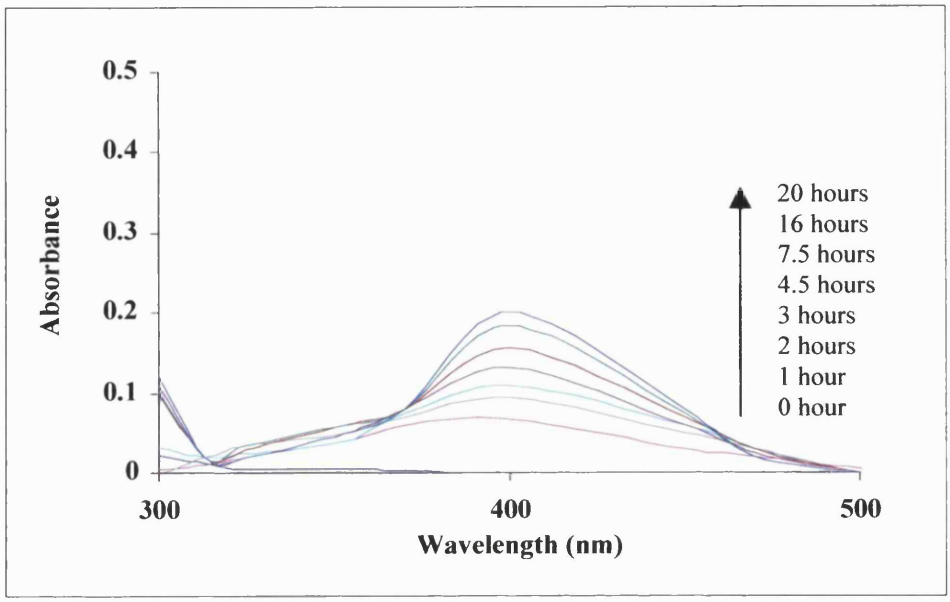


Figure 5.15. The absorption spectra of *p*-nitrophenol production (λ_{max} at 400 nm) from the hydrolysis of the paraoxon incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 for 20 hours. The spectra were recorded at 0, 1, 2, 3, 4.5, 7.5, 16 and 22 hour.

In Figure 5.15, the absorbance (at λ_{400}) of *p*-nitrophenol production from the reaction of paraoxon incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 were recorded at time 0,1, 2, 3, 4.5, 7.5, 16 and 20 hours. The increasing absorbance of *p*-nitrophenol at 400 nm indicated the conversion of the paraoxon to *p*-nitrophenol. The final product of *p*-nitrophenol can also be seen as an appearance of a bright yellow in the tube. The amount of *p*-nitrophenol increase is proportionate to the amount of paraoxon hydrolysed. After an addition of paraoxon into the reaction tube, the *p*-nitrophenol is almost immediately detected. The spectrum at each time point were recorded, and later calculated as an amount of paraoxon hydrolysed. The whole cell activities of *E. coli* BL21 (DE3)CodonPlus-RP harbouring pQR424 toward parathion and methyl parathion were also carried out as in the experiment of paraoxon. All raw data and calculations are shown in Table 5.1.

Substrates and sampling time (hours)	Absorbance reading at λ_{400}	$C_{sample} = A/\varepsilon l$ (ε=13.5 $mM^{-1}cm^{-1}$) (mM)	$C_{original}$ (C×51(dilution factor)) (mM)
Paraoxon			
0	0	0	0
1	0.068	0.0050	0.255
2	0.094	0.0069	0.3519
3	0.110	0.0082	0.4182
4.5	0.129	0.0095	0.4848
7.5	0.157	0.0116	0.5916
16	0.183	0.0135	0.6885
20	0.200	0.0148	0.7584
Methyl parathion			
0	0	0	0
1	0.026	0.0019	0.0969
2	0.029	0.0021	0.1071
3	0.034	0.0025	0.1275
4.5	0.041	0.0030	0.1530
7.5	0.042	0.0031	0.1581
16	0.051	0.0037	0.1887
20	0.051	0.0037	0.1887
Parathion			
0	0	0	0
1	0.01	0.0007	0.0357
2	0.004	0.0003	0.0153
3	0.008	0.0006	0.0306
4.5	0.004	0.0003	0.0153
7.5	0.009	0.0006	0.0306
16	0.016	0.0012	0.0612
20	0.014	0.0010	0.051

Substrates and sampling time (hours)	Absorbance reading at λ_{400}	$C_{sample} = A/\varepsilon l$ (ε=13.5 $mM^{-1}cm^{-1}$) (mM)	$C_{original\ sample}$ (C×51(dilution factor)) (mM)
Control			
0	0	0	0
1	0.016	0.0012	0.0612
2	0.020	0.0015	0.0765
3	0.019	0.0014	0.0714
4.5	0.018	0.0013	0.0663
7.5	0.025	0.0018	0.0918
16	0.026	0.0019	0.0969
20	0.026	0.0019	0.0969

Table 5.1 Raw data and calculation of the paraoxon, parathion and methyl parathion hydrolysed in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 (10 mg-dry cell weight/ml). Paraoxon, parathion and methyl parathion: the incubation of paraoxon, parathion, methyl parathion with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424; Control: the incubation of paraoxon with *E. coli* BL21 (DE3)pLysS harbouring pET21a; C_{sample} : concentration of sample (mM); A: absorbance; ε: extinction coefficient (13.5 $mM^{-1}cm^{-1}$); l: cuvette path length; and $C_{original\ sample}$: actual concentration in the reaction tube (mM)

From Table 5.1, a graph of the molarity of paraoxon, parathion and methyl parathion hydrolysed can be plotted as in Figure 5.16.

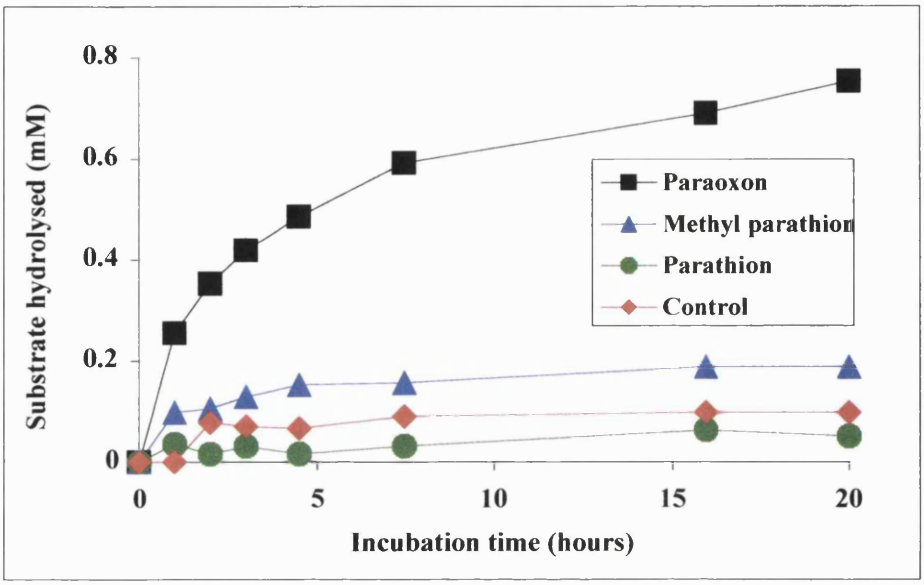


Figure 5.16 The hydrolysis of paraoxon, parathion and methyl parathion by *E. coli* BL21(DE3)Codon-Plus-RP harbouring pQR424 (10 mg-dry cell weight /ml).

As shown in Figure 5.16, paraoxon was degraded rapidly in the first 5 hours in the incubation with the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. Almost 0.5 mM of paraoxon was hydrolysed within 5 hours; however, between time 5 to 20 hours the rate of the hydrolysis of paraoxon gradually decreased, and about 0.3 mM of paraoxon was degraded during this time. About 0.41 mM of paraoxon was completely hydrolysed with in 3 hours. This hydrolysis of paraoxon can be calculated as a reaction velocity of 13.0 μmol/h/mg dry cell weight. The hydrolysis of 0.12 mM of methyl parathion was complete within 3 hours (4.2 μmol/h/mg dry cell weight). In the same conditions, when the amount of cells were doubled, the amount of paraoxon hydrolysed also increased two folds. The hydrolysis of parathion by the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 is very low. In fact, after initial incubation for 3 hours with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424, 0.03 mM of parathion was hydrolysed (1.0 μmol/h/mg dry cell weight). This was less than the amount of paraoxon hydrolysed in the incubation with *E. coli* BL21(DE3)pLysS harbouring pET21a.

5.3.5 Whole cell activities of the *E. coli* BL21(DE3)CodonPlus-RP carrying Orf4 hydrolase toward lactones

P. putida NCIMB 10007 can use either (+) or (-)-camphor as a sole carbon and energy source. In the camphor catabolic pathway of *P. putida* NCIMB 10007, lactones are often found to be intermediates in the steps of this catabolic pathway. It is possible that the *P. putida* may employ various lactone hydrolases to metabolise these intermediate lectones. However, the enzyme that is involved in the hydrolytic ring cleavage of lacotones in the camphor catabolic pathway has not been reported. In this section, we will investigate whether the Orf4 hydrolase is involved in lactone hydrolysis.

The activities of Orf4 hydrolase, whole cell activities, toward γ-butyrolactone, gluconolactone, pantolactone, γ-caprolactone and δ-varelolactone were determined. Briefly, in the reaction tube, the cell pellet of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 (11 mg-dry cell weight) was suspended in 1 ml 0.2 ml Tris-Cl pH 7.9. The lactone was added to a final concentration of 2 mM. The reaction tube was incubated at 37°C with 200 rpm shaking for 30 minutes. After the incubation, the cells were spun down at 14,000 rpm for 30 seconds, and the buffer solution was pipetted to a new tube. The amount of lactone hydrolysed in the reaction tube was determined as described by Fishbein and Bessman (1966). To the tube containing the buffer solution, 1 ml of alkaline hydroxylamine was added, followed by 2 ml of ethanolic ferric chloride. The colour complex of the lactone reacted with the alkaline hydroxylamine and ethanolic ferric chloride was determined spectrophometrically at 520 nm within 10 minutes. The control tubes were the lactone with 1 ml of 0.2 M Tris-Cl buffer pH 7.9 alone and the lactone with the *E. coli* BL21 (DE3)pLysS harbouring pET21a cell suspended in 1 ml of 0.2 M Tris-Cl buffer pH 7.9. The different absorbance at 520 nm between the control and the reaction tube is the amount of lactone hydrolased (materials and methods section 2.36). Figure 5.17 shows the absorption spectra of the colour complex of γ-butyrolactone with alkaline hydroxylamine and ethanolic ferric chloride (γ-butyrolactone incubated

with the *E. coli* harbouring pET21a and with the recombinant *E. coli* harbouring pET21a-*orf4*).

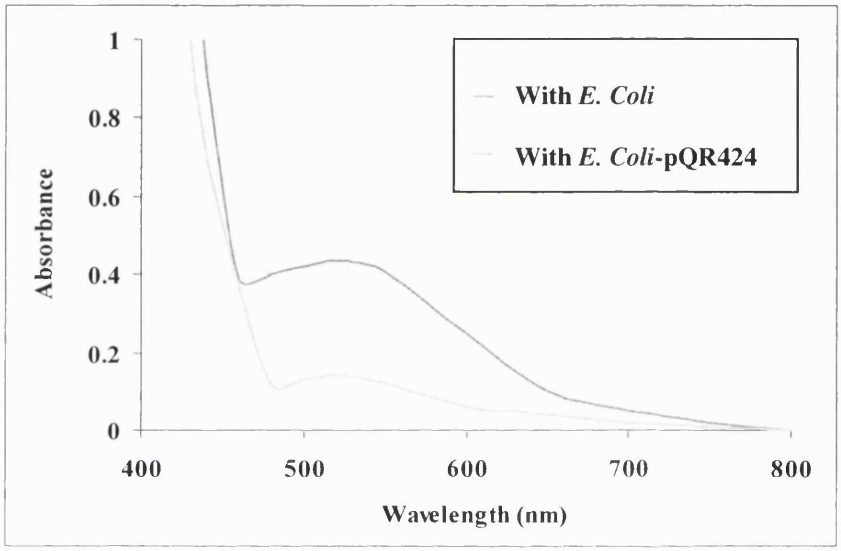


Figure 5.17 The absorption spectra of the colour complex of γ-butyrolactone with alkaline hydroxylamide and ethnolic ferric chloride. The upper spectrum is γ-butyrolactone (2.0 mM) incubated with *E. coli* BL21 (DE3)pLysS harbouring pQR424. The lower spectrum is γ-butyrolactone incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. The incubation time is 30 minutes.

As can be seen in Figure 5.17, the amount of γ-butyrolactone in the incubation with the *E. coli* harbouring pQR424 is significantly less than that of the control, γ-butyrolactone incubated with *E. coli*-pET21a. In fact, the absorbance at 520 nm of detectable γ-butyrolactone in the reaction tube is 74% less than that in the control tube.

This showed a significant reduction of γ-butyrolactone when it was incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. Table 2 shows the results and calculations of the γ-butyrolactone, γ-caprolactone, gluconolactone, pantolactone and δ-valerolactone hydrolysed in an incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 with in 30 minutes.

Absorbance at λ_{520}	γ-Butyrolactone (with the buffer alone = 0.404)			γ-Caprolactone (with the buffer alone = 0.276)			Gluconolactone (with the buffer alone = 0.133)			Pantolactone (with the buffer alone = 0.250)			δ-Valerolactone (with the buffer alone = 0.326)		
Experiment	+	C	%	+	C	%	+	C	%	+	C	%	+	C	%
1	0.140	0.556	65.3	0.307	0.336	0	0.163	0.203	0	0.319	0.325	0	0.383	0.535	0
2	0.260	0.485	35.6	0.336	0.361	0	0.172	0.195	0	0.320	0.359	0	0.346	0.418	0
3	0.164	0.438	59.4	0.278	0.301	0	0.155	0.182	0	0.248	0.288	0.8	-	-	-
Mean			53.4												
SD			4.22												

Table 5.2 Raw data and calculations of the lactones hydrolysed in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 (11 mg-dry cell weight/ml). Absorbance at λ_{520} : the absorbance of the colour complex of the reaction of lactones with alkaline hydroxylamine and ethanolic ferric chloride; + : the lactone incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 in duced with IPTG; C : the lactone incubated with *E. coli* BL21(DE3)pLysS harbouring pET21a; and %: the percentage of lactone hydrolysed compared to the control i.e. lactones incubated with 0.2 M Tris-Cl buffer, pH 7.9 alone. For gluconolactone and pantolactone incubated with the buffer alone, these compounds were also spontaneously hydrolysed. This was also observed by Fishbein and Bessman (1966). Therefore, the calculations of the lactone hydrolysed for all lactones were determined as the percentage of lactone hydrolysed over the spontaneous hydrolysis of the lactone in the control tube. However, for γ-butyrolactone, γ-caprolactone and δ-valerolactone, the compounds are stable in 0.2 M Tris-Cl buffer, pH 7.9. We were able to calculate the mean and standard deviation of γ-butyrolactone hydrolaysed in percentage as 53.4 and 4.22 respectively.

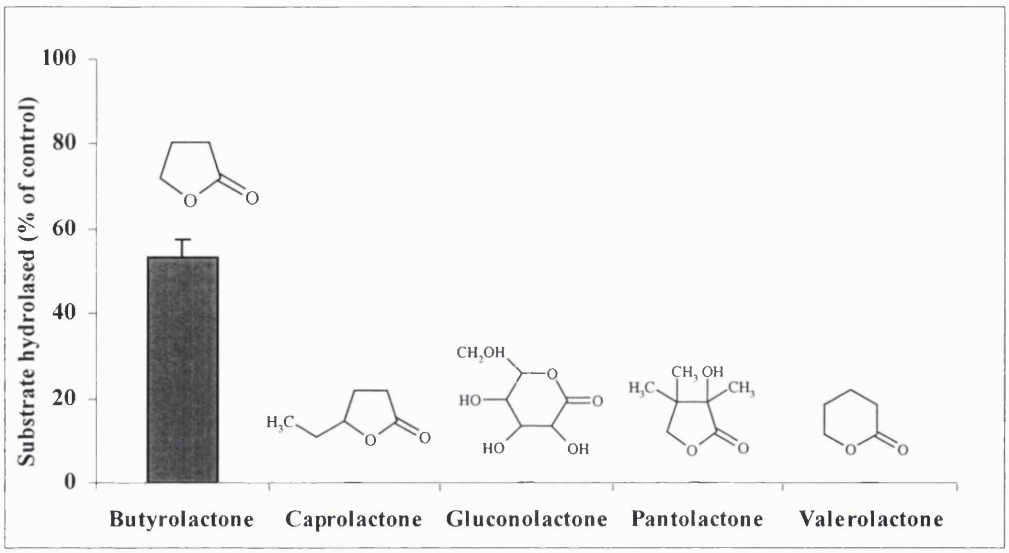


Figure 5.18 The hydrolysis of γ-butyrolactone, γ-caprolactone, gluconolactone, pantolactone and δ-valerolactone (2.0 mM) by the *E. coli* BL21 (DE3)CodonPlus-RP harbouring pQR424 (11 mg-dry cell weight/ml). The incubation period is 30 minutes.

Figure 5.18 represents the degradation of γ-butyrolactone, γ-caprolactone, gluconolactone, pantolactone and δ-valerolactone in the incubation with the *E. coli* BL21 (DE3)CodonPlus-RP harbouring pQR424 cells from Table 5.2. The vertical axis shows the percentage of lactones hydrolysed after incubation with the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 and the horizontal axis compares the different lactones, γ-butyrolactone, γ-caprolactone, gluconolactone, pantolactone and δ-valerolactone. According to Figure 5.18, there is a considerable amount of γ-butyrolactone hydrolysed in related to the other lactones. In fact, 53.4% of γ-butyrolactone was hydrolysed compared to less than 1% of patolactone was hydrolysed in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. However, for γ-caprolactone, gluconolactone and δ-varelolactone, these lactones were not hydrolysed in the incubation with the *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. This shows that the ORF4 hydrolase prefers γ-butyrolactone as a substrate, and it has very

poor activities toward pantolactone, and no activity towards γ-caprolactone gluconolactone and δ-valerolactone. For γ-bytyrolactone, almost 1.0 mM of γ-butyrolactone was hydrolysed within 30 minutes in the incubation with by *E. coli* BL21(DE3)CodonPlus harboring pQR424 (11 mg-dry cell weight). In comparion to the hydrolysis of paraoxon by the same cell, less than 0.2 mM of paraoxon was hydrolysed by *E. coli* BL21(DE3)CodonPlus-RP harboring pQR424 within the same period.

The experiment of the hydrolysis of γ-butyrolactone by *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 for a prolonged period was also carried out. With a similar condition as the above experiment, 2.0 mM of γ-butyrolactone was incubated with the cell suspension of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 in 0.2 M Tris-Cl buffer, pH 7.9. The reaction was monitored every 30 minutes for 3 hours (see Table 5.3).

Time	Absorbance at 520 nm				γ-Butyrolactone hydrolysed (% of control)			
	Experiment				Experiment			
	1 (+)	2 (+)	*E. coli*	Control	1′ (+)	2′ (+)	*E. coli*	Control
0	0.427	0.413	0.443	0.412	0	0	0	0
30	0.381	0.364	0.448	0.414	8.08	12.06	0	0
60	0.319	0.347	0.435	0.405	21.23	14.32	0	0
90	0.302	0.302	0.436	0.392	17.30	17.30	0	0
120	0.303	0.310	0.471	0.432	29.92	28.21	0	0
150	0.238	0.241	0.443	0.392	39.23	38.85	0	0
180	0.269	0.262	0.448	0.428	37.15	38.67	0	0

Table 5.3 γ-Butyrolactone hydrolysed in the incubation with *E. coli* BL21 (DE3)CodonPlus-RP harbouring pQR424 (3 mg-dry cell weight/ ml)1(+) and 2 (+): γ-butyrolactone incubated with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 induced with IPTG (duplicate experiment); *E. coli*: γ-butyrolactone incubated with *E. coli* BL21(DE3)pLysS harbouring pET21a; and Control: γ-butyrolactone in 0.2 M Tris-Cl buffer, pH 7.9 alone. Means of γ-butyrolactone hydrolysed (% of control) in a duplicate experiment (1′(+) and 2′(+)) at time 0 = 0, 30 = 10.07, 60 = 17.78, 90 = 17.3, 120 = 29.06, 150 = 38.85 and 180 = 37.91. Standard deviation (SD) of γ-butyrolactone hydrolysed (% of control) at time 0 = 0, 30 = 1.99, 60 = 2.63, 90 = 0, 120 = 1.3, 150 = 0.87 and 180 = 1.23.

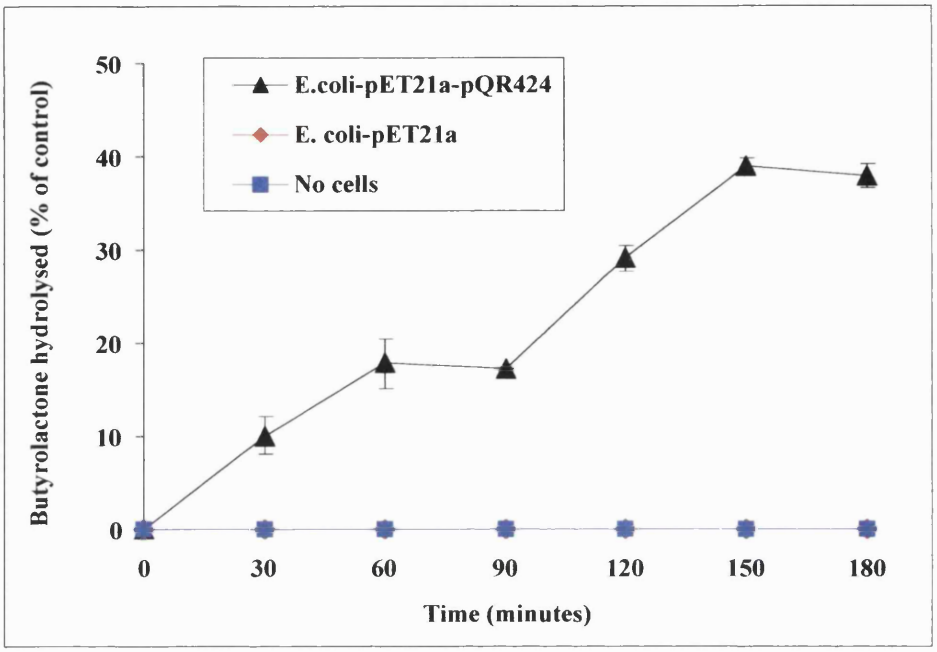


Figure 5.19 The hydrolysis of γ-butyrolactone (2.0 mM) by *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424.

The results from Table 5.3 can be represented as a graph (see Figure 5.19). This graph shows an increased amount of γ-butyrolactone hydrolysed in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424. The γ-butyrolactone of 37.9% (or about 0.8 mM) was hydrolysed within 3 hours (84.2 μmol/h/mg-dry cell weight). It is also clear that γ-butyrolactone is stable in the buffer, and it is not hydrolysed by *E. coli* BL21(DE3)pLysS harbouring pET21a.

In comparison to the Orf4 hydrolase activities toward γ-butyrolactone to paraoxon, γ-butyrolactone is a prefered substrate for the Orf4 hydrolase, as initial hydrolysis rate for γ-butyrolactone and paraoxon are 84.2 and 13.0 μmol/h/mg-dry cell weight respectively.

5.4 Summary

The *orf2*, and *orf2* with the DNA sequence of His in pET21a expression vector were constructed and transformed into *E. coli* BL21(DE3)pLysS. Upon induction with IPTG, a protein band of about 40 kDa of Orf2 monooxygenase and His-tagged Orf2 monooxygenase can be seen prominently on the SDS-PAGE gel. The ability of these monooxygenases to express at a reasonable level in *E. coli* suggests that the codon usage between the *orf2* and *E. coli* genes are very similar. Also, as it has been shown in previous chapter (section 4.3.3), the amino acid composition of *orf2* is similar to the average amino acid compositions of proteins from *E. coli.* In the purification of His-tagged Orf2 monooxygenase using His-bind columns, about 7.3 mg of protein/a litre of culture was obtained. The characteristic spectrum of His-tagged Orf2 monooxygenase showed a light absorbance at 221 nm. This implies that the Orf2 monooxygenase is predominated by α-helical elements, similar to some proteins with helical peptides exposure. The Orf2 monooxygenase activities were tested with (+)-limonene, (-)-limonene, (+)-α-pinene, (-)-α-pinene, cyclohexanone and cyclopentanone in the reaction with NADH or NADPH. However, the reaction of Orf2 monooxygenase with these compounds has not been observed.

The Orf4 hydrolase protein in *E. coli* BL21(DE3)pLysS was expressed at a very low level. The reason for this would be the different codon usage patterns of the *orf4* and that of *E. coli* genes. The expression level of Orf4 hydrolase was improved in *E. coli* BL21(DE3)CodonPlus-RP (containing extra copies of *argU* and *proL* tRNA gene). The Orf4 hydrolase is not cleaved by enzyme peptidases in the *E. coli.* As can be seen on the SDS-polyacrylamide gel, the Orf4 hydrolase has the molecular mass of 36 kDa, corresponding to the predicted molecular mass from the deduced amino acid of *orf4* (35,688 Da). This occurrence is likewise the membrane associated parathion hydrolase from *Flavobacterium* sp. strain ATCC 27551, that when it is expressed in *E. coli*, its signal peptide is not cleaved.

The overexpression and purification of His-tagged Orf4 hydrolase from *E. coli* BL21(DE3)CodonPlus-RP were unsuccessful. Because there was a silent mutation in the nucleotide position 743 (A→G) of *orf4*, which altered the amino acid of tryptophan (W248) of Orf4 hydrolase to a stop codon (TAG). This resulted in a shortened Orf4 hydrolase, as can be seen on the SDS polyacrylamide gels (Figure 5.8 and 5.9), a protein band of about 27 kDa corresponding to the shortened Orf4 hydrolase. This is the reason that we failed to purify the His-tagged Orf4 hydrolase.

The SDS-PAGE gel of membrane and soluble fraction of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 show that the native Orf4 hydrolase is localised to the membrane. The prediction of the signal peptidase cleavage-site, a recognition sequence of membrane proteins by the SignalP, showed a potential signal-peptidase cleavage-site on the Orf4 hydrolase. The prediction indicated that the most likely cleavage-site on the N-terminus of Orf4 hydrolase was between amino acid position 29 and 30: AIA-EP. Equally, the signal peptide of Orf4 hydrolase is in the same length as the signal peptide of the parathion hydrolase from *Flavobacterium* sp. ATCC 27551 (29 residues). In comparison, the signal peptide of Orf4 hydrolase is in similar length with that of the average Gram-possitive signal peptide (32.0 residues), not that of Gram-negative and eukaryotic signal peptide, 25.1 and 22.6 residues respectively (Nielsen et al., 1997). Also, the prediction of subcellular localisation by the PSORT programme revealed that Orf4 hydrolase is localised to periplasmic space.

It was decided to carry out work using the recombinant harbouring the native Orf4 hydrolase (*E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424), and to use whole cells to carry out some substrate activity tests. The initial work revealed that either the membrane or the soluble fraction separated from the disrupted cell of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 were capable of hydrolysing paraoxon. This indicates that the Orf4 hydrolase can be found in either soluble or membrane fraction. However, as it is showed on the SDS-PAGE (Figure 5.11) that the Orf4 hydrolase is localised to the E. coli membrane, the explaination for the existance of Orf4

hydrolase in the soluble fraction would be because of the sonication procedure, which released some Orf4 hydrolases from the *E. coli* BL21(DE3)CodonPlus-RP membrane.

The whole cell activities of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 showed 0.41 mM and 0.12 mM of paraoxon and methyl parathion were hydrolysed within 3 hours (13.0 μmol/h/mg-dry cell weight for paraoxon; and 4.2 μmol/h/mg-dry cell weight for methyl parathion). However, the Orf4 hydrolase showed low activity toward parathion (1.0 μmol/h/mg-dry cell weight). The plotted curve of the hydrolysis of paraoxon in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 cells also showed a hyperbolic curve characteristic, a typical Michaelis-Menten kinetic of an enzyme with a single catalytic site.

The whole cell activity of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 toward lactones showed high activity toward γ-butyrolactone. Within 30 minutes, 53% of γ-butyrolactone was hydrolysed compared to less than 1.0.% of pantolactone hydrolysed within the same period. For γ-caprolactone, δ-valerolactone and gluconolactone, the hydrolysis of these compounds by *E. coli* harbouring pQR424 did not occur. The hydrolysis of γ-butyrolactone was also followed for 3 hours. In initial hydrolysis rate for γ-butyrolactone is 84.2 μmol/h/mg-dry cell weight. In comparison to the hydrolysis of paraoxon to γ-butyrolactone, it is clear that the Orf4 hydrolase prefers γ-butyrolactone rather than paraoxon as a substrate.

Chapter 6

Discussion

6.1 Introduction

The ability of *P. putida* NCIMB 10007 to grow on the monoterpene ketone, camphor, is conferred by the CAM plasmid. The studies of *camR*, *camD*, *camC*, *camA* and *camB* from the CAM plasmid, which are responsible for early steps of camphor degradation, have been previously described (Koga *et al.*, 1985; Koga *et al.*, 1989; Aramaki *et al.*, 1993; and Aramaki *et al.*, 1994). The CAM plasmid, based on biochemical studies of the camphor metabolic pathway, contains at least three Baeyer-Villiger monooxygenases; 2,5 diketocamphane 1,2-monooxygenas, 3,6 diketocamphane 1,6-monooxygenase and 2-oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-CoA 1,2-monooxygenase (Ougham *et al.*, 1983; Taylor and Trugill, 1986; and Jones *et al.*, 1993). These Baeyer-Villiger monooxygenases are enzymes inserte one atom of O_2 into a ketone creating a lactone and often only creating one particular isomer of the lactone, and as such the Baeyer-Villiger monooxygenase are having increasing uses in biotransformations (Kelly *et al.*, 1998; and Roberts and Willetts, 1993). We originally assumed that one or two of these Baeyer-Villiger monooxygenases would lie on either the left or the right-hand side of *cam* operon (see Figure 1.2 and 1.7, Chapter1), since it is often found that the genes encoding for enzymes in sequential catabolic steps are usually clustered in one or adjoining operon (van der Meer *et al.*, 1992). For an example, in TOL plasmid, thirteen genes that are responsible for the degradation of benzoate and toluate to TCA cycle intermediates are clustered in the same operon (*meta* operon) (*xylXYZLTEGFJQKIH*), and this operon is located approximately 10 kb downstream of the upper operon, which contains five genes of *xylCMABN* (Hugouvieux-Cotte-Pattat, *et al.*, 1990). Also, Chakrabarty (1971) showed that the ketolactonase gene in camphor catabolic pathway is located on the right-hand side of *cam* operon. In the study, the *Bam*HI-DNA fragment extending from the left-hand side of *cam* operon was cloned. In

addition, we used the DNA probe derived from the N-terminal sequence of 3,6-diketocamphane 1,6-monooxygenase, the enzyme in the third step of camphor catabolic pathway, to identify and clone the DNA fragment containing this Baeyer-villiger monooxygenase gene. We were able to clone the *Bam*HI DNA fragment on the left-hand side of *cam* operon on CAM plasmid. However, we have failed to clone the *Kpn*I DNA region of 3,6-diketonecamphane 1,6-monooxygenase.

6.2 Cloning of the CAM plasmid DNA

We have determined the DNA sequence of the *Bam*HI fragment on the left-hand side of *cam* operon. Yet, only 4485 bp from about 7.1 kb DNA of the *Bam*HI DNA fragment has been sequenced because the rest of the DNA contains the known *camR* and *camD*. The restriction map of the 3′ terminal of 4485 bp DNA fragment is identical to that of previous literature (Aramaki *et al.*, 1994) (see also Figure 3.8, Chapter 3). In addition, the DNA sequence of the 3′ terminal of 4484 bp DNA fragment is identical to the DNA sequence of the 3′ flanking region of the camR gene (Aramaki *et al*, 1994). This verifies that the 4485 bp DNA fragment is adjacent to the *cam* operon on the CAM plasmid. The overall G+C content of the 4485 bp DNA *Bam*HI fragment is 56.3%, similar to that of NAH plasmid (58.3%)(Harayama *et al.*, 1987), and lower than the G+C content of *P. putida* chromosome (60.7-62.5%) (Mandel, 1966).

We were able to identify the DNA fragment, which bound with the pQR277 DNA probe derived from 3,6-diketocamphane 1,6-monooxygenase. This DNA fragment is 4201 basepairs in length. As the alignment result of the pQR277 DNA probe and the 4201-bp DNA fragment, only 52.4% of the DNA sequence of the probe is identical to the 4201-bp DNA fragment (Figure 3.21, Chapter 3). The G+C content of the 4201 bp DNA fragment is 61.3%, similar to that of *P. putida* chromosomal DNA (60.18%) (Nakamura *et al.*, 1997). This indicates that the 4201 bp DNA fragment is probably the chromosomal DNA. An analysis of the genes on this 4201 bp DNA fragment will be discussed later.

6.3 Organisation of genes on the CAM plasmid

Both *P. putida* NCIMB 10007 and *Rhodococcus* sp. NCIMB 9784 can use camphor as a sole carbon and energy source. However, the camphor catabolic pathways in the *Pseudomonas* and the *Rhodococcus* are different. The camphor degradation pathway in *P. putida* NCIMB 10007 proceeds via hydroxylation, dehydrogenation and oxygenation leading to the lactone 5-oxo-1,2-campholide, which is spontaneously rearranged to 2-oxo-Δ^3-4,5,5-trimethyl cyclopentenyl acetic acid (Jones, *et al.*, 1993). In contrast, camphor catabolism by *Rhodococcus* sp. NCIMB 9784 proceeds via hydroxylation, dehydrogenation and hydrolytic cleavage of 6-oxohydroxycamphor to campholinic acid (Figure 6.1).

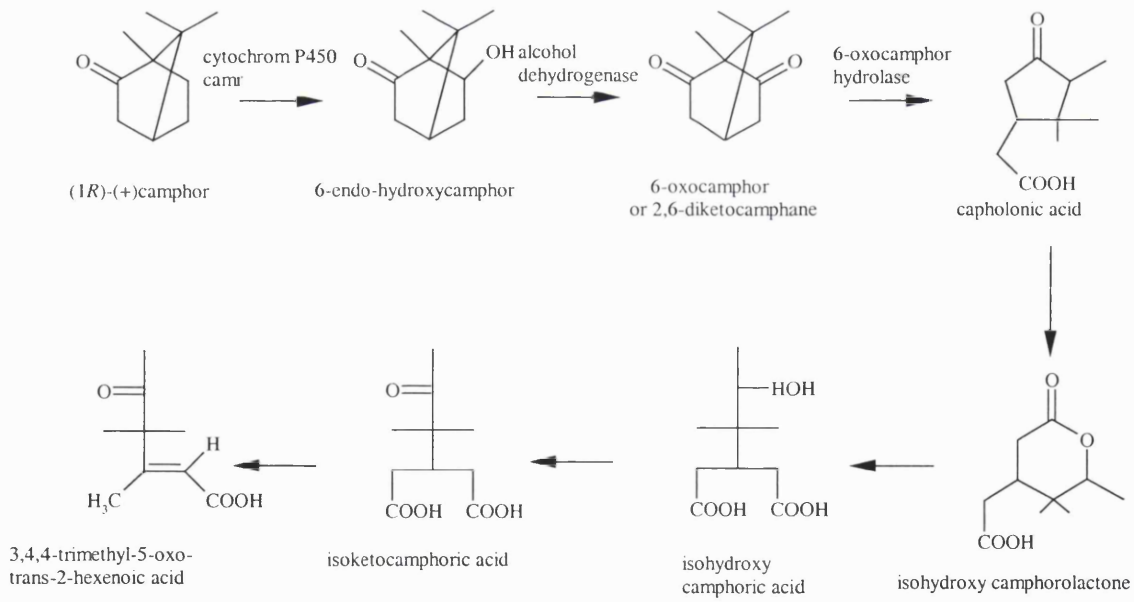


Figure 6.1 (+)-camphor degradative pathway in *Rhodococcus* sp. NCIMB 9784 (*Mycobacterium rhodochrous* strain T_1).

The biochemical and genetic studies of the camphor degradation pathway in *P. putida* are much more detailed than that in the *Rhodococcus* sp.. The organisation of known genes on the *cam* operon plus the novel genes obtained from our study is compared to that of the 6-oxo-camphor hydrolase DNA region on the chromosome of *Rhodococus* sp. NCIMB 9784 (see Figure 6.2).

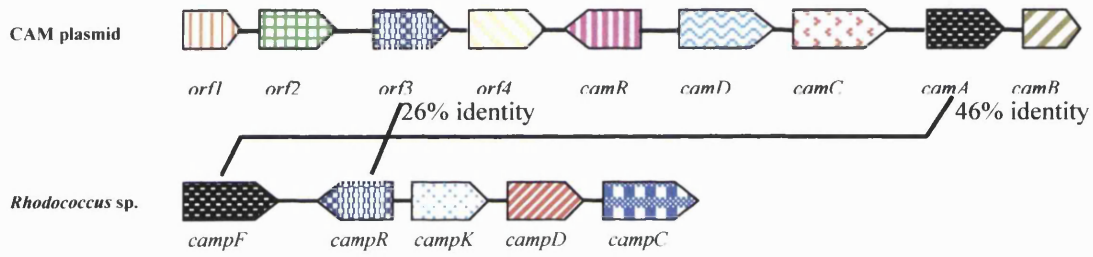


Figure 6.2 Genetic organisations of genes on the CAM plasmid of *P. putida* NCIMB 10007 and on that of the 6-oxo-camphor hydrolase gene region on the DNA chromosome of *Rhodococcus* sp. NCIMB 9784 (Grogan *et al.*, 2000b). Non-coding regions are indicated by space bars. Two genes that share homology are linked by drawn lines. *orf1*: putative Baeyer-Villiger monooxygenase gene; *orf2*: putative monooxygenase gene; *orf3*: putative regulatory gene; *orf4*: hydrolase gene; *camR*: CamR repressor gene; *camD*: 5-oxo-hydroxycamphor dehydrogenease gene; *camC*: cytochrome P450cam gene; *camA*: putidaredoxin reductase gene; *camB*: putida redoxin gene; *campF*: putative feredoxin reductase gene; *campR*: putative regulatory gene; *campK*: 5-oxo-camphor hydrolase; *campD*: putative dehydrogenase; and *campC*: putative lipid transfer protein.

Only two proteins encoded by *orf3* and *camA* from the CAM plasmid that share homology with proteins encoded from the camphor genes of *Rhodococcus* sp. NCIMB 9784. The deduced amino acid of *orf3*, a putative regulatory protein, shares 26% of its

identity to the deduced amino acid of *campR* on the camphor gene region of *Rhodococcus* sp. NCIMB 9784. The CampF protein is found to have homology (46%) with the putidaredoxin reductase encoded by *camA* (Grogan et al, 2001b). Both the CAM plasmid and the *Rhodococcus* sp. NCIMB 9784 gene cluster contain enzymes that catalyse similar reactions *i.e.* 5-*exo*-hydroxycamphor dehydrogenase (*camD*) and putative dehydrogenase (*campD*); putidaredoxin reductase (*camA*) and ferredoxin reductase (*campF*); and Orf4 hydrolase (the coding protein of *orf4*) and 6-oxo-camphor hydrolase (*campK*).

6.4 Analysis of the *orf1*

orf1 is gene on the far left of 4485 bp DNA fragment on the CAM plasmid. This ORF is 684 basepairs in length, and its deduced amino acid sequence is 227 residues. There are two inverted repeat sequences of **GCGAAC**GCAC**GTTCGC**, with the free energy of –5.7 kcal and **GCACGTTCGC**TCTCC**GCGAACGTGC**, with the free energy of –12.8 kcal, downstream of the *orf1*. However, the one that can lead to efficient transcription termination of the *orf1* would be the latter. Because the latter inverted repeat is also proceeded by a 3′ AT rich region (AACCAATAA), this could lead to rho-idependent transcription termination of the *orf1*. However, the significance of these inverted repeats is not yet fully understood. Although the *orf1* is incomplete, we expected that this gene would encode for a protein of about 540-550 amino acids, which is speculated from the multiple alignment of the deduced amino acid of *orf1* with the known proteins (Figure 4.3, Chapter 4). The Orf1 protein can be classified as Baeyer-Villiger monooxygenase, in the same class as the NADPH dependent cyclohexanone 1,2-monooxygenase from *Acinetobacter* sp. NCIMB 9871. It is also assumed that the Orf1 protein is a novel Bayer-Villiger monooxygenase, neither 2,5-diketocamphane 1,2-monooxygenase nor 3,6-diketocomphane 1,6-monooxygenase. The reason for this is that the probable deduced amino acid of *orf1* is around 60 kDa (for a protein of about 540-550 amino acid residues), which is much higher than that of 2,5-and 3,6-diketocamphane monooxygenase (38-40 kDa). The C-terminal sequence of the Orf1 protein contains a highly conserved domain, ATG-motif, FAD and NAD(P)H-binding domain. This ATG-

motif on the Orf1 protein is DVIIY**ATG,** similar to the consensus sequence of oohhh**ATG** (o is any of D, E, K, R, H, N, S, P or A; and h is any of I, V, L, F, Y or A) presenting in flavoproteins (Vallon *et al.*, 2000).

We were able to use the 227-deduced amino acid of *orf1* in a sequence homology search against proteins in the GenBank database. The result recently showed that the most highly related protein to the deduced amino acid of *orf1* is the steroid monooxygenase from *R. rhodochrous* (43% identity). The steroid monooxygenase from *R. rhodocochrous* is a protein of 60.1 kDa (549 amino acids) encoded by the *smo* gene 1,650 nucleotides in length. This protein is an NADH dependent Baeyer-Villiger monooxygenase catalysing the insertion of an oxygen atom between the C_{17} and C_{20}-carbons of progesterone to form testosterone acetate (Morii *et al.*, 1999 and Miyamoto *et al.*, 1995).

6.5 Analysis of the *orf2*

The *orf2* consists of 1092 nucleotides or encodes a protein of 363 amino acid residues. This ORF is located 88 basepairs downstream of the *orf1*. The start codon (ATG) and stop codon (TGA) of *orf2* are universal codons, as are most used bacterial start and stop codons. A ribosome binding site, AGGACA, 6 base pair upstream of *orf2*, shares a common feature with the ribosome binding site of the *camB*, GAGGA (Peterson *et al.*, 1990). No potential promoter sequences have been found on the upstream region between *orf1* and *orf2*, therefore, the transcription control of the *orf2* is not yet fully understood. With the free energy of –11.9 kcal, there is a tendency that the inverted repeat sequence **CCACGCCGC**TAT**GCGGCGTGG** 27 nucleotides downstream of the *orf2* could lead to efficient transcription termination for the *orf2*. The inverted repeat downstream of the *orf2* is also followed by a short ATT and TCTCTC, very similar to the 3′ region of the inverted repeat sequence of *camB* (an AT followed by TCTCTC sequence) (Peterson *et al.*, 1990). The G+C content of *orf2* was found to be 59.25%, very similar to that of the *camC* gene (59.0%) (Unger *et al*, 1986). The high G+C

distribution at the third letter of *orf2* genetic codes (76.92%) shows biased codon usage. This is similar to the G+C content in the third codon position of genetic codes of *Pseudomonas* sp. Cam-1 (76.79%) and *P. putida* (73.49%), yet significantly lower than that of *P. aeruginosa* (86.30%) (CUTG database) (Nakamura *et al.*, 1997). However, the codon usage of *orf2* showed that it preferred T to C in the third codon position for His and Cys. The amino acid composition of Orf2 protein is also similar to that of proteins from *E. coli* (Doolitle, 1986).

The *orf2* encodes a protein of 363 amino acids with a theoretical protein molecular weight of 40,704. The conserved domain of FMN-binding bacterial luciferase, a luciferase-like monooxygenase, found on the Orf2 protein implies a common evolutionary origin of this Orf2 protein with proteins in the luciferase and luciferase-like monooxygenase family.

A closely related protein to the Orf2 protein is the limonene 1,2-monooxygenase, NADH and FAD-dependent enzyme catalysing (+)-(4R)-limonene to (R)-limonene 1,2-epoxide from *R. Erythropolis* (van der Warf *et al*, 1999) (see also Figure 6.3). The limonene 1,2-monooxygenase is an oxidation enzyme which oxidases the C=C bond of (+)-(4R)-limonene to form an epoxide group (c  c). Whether the Orf2 protein can perform a catalytic reaction in similar pattern to the limonene 1,2-monooxygenase, i.e. in the oxidation of the C=C bond of (+)-(4R)-limonene, and with NADH and FAD as cofactors, it is also subjected to experiment, which is discussed later.

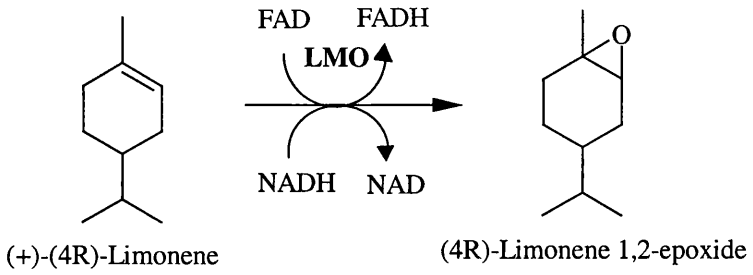


Figure 6.3 Oxygen insertion of (+)-(4R)-limonene to form (4R)-limonene 1,2-epoxide by limonene monooxygense (LMO) from *R. erythropolis* (van der Warf *et al.*, 1999).

The Orf2 protein also shares its identity with several luciferase and putative luciferase proteins. The luciferase is dimeric enzyme and consists of α and β subunit with molecular weight of 40-45 and 35-40 kDa (depending on bacteria species) respectively (Meighen, 1991). The gene of the β subunit of luciferase is believed to evolve from the α-subunit by gene duplication (Kane and Prasher, 1992). The bioluminescence reaction involves in the oxidation of aldehyde and requires $FMNH_2$, as an electron donor, and O_2 to yield carboxyl product of the aldehyde (Figure 6.4).

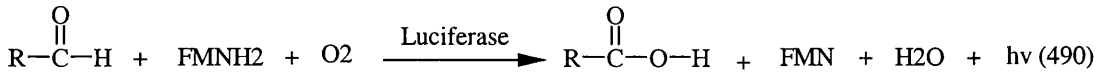


Figure 6.4 The oxidation reaction of aldehyde to the corresponding carboxyl product by a luciferase enzyme (Fisher *et al.*, 1996).

Most of the predicted structures of Orf2 protein appear to coincide with the protein secondary structure of the luciferase-β-subunit from *V. have*yi (Fisher *et al.*, 1996). The bacterial luciferase is α-β heterodimer, consisting of α-and β-subunit, which has a topology of $(\alpha\beta/)_8$ barrel structure similar to a triose phosphate isomerase (TIM). For enzymes with a barrel structure, it is known that the active sites reside at the C-terminal end of the β-barrel (Farber *et al.*, 1990). For the luciferase-β-subunit from *V. have*yi, similarly, the active site is located in the α subunit's β-barrel (Fisher *et al.*, 1996). The role of luciferase-α-subunit is thought to be an essential for high enzyme activity (Baldwin and Ziegler, 1992). The coincidence between the predicted protein secondary structure of Orf2 protein and the known secondary structure of the luciferase-β-subunit indicates a similar topology between these two proteins.

6.6 Analysis of the *orf3*

The *orf3* is a gene of 645 basepairs in length, and it encodes a 23.6 kDa protein. The *orf3* is the smallest ORF on the 4485 bp DNA fragment. This gene is located down stream of the *orf2* and transcribed in the same direction as *orf1* and *orf2*. Although there is a long intercistronic region between *orf3* and *orf4* (735 basepairs), no coding region that possibly codes for an active protein is found (as described in section 3.2, Chapter 3). The potential promoter sequences of TTAGAC (-35 region) and TATTAT (-10 region) upstream of the *orf3* showed some degrees of similarity to that of *E. coli*, which has the promoter sequences of TTGACA (-35 region) and TATAAT (-10 region) (Rosenberg, 1979). The almost perfect inverted repeat of **CAACTCTC**GG**GAGGTTTG** between the promoter regions and the initiation codon of *orf2* may function as the operator for the *orf3*. With regard to the G+C content, the G+C content of *orf3* (50.70%) is significantly less than that of other genes on the CAM plasmid. The G+C content of *camR*, *camD*, *camC*, *camA* and *camB* is 59.32%, 61.69%, 58.97%, 58.43% and 57.10% respectively. However, the G+C content of *orf3* is closer to that of *E. coli* chromosome (50-51%) (Nakamura, 1996). The A or T nucleotide in the third codon position of *orf3* genetic codes is preferred by several codons, *viz.* Asp, Val, Ala, Try, Phe, Ser and Gln. This

pattern gives a low G+C content in the third codon position (52.57%) for the *orf3*. This may imply that the *orf3* is probably from a source, which is different from other genes on the CAM plasmid and integrated into the CAM plasmid by gene transfer processes.

The conserved domain of TetR, a bacterial regulatory protein in the TetR family, is seen on the Orf3 amino acid and this indicates that the Orf3 protein shares a common ancestor with TetR regulatory proteins. When the deduced amino acid of *orf3* is used in a BLAST protein-protein search, it is revealed that the Orf3 protein is related to the transcriptional regulator of *campR* (26% identity), the putative transcriptional regulator of 6-oxo-camphor hydrolase (*Rhodococcus* sp. NCIMB 9784); AmtR (28% identity), the transcriptional repressor in the nitrogen regulation system (*C. glutamicum*), and ScbR (28% identity), the protein which binds to γ-butylolactone, the signal molecule in antibiotic synthesis (*S. coelicolor*) (see also section 4.4.5, Chapter 4).

The putative transcriptional regulator CampR protein is the closest homologous protein to the Orf3 protein. *CampR* has recently been identified as an adjoining gene along with the 6-oxo-camphor hydrolase gene (*camK*). *CampR* is on the upstream region of *camK*, transcribed in the opposite direction to *camK* (see also Figure 6.2) (Grogan *et al.*, 2001). There is; however, no relation between the Orf3 protein and the *camR* protein (cam repressor on the *cam* operon).

The similarities between the AmtR and ScbR protein are they both are dimeric and repressors. The AmtR repressor binds to the AmtR-binding motif and represses the transcription of the *amt* gene and *amtB-glnK-glnD* operon, the key regulation genes of the ammonium uptake system in *C. glutamicum* (Jakoby *et al.*, 2000). The ScbR repressor is its own negative regulator, binding to its transcriptional start site. With an addition of γ-butyrolactone SCB1 (2*R*, 3*R*, 1′*R*)-2-(1′-hydroxy-6-methylheptyl)-3-hydroxymethylbutanolide), it is shown that the ScbR repressor lost it ability to bind to DNA (Tanaka *et al.*, 2001).

At the N-terminus of the Orf3 protein, the polypeptide sequence of LESIADALGVTKQFIYSR has the characteristics of helix-turn-helix (HTH) motif, a DNA-binding domain. A common structure of HTH motif consists of an α-helix, a turn, and a second α-helix. Glycine and hydrophobic amino acids of A, C, F, I, L, M, V and Y are usually found in the HTH motif (Pabo, 1992). An amino acid at position five of the HTH motif is also often a G or an A (Harrison and Aggarwal, 1990). In Figure 6.5, the HTH motif of the Orf3 protein is compared to that of TetR, the tetracycline repressor (Orth *et al.*, 2000) and QacR, the multidrug binding protein from *S. aureus* (Schumacher *et al.*, 2002).

```
                 .  ..  ....   ...
Orf3      43  LESIADALGVTKQFIYSR  60
QacR      25  TGEIVKLSESSKGNLYYH  42
TetR      27  TRKLAQKLGIGQPTLYWH  44
```

Helix-turn-helix

Figure 6.5 A comparison of the helix-turn-helix motifs in Orf3, QacR and TetR. The amino acid at position 5 is underlined, and identical amino acids are shaded. The amino acids of A, C, F, G, I, L, M, V and Y (Pabo, 1992) on the HTH motif of the Orf3 protein are marked by dots.

In the prediction of the protein secondary structure of the deduced amino acid of Orf3, however, the polypeptide of FIYSR (the second helice) was predicted to be a β-sheet. The prediction is contrary to the HTH motif prediction. It remains to be determined whether the predicted HTH motif of the Orf3 protein is correct.

6.7 Analysis of the *orf4*

The *orf4* is a gene (984 nucleotides) on the far right of the 4485 bp DNA fragment. It is located 21 basepairs down stream of the *camR*. This gene is transcribed in the same direction as *orf1*, *orf2* and *orf3*, but in the opposite direction to the *camR*. The preference of ATG and TGA as start and stop codons in the *orf4* is similar to that of *P. aeruginosa* genes (Nakamura *et al.*, 1996). The *orf4* has the potential promoter sequences of TATAGC (-10 region) and TTCTCA (-35 region), separated by 14 nucleotides, 116 nucleotides upstream of the initiation codon of *orf4*. The position of these promoter sequences is found to be quite distant from the start codon of *orf4*. A similar range is also observed in the *camR* promoter sequences which are located 223 nucleotides upstream of the initiation codon of *camR* (Fujita *et al.*, 1993 and Aramaki *et al.*, 1994). An inverted repeat sequence of **TCAAC**CATATT**GTTGA**, probably a operator for *orf4*, positions 2 nucleotides upstream of the –35 region promoter. This putative operator, found upstream of the *orf4*, could indicate that the transcription of the gene is controlled by a specific regulatory protein. Such an operator could be a binding site of a repressor protein, which then prevents RNA polymerase to bind to the promoter of *orf4*. The putative ribosome binding site of *orf4*, GAGAA, 6 basepairs upstream of the start codon of *orf4*, shares a slight similarity to that of the *camC* gene, AGGAGA (Unger *et al.*, 1986).

The G+C content of the *orf4* (58.64%) is similar to that of *camC* (59.0%) (Unger *et al.*, 1986). The codon usage of the *orf4* is similar to that of *P. putida* genes , which prefer the G+C nucleotide in the third codon position (Nakamura *et al.*, 1997). G+C contents of the third position for *orf4* and *P. putida* genes are 65.55% and 73.49% respectively.

The conserved domain of β-lactamase of the metallo-β-lactamase family, indicates an evolutionary relationship of the Orf4 protein with proteins in the metallo-beta-lactamase family. A sequence homology search showed that the Orf4 protein is strongly related to methyl parathion hydrolase (*Plesiomonas* sp. M6) (40% identity) and

methyl parathion degrading protein (*Plesiomonas* sp. DLL1) (39% identity). Studies of organophosphate hydrolases, such as parathion hydrolase encoded by *opd* (organophosphate-degrading gene) from *P. diminuta* GM and *Flavobacterium* sp. ATCC 27551 and methyl parathion hydrolase encoded by the *mpd* (methyl parathion degrading gene) from *Plesiomonas* sp. strain M6, have been described previously (McDaniel *et al.*, 1988; Mulbry *et al.*, 1986; and Zhongli *et al.*, 2001). The Opd enzyme of *P. diminuta* GM has a molecular mass of 35.4 kDa. However, the active enzyme has molecular weight of 60.0 kDa, suggesting that the holoenzyme is dimeric. This enzyme is a membrane bound enzyme (McDaniel *et al.*, 1988). The methyl parathion hydrolase from *Plesiomonas* sp. M6 has a molecular mass of 35.0 kDa. The deduced amino acid of the *mpd* from the *Plesiomonas* sp. M6 has no homology to that of the *opd* from *P. diminuta* GM or *Flavobacterium* sp. ATCC 27551. However, it shares homology with the hypothetical beta-lactamase SCJ21.1 from *S. coelicolor* with 31% identity.

In the deduced amino acid of *orf4*, the amino acid sequence of HLHPDH, the putative zinc-binding motif of metallo hydrolase and beta-lactamase proteins, is in agreement with the consensus sequence of the zinc-binding motif of HxHxDH (x is any amino acid) (Dong *et al.*, 1999 and Melino, 1998). This HxHxDH pattern is highly conserved in the enzymes glyoxylase II, arylsulfatases and AHL lactonases (Dong *et al.*, 2000).

An AHL lactonase (AiiA) from *Bacillus* sp. strain 240B1 has been recently reported (Dong *et al.*, 2000; Dong *et al.*, 2001a and Dong *et al.*, 2001b). This AHL lactonase catalyse the hydrolysis of the ester bond in the lactone ring of several AHLs. The AiiA lactonase has two highly conserved motifs of $_{104}$HLHFDHAG$_{111}$ (first motif) and $_{165}$HTPGHTPGH$_{178}$ (second motif). Site-directed mutagenesis of amino acids in the first motif demonstrated that this motif is important for its enzyme activity (Dong *et al*, 2000). In Figure 6.6, the active site of AiiA lactonase is compared to the conserved domain of proteins in the metallo-beta-lactamase family and to our Orf4 protein.

```
Conserved   48 IILTHAHADHIGGVPEL 64 -----112 EVIHTPGHTPGSIVY 126
AiiA       100 IISSHLHFDHAGGNGAF 116-----162 QLLHTPGHTPGHQSL 176
Orf4       140 VLLTHLHPDHACGLANA 156-----222 ESVPTYGHTPGHSAY 236
                   HxHxDHxxG                     TxGHTPG
```

Figure 6.6 Alignment comparison of the active site of AiiA (underlined) to the Orf4 protein and the conserved domain of proteins in the metallo-beta-lactamase family (from the Conserved domain database). Identical amino acids are shaded, and the highly conserved amino acids are shown below the alignment (x is any amino acid).

A predicted protein secondary structure (using DSC, PHD and Jpred method) showed that the Orf4 structure is comparable with the α/β barrel structure of L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* (Ullah *et al.*, 1998). The probable zinc-binding motif (HLHPDH) of Orf4 protein is located in the loop region between the β_3-strand and the α_8-helice. This characteristic is similar to the active site of all proteins with a α/β barrel structure, which is formed by loops between β-strands and α-helices at the end of the barrel (Branden and Tooze, 1999).

6.8 Analysis of the *orf5*, *orf6*, *orf7* and *orf8*

The 4201 bp *Kpn*I DNA fragment is the DNA fragment that is identified by pQR277 probe (0.7 kb DNA fragment derived from the N-terminal sequence of 3,6-diketocamphane 1,6-monooxygenase gene). The high G+C content (61.3%) of 4201 bp *Kpn*I DNA fragment represents a characteristic of chromosomal DNA in *P. putida* (60.7-62.5 %) (Mandel, 1969). Open reading frame analysis of this DNA fragment showed that the DNA contains 4 open reading frames, designated *orf5*, *orf6*, *orf7* and *orf8*. The sequence homology search showed that the deduced amino acids from *orf5*, *orf6*, *orf7* and *orf8* are similar to several membrane proteins in the recent GenBank protein

database, but none have homology to proteins in the Baeyer-Villiger monooxygenase family. In addition, a search for a similar DNA sequence to the 4201 bp *Kpn*I fragment in unfinished genome databases (http://tigrblast.tigr.org) revealed that this DNA fragment is homologous to the DNA contig 13538 of *P. putida* KT 2400 (85% identity). In the experiment, which was used to identified a restriction fragment of a total DNA-*Kpn*I digest of *P. putida* NCIMB 10007, the Southern hybridisation using pQR277 might hybridise to a partially homologous DNA or DNA with a highly similar sequence from the chromosomal DNA of *P. putida* NCIMB 10007. In spite of the deduced amino acid of *orf5*, *orf6*, *orf7* and *orf8* having similarities to several proteins (as listed in Table 4.11, Chapter 4), further investigation of these genes has not been carried out.

6.9 Characteristics of the Orf2 monooxygenase

The Orf2 monooxygenase has been shown in the Chapter 5 to be expressed to a high level in *E. coli*. The molecular mass of Orf2 monooxygenase seen on SDS-PAGE is approximately 40 kDa, consistent with the predicted molecular weight mass from the deduced amino acid (40,704 Da). We were able to purify the His-tagged Orf2 monooxygenase in a single step to apparent homogeneity, and the purified His-tagged Orf2 monooxygenase exhibited an absorption spectrum at 221 nm. This absorbance shows helical peptides exposure of the Orf2 monooxygenase in the solvent. This characteristic is shown in proteins with predominant α-helical elements (Monera *et al.*, 1993; and Horn *et al.*, 1999).

As the deduced amino acid of *orf2* is closely related to the limonene monooxygenase from *R. erythropolis*, we conducted an experiment to test whether this protein can perform a reaction similar to that of the limonene monooxygenase. Enzyme assays of ORF2 monooxygenase with (+)-limonene, (-)-limonene, (+)-pinene, (-)-pinene, cyclohexane and cyclopentanone combined with NADH or NADPH showed that none of these compounds can serve as a substrate for the Orf2 monooxygenase.

With regard to the N-terminal sequence, the N-terminal sequence of Orf2 monooxygenase is highly homologous (80% identity) to that of 2,5-diketocamphane 1,2-monooxygenase. The 2,5-diketocamphane 1,2-monooxygenase is a complex consisting of two proteins: a FMN-binding oxygenating component and NADH dehydrogenase (Taylor and Trudgill, 1986). In the oxygenation of 2,5-diketocamphane, NADH dehydrogenase reduces NADH to NAD^+ and transports electrons to the FMN bound 2,5-diketocamphane 1,2-monoxygenase. This enzyme-cofactor complex reacts with O_2 to form an enzyme-hydrogen peroxide, which then inserts one oxygen between C1 and C2 of 2,5-diketocamphane to form 5-oxo-campholide lactone (see also Figure 1.5, Chapter 1). This might imply that the Orf2 monooxygenase requires another protein component to act as an active complex in the lactonisation of its substrate.

6.10 Characteristics of the Orf4 hydrolase

The expression level of the *orf4* in *E. coli* was very low. This indicates the restricted expression of the *orf4* in *E. coli*. The restricted expression of the *Pseudomonas* genes in *E. coli* was also reported when the *trpAB* (tryptophan genes from *P. aeruginosa*) was cloned into *E. coli* (Hedges *et al.*, 1977). However, we were able to improve the expression level of the *orf4* in *E. coli* by introducing the *orf4* into *E. coli* containing the extra *argU* and *pro*L tRNA genes. This indicates the different codon usage patterns of the *orf4* and *E. coli* genes for Arg and Pro. This difference can be seen as the difference between the number of AGA and AGG codons in *E. coli* genes and the *orf4*. The number of AGA and AGG codons are 3-4 per 1,000 codons in *E. coli* genes and 3 (AGA) and 12 (AGG) per 1,000 codons in the *orf4* (see also section 4.5.2, Chapter 4).

The SDS-PAGE of Orf4 hydrolase showed that the protein has a molecular mass of about 36 kDa, consistent with the predicted molecular weight from the deduced amino acid (35,688 Da).

In the attempt to purify the His-tagged Orf4 hydrolase, we failed to obtain the purified His-tagged Orf4 hydrolase. We found that there was a silent mutation during the PCR amplification of *orf4*, which altered a G at the position 743 to an A. This mutation then altered a tryptophan (W248) of Orf4 hydrolase to a stop codon (TAG), which then generated a corresponding protein of about 27 kDa, as can be seen in Figure 5.8 and 5.9, Chapter 5. This is the reason that we failed to purify the His-tagged Orf4 hydrolase.

As we could not isolate a pure Orf4 hydrolase, we carried out some preliminary characterisation of the Orf4 hydrolase native recombinant enzyme in isolated fractions of whole cells. The soluble and membrane fractions of *E. coli* BL21(DE3)CodonPlus-RP harbouring pQRX7 were separated from the disrupted cells and run on the SDS-PAGE gel. By monitoring the protein on the SDS-PAGE, it is found that the native Orf4 hydrolase is associated with the membrane fraction. Morover, the molecular mass of about 36 kDa of the Orf4 hydrolase on SDS-PAGE is consistent with the predicted molecular mass (from the nucleotide sequence) of 35,688 Da.

The prediction of PSORT (Nakai and Hortan, 1999) showed that Orf4 hydrolase is localised to the periplasmic space. As it is shown on the SDS-PAGE gel that the molecular mass of Orf4 hydrolase is 36 kDa, consistent with that of the theoretical molecular mass from the deduced amino acid, this indicates that the Orf4 hydrolase is not modified in the *E. coli*. Similarly, the membrane bound parathion hydrolase from *Flavobacterium* sp. ATCC 27551 is not recognised by signal peptidases when it is expressed in *E. coli* (Mulbry and Karns, 1989). However, when this parathion hydrolase is expressed in gram-positive bacteria, *S. lividans*, it is found that the protein is secreted out of the cells (Steiert *et al.*, 1989).

The 29-amino acid region at the N-terminus of Orf4 hydrolase exhibits characteristics for a signal peptide (von Heijne, 1988). The positively charged N-terminal of Orf4 hydrolase (MRKFRS) comprises three charged amino acids: 2R and K. The central hydrophobic region (FAFQLTLVTVTVGCGM) is dominated by the hydrophobic amino acids of A, F, L, M and V. This central hydrophobic peptide is also

predicted as a transmembrane region by the transmembrane topology prediction of TMHMM (Sonnhammer *et al*, 1998). The hydrophilic C-terminal (NTIP) of the leader peptide comprises the hydrophilic amino acids of N and T. The cleavage site contains a predicted cleavage site of AIA-EP (by the SignalP program) (see Figure 6.7). The pattern for the cleavage by signal peptidase is identical to a consensus pattern specific for the cleavage by signal peptidase type I (AXA) (X is a large amino acid) (Neilsen *et al.*, 1997). The length of the leader peptide of Orf4 hydrolase (29 residues) also identical to that of the parathion hydrolase from *Flavobacterium* sp. ATCC 27551, and in similar length with the average gram-positive signal peptide (32.0 residues). However, it is unlike Gram-negative and eukaryotic signal peptides, which have the average length of a signal peptide of 25.1 and 22.6 residues respectively (Nielsen *et al.*, 1997).

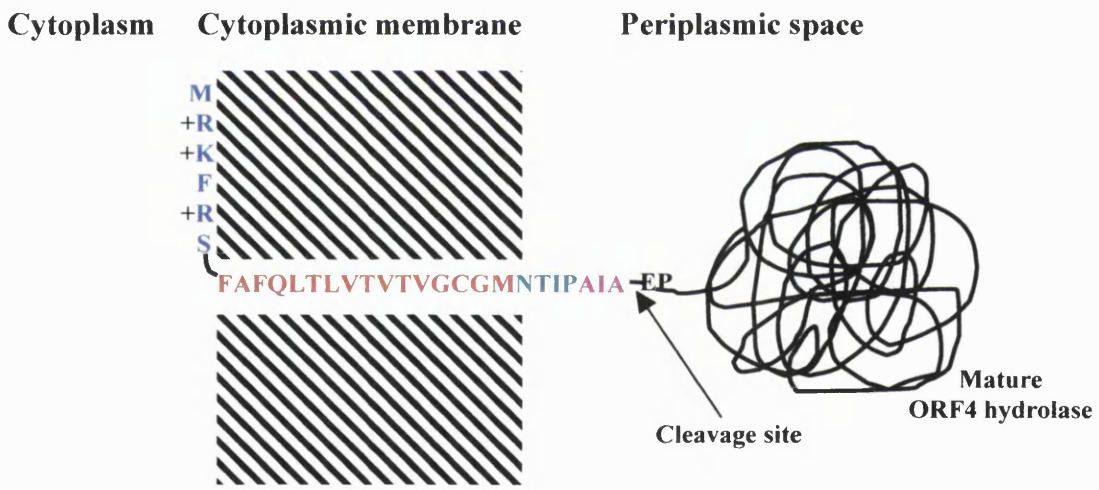


Figure 6.7 Characteristics of the signal peptide of Orf4 hydrolase.

The Orf4 hydrolase was able to hydrolyse our tested substrates. It can hydrolyse both γ-butyrolactone and paraoxon but with different activities (Figure 6.8). For the tested organophosphorous compounds, the Orf4 hydrolase has high activity in hydrolysing paraoxon and methyl parathion. The initial hydrolysis rates of paraoxon

(13.0 μmol/h/mg-dry cell weight) and methyl parathion (4.2 μmol/h/mg-dry cell weight) by a whole cell of *E. coli* BL21(DE3)CodonPlus-RP harbouring pET21a-*orf4* are faster than the rate of 9.0 and 0.6 μmol/h/mg-dry cell weight of surfaced-expressed organophosphorous hydrolase *Moraxella* sp. cells (Shimazu *et al*, 2001). The Orf4 hydrolase has low activity towards parathion. The initial hydrolysis rate for parathion in the incubation with *E. coli* BL21(DE3)CodonPlus-RP harbouring pQR424 is only 1.0 μmol/h/mg-dry cell weight. It is noteworthy that the ORF4 hydrolase prefers paraoxon to methyl parathion and parathion, respectively, as a substrate. This preference in the hydrolysis of organophosphorous compounds is different from the organophosphorous hydrolase from *P. diminuta*, which prefers hydrolysing paraoxon to parathion and methyl parathion respectively (Dumas *et al.*, 1993).

There has been a report of lactonohydrolase from *Fusarium oxysporum* AKU 3702, which has a protein sequence similar to the gluconolactonase from *Zymomonas mobilis*, human serum paraoxonase and strictosidine synthase from *Catharanthus roseus* (Kobayashi *et al.*, 1998). Although the lactonohydrolase from *F. oxysporum* AKU 3702 showed a protein sequence similar to the paraoxonase, no activity of this hydrolase towards paraoxon has been detected. By contrast, *orf4* has a 40% sequence identity to the parathion hydrolase from *Plesiomonas* sp. M6 and the enzyme is also capable of hydrolysing paraoxon and methyl parathion in high activity.

For lactones, the Orf4 hydrolase is capable of hydrolysing γ-butyrolactone, but not γ-caprolactone, gluconolactone, pantolactone and δ-valerolactone. The initial hydrolysis rate for γ-butyrolactone is 84.2 μmol/h/mg-dry cell weight which is considered very high. Because there is the high activity towards γ-butyrolactone compared to paraoxon of the Orf4 hydrolase, it may appear that a compound similar to γ-butyrolactone may be the primary substrate for the enzyme. It is unlikely that γ-butyrolactone is a natural substrate for the Orf4 hydrolase. Predicted compounds that would serve as a substrate for Orf4 hydrolases are such as 1,2-campholide, pimelyl lactone and the other lactone intermediates in the (+) or (-) camphor degradative pahtway (see also Figure 1.2, Chapter 1), which are compounds structurally similar to γ-butyrolactone.

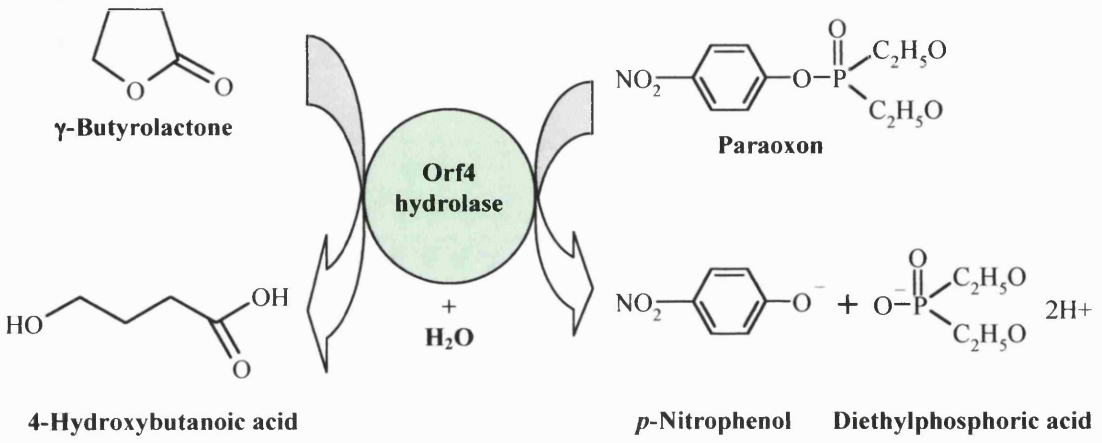


Figure 6.8 Hydrolysis of γ-butyrolactone and paraoxon by the Orf4 hydrolase. Either γ-butyrolactone or paraoxon can serve as a substrate for the Orf4 hydrolase.

A γ-butyrolactone has also been reported as an intermediate detected in the degradation of 1-chlorobutane and 1-chlorohexadecane by *R. rhodochrous* NCIMB 13064 (Curragh *et al.*, 1994). This intermediate compound is consequently metabolised to 4-hydroxybutyric acid and succinic acid. A γ-Butyrolactone and 4-hydroxybutyric acid is also implicated in the degradation pathway of tetrahydrofuran by *Rhodococcus* sp. (Berhardt and Diekmann, 1991).

One γ-butyrolactone derivative is *N*-Acyl-homoserine lactone (AHL). An AHL molecule has a structure of cyclic lactone bridged with 4-14 carbon acyl side chains by an amide bond (Pearson *et al*, 1995). The AHL is a signal molecule, also known as autoinducers (AI), involved in the induction of virulence genes in *P. aeruginosa* and *Erwinia carotovara*, conjugation transfer of Ti plasmid in *Agrobacterium tumefaciens,* bioluminescence in *Vibrio* species and antibiotic production in *E. carotovara* (Robson *et al.*, 1997). The AHL lactonase from *Bacillus* sp. 240B1 has been described recently

(Dong *et al.*, 2000). This AHL lactonase is an enzyme capable of hydrolysing *N*-(3-oxohexanoyl)-L-homoserine lactone to form *N*-(3-oxohexanoyl)-L-homoserine (Figure 6.9), and also able to cleave other AHL signal molecules (Dong *et al.*, 2001).

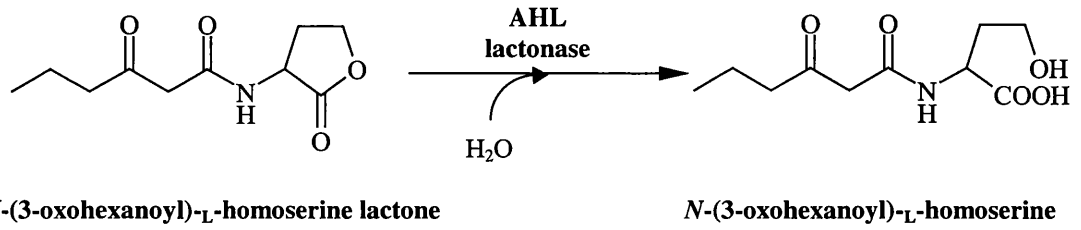


Figure 6.9 Hydrolysis of the *N*-(3-oxohexanoyl)-L-homoserine lactone to *N*-(3-oxohexanoyl)-L-homoserine by AHL lactonase from *Bacillus* sp. 240B1.

To test whether *P. putida* NCIMB 10007 can use γ-butyrolactone or paraoxon as a sole carbon source, *P. putida* NCIMB 10007 was incubated in M9 medium supplemented with glucose and γ-butyrolactone or paraoxon, and the growth rate was monitored for 7 days. However, no growth was observed when the *P. putida* NCIMB 10007 was grown on either γ-butyrolactone or paraoxon as a sole carbon source. Therefore, in spite of the high activity in hydrolysing γ-butyrolactone and paraoxon, these compounds are not a substrate for the *P. putida* NCIMB 10007.

In our study, the Orf4 hydrolase was tested with limited lactones due to very few lactones being available commercially. Also, it would be very interesting to test AHL as a substrate for the Orf4 hydrolase because AHL is structurally related to γ-butyrolactone and it has previously been shown that the Orf4 hydrolase has a central polypeptide sequence similar to the highly conserved amino acids of the active site of the AHL lactonase from *Bacillus* sp. 240B1.

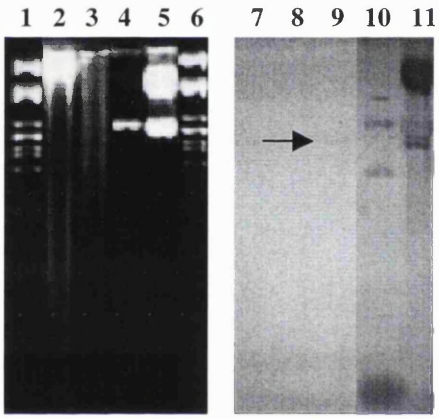
To summarise, we have achieved the cloning of genes on the left-hand side of *cam* operon. However, we failed to clone the 3,6-diketocamphane 1,6-monooxygenase gene region by using the pQR277 probe derived from the 3,6-diketocamphane 1,6-monooxygenase sequence. The complete gene sequences of the monooxygenase gene (*orf2*), the regulatory gene (*orf3*), the hydrolase gene (*orf4*), and the partial DNA sequence of novel Baeyer-Villiger monooxygenase gene (*orf1*) were determined. The Orf2 monooxygenase and Orf4 hydrolase were overexpressed, and their characteristics were determined. The Orf4 hydrolase is characterised as a membrane associated protein. This Orf4 hydrolase is not only capable of hydrolysing paraoxon, but also γ-butyrolactone.

6.11 Future work

In order to obtain the remaining DNA region of *orf1* coding for a novel Baeyer-Villiger monooxygenase, the fragment of 192 bp *Eco*RV-*Bam*HI at the end of the *orf1* nucleotide sequence was cloned and ligated into a pUC19 vector. This 192 bp *Eco*RV-*Bam*HI fragment was subjected to a DIG DNA labelling procedure. In the preliminary experiment, the restriction endonuclease of *Nar*I was used to digest the genomic DNA from *P. putida* NCIMB 10007. The Southern hybridisation of genomic DNA-*Nar*I digest probed with 192 bp *Eco*RV-*Bam*HI DNA probe showed a positive band at about 2.2 kb (see Figure 6.10). For further study, cloning of genes on the right-hand side of the cytochrome P450cam operon is also important. We might discover the Baeyer-Villiger monooxygenase genes that we have not yet found on this side.

We have also cloned and constructed a whole *cam* operon: *camR, camD, camC, camA* and *camB* in one plasmid. This construct will be used to study the relationship of the *cam* operon and our construct of *orf1234* to investigate if there is any relation between these two operons. What if an existence of the 4485 bp *Bam*HI fragment and the *cam* operon in the same cell could produce other chemicals from (+)-camphor. It is also found that there is a regulatory gene (*orf3*) on the 4485 bp *Bam*HI fragment. Therefore,

there are the questions that 1) whether this gene encodes a protein in the regulation of genes (*camRDCAB*) in the camphor degradative pathway; or 2) this gene is regulated by CamR repressor encoded by *camR* gene on the *cam* operon. For future investigation, having the *orf2* and *orf4* in an expression vector will also be useful in the characterisation of the enzymes in relation to the change in amino acids by site-directed mutagenesis.



Electrophoresis gel Southern hybridisation

Figure 6.10 Southern blotting and colorimetric detection of the total DNA of *P. putida* NCIMB 10007 *Nar*I digest hybridised with the 192 bp *Eco*RV-*Bam*HI probe. Lane 1 and 6: λ-*Pst*I DNA marker; Lane 2: *P. putida* NCIMB 10007 genomic DNA; Lane 3: *P. putida* NCIMB 10007 genomic DNA cut with *Nar*I; Lane 4: 192 bp *Eco*RV-*Bam*HI probe; Lane 5: the p138 DNA *Bam*HI (7.0 kb *Bam*HI DNA on the left hand side of *cam* operon); Lane 7-11: Southern hybridisation and colorimetric detection of the electrophoresis gel of Lane 1-5. The positive band of about 2.2 kb *Nar*I frangment is indicated by an arrow.

Appendix

Bacterial media, buffers and essential reagents

Media

SOC medium (per litre)

20 g of tryptone

5 g of Yeast extract

0.5 g of NaCl

Autoclave

Add 10 ml of 1M $MgSO_4$ (filter sterilised) and 20 ml of 20% (w/v) glucose

M9 Medium (per litre)

Make following solution;

6 g of Na_2HPO_4

3 g of KH_2PO_4

1 g of NH_4Cl

Add water to 1 litre

Autoclave

Make following solution;

1 ml of 1M $MgSO_4$

2 g of glucose

0.1 ml of 1M $CaCl_2$

1.0 ml of 1M thiamine-HCl

Add water to 10 ml

Filter sterilise

Add this solution to the above solution to make M9 media

Buffers

TBE buffer (5× stock solution)

54 g Tris-Cl

27.5 g boric acid

20 ml of 0.5 M EDTA, pH 8.0

Add water to 1 litre

(0.5× working solution)

SDS-PAGE buffer (5× stock solution)

15.1 g Tris-base

72.0 g Glycine

5.0 g SDS

Add distilled water to 1 litre

(1× working solution)

Gel loading buffer

0.25% bromophenol blue

0.25% xylene cyanol

30% glycerol in distilled water

0.5 M Tris-HCl buffer, pH 6.8 (per litre)

60.6 g Tris-HCl

Add HCl to pH 6.8

Add distilled water to 1000 ml

2×SDS-PAGE sample buffer (per 10 ml)

2.5 ml of 0.5 M Tris-HCl, pH 6.8

10% SDS

20% (v/v) glycerol

2% (v/v) 2-mercaptoethanol

0.2 mg Bromphenolblue

Add water to 10 ml

Buffer 1 (per 800 ml)

100 ml Tris-HCl, pH 7.5

8.75 g NaCl

Adjust pH to 7.5

Buffer 2 (per 100 ml)

0.5 g skimmed milk powder

Add 100 ml of Buffer 1

Dissolve at 65°C

Buffer 3 (per 400ml)

40 ml Tris-HCl, pH 9.5

2.32 g NaCl

4.08 g MgCl

Buffer 4 (per 400 ml)

3 ml Tris-HCl, pH 8.0

0.14 g EDTA

8×Binding buffer (per litre)

40 mM imidazole

4 M NaCl

160 mM Tris-HCl, pH 7.9

8×Wash buffer (per litre)

480 mM imidazole

4 M NaCl

160 mM Tris-HCl, pH 7.9

8×Elute buffer (per litre)

4 M imidazole

1 M NaCl

80 mM Tris-HCl, pH 7.9

Essential reagents

Coomassie staining solution (per litre)

500 ml (v/v) methanol

0.5 g (v/v) Coomessie brilliant blue R-250 (Biorad)

100 ml acetic acid

400 ml water

20×SSC (per litre)

173.3 g NaCl

88.2 g tri-sodium citrate

Adjust pH to 7.0

Add water to 1 litre

Depurination solution (0.2 M HCl) (per 0.5 litre)

4.55 ml HCl

Add water to 500 ml

Denaturation solution (per 0.5 litre)

10 g NaOH

43.75 g NaCl

Add water to 500 ml

Neutralisation solution (per 0.5 litre)

60.55 g Tris-HCl

43.75 g NaCl

Adjust pH to 8.0

Add water to 500 ml

Hybridisation solution (per 100 ml)

1 ml 20×SSC

0.1% L-sacosine

0.02% SDS

Add water to 100 ml

Washing solution (per 400 ml)

- low stringency solution

40 ml 20×SSC

0.1% SDS

Add water to 400 ml

-high stringency solution

1 ml 20×SSC

0.1% SDS

Add water to 400 ml

References

Adger B., Bes M.T., Grogan G., McCague R., Pedragosa-Moreau S., Roberts S.M., Villa R., Wan P.W., Willetts A.J. (1997) The synthesis of (R)-(+)-lipoic acid using a monooxygenase-catalysed biotransformation as the key step. *Bioorg. Med. Chem.* (2):253-61.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389-402.

Altschul S.F. and Koonin E.V. (1998) Iterated profile searches with PSI-BLAST a tool for discovery in protein databases. *Trends Biochem. Sci.* 11:444-7.

Aramaki H. Sagara Y., Hosoi M., and Horiuchi T. (1993a) Evidence for autoregulation of *camR*, which encodes a repressor for the cytochrome P450cam hydroxylase operon on *Pseudomonas putida* CAM plasmid. *J. Bacteriol*, 175(24): 7828-7833.

Aramaki H., Koga H., Sagara Y., Hosoi M. and Horiuchi T.(1993b) Complete nucleotide sequence of the 5-*exo*-hydroxycamphor dehydrogenase gene on the CAM plasmid of *Pseudomonas putida* (ATCC17453). *Biochim. Biophy. Acta.* 1174: 91-94.

Aramaki H., Sagara Y., Kabata H., Shimamoto N. and Horiuchi T. (1995) Purification and characterisation of camR repressor (*camR*) for the cytochrome P-450cam hydroxylase operon the *Pseudomonas putida* CAM Plasmid. *J. Bacteriol.* 177(11): 3210-3217.

Aramaki H., Sagara Y., Takeuchi K., Koga H. and Horiuchi T. (1994) Nucleotide sequence of the gene encoding a repressor for the cytochrome P-450cam hydroxylase operon on the *Pseudomonas putida* CAM plasmid. *Biochimie.* 76: 63-70.

Atlas R. M. (1997) *Handbook of microbiological media.* 2nd ed. CRC Press.

Baldwin T. O., and Ziegler, M. M. (1992) *Chemistry and biochemistry of flavoenzymes.* (Muller, F., ed) Vol. 3, CRC Press, Inc, Boca Raton, FL.

Benning M.M., Kuo J.M., Raushel F.M. and Holden H.M. (1995) Three-dimensional structure of the binuclear metal center of phosphotriesterase. *Biochemistry.* 34(25):7973-8.

Benning M.M., Hong S.B., Raushel F.M. and Holden H.M. (2000) The binding of substrate analogs to phosphotriesterase. *J. Biol. Chem.* 275(39):30556-60.

Benning M.M., Shim H., Raushel F.M. and Holden H.M. (2001) High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from *Pseudomonas diminuta. Biochemistry.* 40(9):2712-22.

Bernhardt D. and Diekmann H. (1991) Degradation of dioxane, tetrahydrofuran and other cyclic ethers by an environmental *Rhodococcus* strain. *Appl. Microbiol. Biotechnol.* 36(1):120-3.

Black M.T., Munn J.G., Allsop A.E. (1992) On the catalytic mechanism of prokaryotic leader peptidase 1. *Biochem J.* 282 (Pt 2):539-43.

Branden C. and Tooze J. (1991) *Introduction to protein structure.* Garland Pubishing, Inc. New York.

Brandshaw W.H., Conrad H. E., Corey E. J., Gunsalus I.C. and Lednicer D. (1959) Microbial degradation of (+)-camphor. *J. Am. Chemi. Soc.*, 81:5507.

Brzostowicz P.C., Blasko M.S. and Rouviere P.E. (2002) Identification of two gene clusters involved in cyclohexanone oxidation in *Brevibacterium epidermidis* strain HCU. *Appl. Microbiol. Biotechnol.* 58(6):781-9.

Bukhari A.I., Shapiro J.A. And Adhya S.L. (1977) *DNA insertion element, plasmid, and episomes.* Cold Sprinf Harbor Laboratory, USA.

Bukhari A.I. Brookhaven (1977) Transposition of DNA sequences. *Symp. Biol.* 20;(29):218-32 .

Chakrabarty A.M. and Gunsalus I.C. (1971) CAM plasmid in Pseuomonas: transfer polarity and genetic circularity. *Bacteriol. Proc.* 46.

Chakrabarty A.M. (1976) Plasmids in Pseudomonas. *Annual Review of Genetics*, 10: 7-30.

Chapman P. J., Kuo Jyh-Fa and Gunsalus I.C. (1963) Camphor oxidation: 2,6-diketocamphane pathway in Diptheroid. *Federation Proceedings.* 22: 296.

Chapman J.P., Meerman G., Gunsalus I.C., Srinivasan Rangaswamy and Rinehart K.L. (1966) A new acyclic acid metabolite in camphor oxidation. *J. Am. Chemi. Soc.* 88(3); 618-619.

Chen Y.C., Peoples O.P., Walsh C.T. (1988) *Acinetobacter* cyclohexanone monooxygenase: gene cloning and sequence determination. *J. Bacteriol.* 170(2):781-9.

Cheng T., Liu L., Wang B., Wu J., DeFrank J.J., Anderson D.M., Rastogi V.K. and Hamilton A.B. (1997) Nucleotide sequence of a gene encoding an organophosphorus nerve agent degrading enzyme from *Alteromonas haloplanktis. J. Ind. Microbiol. Biotechnol.* 18(1):49-55.

Cline T.W. and Hastings J.W. (1972) Mutationally altered bacterial luciferase. Implications for subunit functions. *Biochemistry.* 11(18):3359-70.

Conrad H.E, Lieb K. and Gunsalus I.C.(1965a) Mixed function oxidation. 3. An electron transport complex in camphor ketolactonization. *J. Biol. Chem.* 240(10):4029-37.

Conrad, H.E., DuBus, R., Namtvedt, M. J., and Gunsalus, I.C. (1965b) Mixed function oxidation II. Separation and properties of the enzymes catalysisng camphor lactonization. *J. Biol. Chem.* 240: 495-503.

Croteau R., El-Bialy H. and El-Hindawi S. (1984) Metabolism of monoterpenes: lactonization of (+)-camphor and conversion of the corresponding hydroxy acid to the glucoside-glucose ester in sage (*Salvia officinalis*). *Arch. Biochem. Biophys.* 228(2):667-80.

Cuff J.A., Clamp M.E.., Siddiqui A.S., Finlay M. and Barton G.J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics.* 14(10):892-3.

Cuff J. A. and Barton G.J (1999) Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* 40:502-511.

Curragh H., Flynn O., Larkin M.J., Stafford T.M., Hamilton J.T. and Harper D.B. (1994) Haloalkane degradation and assimilation by *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology.* 140 (Pt 6):1433-42.

Dave K.I., Miller C.E. and Wild J.R. (1993) Characterization of organophosphorus hydrolases and the genetic manipulation of the phosphotriesterase from *Pseudomonas diminuta. Chem. Biol. Interact.* 87(1-3):55-68.

DeFrank J.J. and Cheng T.C. (1991) Purification and properties of an organophosphorus acid anhydrase from a halophilic bacterial isolate. *J. Bacteriol.* 173(6):1938-43.

DeFrank J.J., Beaudry W.T., Cheng T.C., Harvey S.P., Stroup A.N. and Szafraniec L.L. (1993) Screening of halophilic bacteria and *Alteromonas* species for organophosphorus hydrolyzing enzyme activity. *Chem. Bio.l Interact.* 87(1-3):141-8.

Dong Y.H., Xu J.L., Li X.Z. and Zhang L.H. (2000) AiiA, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of Erwinia carotovora. *Pro.c Natl. Acad. Sci. U .S. A.* 97(7):3526-31.

Dong Y.H., Wang L.H., Xu J.L., Zhang H.B., Zhang X.F. and Zhang L.H. (2001) Quenching quorum-sensing-dependent bacterial infection by an N-acyl homoserine lactonase. *Nature.* 411(6839):813-7.

Dong Y.H., Gusti A.R., Zhang Q., Xu J.L. and Zhang L.H. (2002) Identification of quorum-quenching N-acyl homoserine lactonases from *Bacillus* species. *Appl. Environ. Microbiol.* 68(4):1754-9.

Doolittle F. R. (1986) *Of URFs and ORFs*; *A primer on how to analyse derived amino acid sequences*, University science books, CA. USA.

Dumas D.P., Caldwell S.R., Wild J.R. and Raushel F.M. (1989) Purification and properties of the phosphotriesterase from *Pseudomonas diminuta. J. Biol. Chem.* 264(33):19659-65.

Dumas D.P., Durst H.D., Landis W.G., Raushel F.M. and Wild J.R. (1990) Inactivation of organophosphorus nerve agents by the phosphotriesterase from *Pseudomonas diminuta. Arch. Biochem. Biophys.* 277(1):155-9.

Farber K.G. and Petsko G.A. (1990) The evolution of alpha/beta barrel enzymes. *Trends Biochem. Sci.* 15(6):228-34.

Fishbein W.N. and Bessman S.P. (1966) Purification and properties of an enzyme in human blood and rat liver microsomes catalyzing the formation and hydrolysis of gamma-lactones. I. Tissue localization, stoichiometry, specificity, distinction from esterase. *J. Biol. Chem.* 241(21):4835-41.

Fisher A.J., Thompson T.B., Thoden J.B., Baldwin T.O. and Rayment I. (1996) The 1.5-A resolution crystal structure of bacterial luciferase in low salt conditions. *J. Biol. Chem.* 271(36):21956-68.

Fujita M., Aramaki H., Horiuchi T. and Amemura A. (1993) Transcription of the cam operon and camR genes in *Pseudomonas putida* PpG1. *J. Bacteriol.* 175(21): 6953-6958.

Funk C., Koepp A.E., Croteau R. (1992) Catabolism of camphor in tissue cultures and leaf disks of common sage (*Salvia officinalis*). *Arch Biochem Biophys.* 294(1):306-13.

Gagnon R.,Grogan G., Groussain E., Pedragosa-Moreau S., Richardson FP., Roberts MS., Willetts JA., Alphand V., Lebreton J. And Furstoss R. (1995) Oxidation of some prochiral 3-substituted cyclobutanones using monooxygenase enzymes: a single-step method for the synthsis of optically enriched 3-sustituted γ-lactones. *J. Chem. Soc. Perkin. Trans.* 1:2527-2528.

Galli E., Silver S. and Witholt B. (1992) *Pseudomonas: Molecular biology and biotechnology.* American Society for Microbiology, Washington.

Gelb M.H., Malkonen P. and Sligar S.G. (1982) Cytochrome P450cam catalyzed epoxidation of dehydrocamphor. *Biochem. Biophy.s Res. Commun.* 104(3):853-8.

Gill I. and Ballesteros A. (2000) Degradation of organophosphorous nerve agents by enzyme-polymer nanocomposites: efficient biocatalytic materials for personal protection and large-scale detoxification. *Biotechnol. Bioeng.* 70(4):400-10.

Glavac D., Potocnik U., Podpecnik D., Zizek T., Smerkolj S. and Ravnik-Glavac M. (2002) Correlation of MFOLD-predicted DNA secondary structures with separation patterns obtained by capillary electrophoresis single-strand conformation polymorphism (CE-SSCP) analysis. *Hum. Mutat.* 19(4):384-94.

Gorgan G., Roberts M.S. and Willetts J.A. (1992). Biotransformation by microbial Baeyer-Villiger monooxygenase stereoselective lactone formation *in vitro* by couple enzyme systems. *Biotech. Lett.* 14(12); 1125-1130.

Gogan G., Roberts M. Stanley and Willetts J. Andrew (1993) Some Baeyer-Villiger oxidations a monooxygenase enzyme from *Pseudomonas putida* NCIMB 10007. J. *Chem. Soc., Chem. Commun*; 699-701.

Grogan G., Graf J., Jones A., Parsons S., Turner N.J. and Flitsch S.L. (2001a) An asymmetric anzyme-catalyzed Retro-Claisen reaction for the desymmetrization of cyclic beta-diketones *Angew Chem Int Ed Engl.* 40(6):1111-1114.

Grogan G., Roberts G.A., Bougioukou D., Turner N.J. and Flitsch S.L. (2001b) The desymmetrization of bicyclic beta-diketones by an enzymatic retro-Claisen reaction. A new reaction of the crotonase superfamily. *J. Biol. Chem.* 276(16):12565-72

Grogan G., Roberts G.A., Parsons S., Turner N.J. and Flitsch S.L. (2002) P450(camr), a cytochromeP450 catalysing the stereospecific 6-*endo*-hydroxylation of (1R)-(+)-camphor. *Appl. Microbiol. Biotechnol.* 59(4-5):449-54

Gunsalus I.C. and Marshall V.P. (1971) Monoterpene dissimilation: chemical and genetic models. *CRC critical reviews in microbiology*, 1:291-310.

Haniu M., Armes L.G., Yasunobu K.T., Shastry B.A. and Gunsalus I.C. (1982) Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen bromide peptides, acid cleavage peptides, and the complete sequence. *J. Biol. Chem.* 257(21):12664-71.

Hansen J.B. and Olsen R.H. (1978) IncP2 group of *Pseudomonas*, a class of uniquely large plasmids. *Nature*, 274: 715-717.

Hanson J.R. (1995) *An introduction to biotransformations in organic chemistry.* W.H. Freeman and company limited, Oxford.

Harrison S.C. and Aggarwal A.K. (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* 59:933-69.

Hedegaard J. and Gunsalus I.C. (1965) Mixed function oxidation. IV. An induced methylene hydroxylase in camphor oxidation. *J. Biol. Chem.* 240(10):4038-43.

Hedges R.W., Jacob A.E. and Crawford I.P. (1977) Wide ranging plasmid bearing the Pseudomonas aeruginosa tryptophan synthase genes. *Nature*. 267(5608):283-4.

Hill C.M., Wu F., Cheng T.C., DeFrank J.J. and Raushel F.M. (2000) Substrate and stereochemical specificity of the organophosphorus acid anhydrolase from Alteromonas sp. JD6.5 toward p-nitrophenyl phosphotriesters. *Bioorg. Med. Chem. Lett.* 10(11):1285-8.

Hong S.B. and Raushel F.M.(1996) Metal-substrate interactions facilitate the catalytic activity of the bacterial phosphotriesterase. *Biochemistry*. 35(33):10904-12.

Hong S.B. and Raushel F.M. (1997) Inhibitors directed towards the binuclear metal center of phosphotriesterase. *J. Enzyme Inhib.* 12(3):191-203.

Horn C., Namane A., Pescher P., Riviere M., Romain F., Puzo G., Barzu O., Marchal G. (1999) Decreased capacity of recombinant 45/47-kDa molecules (Apa) of Mycobacterium tuberculosis to stimulate T lymphocyte responses related to changes in their mannosylation pattern. *J. Biol. Chem.* 274(45):32023-30.

Horne I., Sutherland T.D., Harcourt R.L., Russell R.J. and Oakeshott J.G. (2002a) Identification of an *opd* (organophosphate degradation) gene in an *Agrobacterium* isolate. *Appl. Environ. Microbiol.* 68(7):3371-6.

Horne I., Sutherland T.D., Oakeshott J.G. and Russell R.J. (2002b) Cloning and expression of the phosphotriesterase gene *hocA* from *Pseudomonas monteilii* C11. *Microbiology.* 148(Pt 9):2687-95.

Huang X. A. (1992) Contig assembly program based on sensitive detection of fragment overlaps. *Genomics.*14(1):18-25.

Hugouvieux-Cotte-Pattat N., Kohler T., Rekik M., Harayama S. (1990) Growth-phase-dependent expression of the *Pseudomonas putida* TOL plasmid pWW0 catabolic genes. *J. Bacteriol.* 172(12):6651-60.

Ishikawa J. and Hotta K. (1999) FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol Lett.*174(2):251-3.

Jack Chen Y.-C. Peoples O. and Walsh C. (1988) Acinetobacter Cyclohexanone monooxygenase: gene cloning and sequence determination. *J. Bacteriol.* 170(2):781-789.

Jakoby M., Nolden L., Meier-Wagner J., Kramer R., Burkovski A. (2000) AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum. Mol Microbiol.* 37(4):964-77.

Jeffrey B. Hansen and Ronald H. Olsen (1978) $IncP_2$ group of *Pseudomonas*, a class of uniquely large plasmids. *Nature*. 274: 715-717.

Jones H. K., Smith R.T.and Trudgill P. W. (1993) Diketocamphane enantiomer-specific 'Baeyer-Villiger' monooxygenases from camphor-grown *Pseudomonas putida* ATCC17453. *J. Gen. Microbiology*. 139: 797-805.

Kabsch W and Sander C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22(12):2577-637.

Katagiri Masayuki, Gunguli N. B. And Gunsalus I.C. (1968) A soluble cytochrome P-450 function in methylene hydroxylation. *J. Biol. Chem.* 243 (12); 3543-3546.

Kelly R.D., Knowles J.C., Mahdi G.J., Wright A.M., Taylor N.I., Roberts M.S., Wan WP., Grogan G. Pedrgosa-Moreau S. and Willetts J.A. (1996) Stereochemical congruence of Baeyer-Villigerases. *Chem. Commun.* 2333.

Kelly D.R., Wan Peter W.H. and Tang J. (1998) Flavin monooxygenase-use as catalyst for Baeyer-Villiger ring expansion and heteroatom oxidation. *Biotechnology 8a*: 536-587, Wiley-VCH.

King R.D., Saqi M., Sayle R. And Sternberg M.J. (1997) DSC: public domain protein secondary structure predication. *Comput. App.l Biosci.* 13(4):473-4.

Kobayashi M., Shinohara M., Sakoh C., Kataoka M. and Shimizu S. (1998) Lactone-ring-cleaving enzyme: genetic analysis, novel RNA editing, and evolutionary implications. *Proc. Natl. Acad. Sc.i U S A.* 95(22):12787-92.

Koga H., Rauchfuss B. and Gunsalus I. C. (1985) P450cam gene cloning and expression in *Pseudomonas putida* and *Escherichia coli. Biochem Biophy. Res. Commun.* 130(1): 412-417.

Koga H. Aramaki H., Yamaguchi E., Takeuchi K., Horiuchi T. and Gunsalus I. C. (1986) *camR*, A Negative Regulator Locus of the Cytochrome P-450cam Hydroxylase Operon. *J. Bacteriol.* 166: 1086-1095.

Koga H., Yamaguchi E., Matsunaga K. Aramaki H. and Horiuchi T. (1989) Cloning and neucleotide sequences of NADH- Putidaredoxin reductase gene (camA) and putidaredoxin gene (camB) involved in Cytochrome P450cam hydroxylase of *Pseudomonas putida. J. Biochemistry.* 106: 831-836.

Kuo A., Blough N.V. and Dunlap P.V. (1994) Multiple N-acyl-L-homoserine lactone autoinducers of luminescence in the marine symbiotic bacterium *Vibrio fischeri. J. Bacteriol.* 176(24):7558-65.

Kuo J.M. and Raushel F.M. (1994) Identification of the histidine ligands to the binuclear metal center of phosphotriesterase by site-directed mutagenesis. *Biochemistry.* 33(14):4265-72.

Kurt F. (1992) *Biotransformations in organic chemistry.* Springer-Verlag Berlin Heidelberg, Germany.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227(259):680-5.

Lai K., Stolowich N.J. and Wild J.R. (1995) Characterization of P-S bond hydrolysis in organophosphorothioate pesticides by organophosphorus hydrolase. *Arch. Biochem. Biophys.* 318(1):59-64.

LeGall, Bertland A. U., Namtredt M.J. and Conrad H.E. (1963) Enzyme and intermediates from D- and L-camphor induced *Pseudomonas*. Federation Proceedings 22: 295.

Li S., Wackett L.P. (1993) Reductive dehalogenation by cytochrome P450CAM: substrate binding and catalysis. *Biochemistry*. 32(36):9355-61.

Loida P.J., Sligar S.G., Paulsen M.D., Arnold G.E. and Ornstein RL. (1995) Stereoselective hydroxylation of norcamphor by cytochrome P450cam. Experimental verification of molecular dynamics simulations. *J. Biol. Chem.* 270(10):5326-30.

Lombard V., Camon E.B., Parkinson H.E., Hingamp P., Stoesser G. and Redaschi N. (2002) EMBL-Align: a new public nucleotide and amino acid multiple sequence alignment database. *Bioinformatics*. 18(5):763-4.

Mandel M. (1996) Deoxyribonucleic acid base composition in the genus Pseudomonas. *J. Gen. Microbiol.* 43(2):273-92.

McDaniel C.S., Harper L.L. and Wild J.R. (1988) Cloning and sequencing of a plasmid-borne gene (opd) encoding a phosphotriesterase. *J. Bacteriol.* 170(5):2306-11.

Meighen E.A. and Bartlet I. (1980) Complementation of subunits from different bacterial luciferases. Evidence for the role of the beta subunit in the bioluminescent mechanism. *J Biol Chem.* 255(23):11181-7.

Meighen E.A. (1991) Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* 55(1):123-42.

Mitsuru H. Tanaka M. and Yasunobu K. T. (1982) Amino acid sequence of the *Pseudomonas putida* Cytochrome P-450; I. Sequences of tryptic and clostripain peptides. *J. Biol. Chem.* 257(21): 12657-12663.

Mitsuru H., Tanaka M. and Yasunobu K. T. (1982) Amino acid sequence of the *Pseudomonas putida* Cytochrome P-450; II. Cyanogen bromide peptides, acid cleavage peptides, and the complete sequence. *J. Biol. Chem.* 257(21): 12664-12671.

Morii S., Sawamoto S., Yamauchi Y., Miyamoto M., Iwami M., and Itagaki E. (1999) Steroid monooxygenase of *Rhodococcus rhodochrous*: sequencing of the genomic DNA, and hyperexpression, purification, and characterization of the recombinant enzyme. *J. Biochem.* (Tokyo) 126:624-31.

Mulbry W.W., Karns J.S., Kearney P.C., Nelson J.O., McDaniel C.S. and Wild J.R. (1986) Identification of a plasmid-borne parathion hydrolase gene from Flavobacterium sp. by southern hybridization with opd from *Pseudomonas diminuta*. *Appl. Environ. Microbiol.* 51(5):926-30.

Mulbry W.W., Kearney P.C., Nelson J.O. and Karns J.S. (1987) Physical comparison of parathion hydrolase plasmids from *Pseudomonas diminuta* and *Flavobacterium sp. Plasmid.* 18(2):173-7.

Mulbry W.W. and Karns J.S. (1989a) Parathion hydrolase specified by the *Flavobacterium opd* gene: relationship between the gene and protein. *J. Bacteriol.* 171(12):6740-6.

Mulbry W.W. and Karns J.S. (1989b) Purification and characterization of three parathion hydrolases from gram-negative bacterial strains. *Appl. Environ. Microbiol.* 55(2):289-93.

Mulchandani A., Mulchandani P., Kaneva I. and Chen W. (1998) Biosensor for direct determination of organophosphate nerve agents using recombinant *Escherichia coli* with surface-expressed oganophosphorus hydrolase. 1. Potentiometric microbial electrode. *Anal. Chem.* 70(19):4140-5.

Mulchandani A., Kaneva I. and Chen W. (1998) Biosensor for direct determination of organophosphate nerve agents using recombinant *Escherichia coli* with surface-expressed organophosphorus hydrolase. 2. Fiber-optic microbial biosensor. *Anal. Chem.* 70(23):5042-6.

Nakai K. and Horton P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24(1):34-6.

Nakamura Y., Wada K., Wada Y., Doi H., Kanaya S., Gojobori T. and Ikemura T. (1996) Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Res.* 24(1):214-5.

Newton R.F. And Roberts S.M. (1980) Steric control in prostaglandin synthesis involving bicyclic and tricyclic intermediates. *Tetrahedron.* 36: 2163-2196.

Nielsen H., Engelbrecht J., Brunak S. and von Heijne G. (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* 8(5-6):581-99.

Nielsen H., Brunak S., von Heijne G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 12(1):3-9.

Oikawa E., Shimizu A. and Ishibashi Y. (1995) 2-Methylisoborneol degradation by the CAM operon from *Pseudomonas putida* PpG1. *Wat. Sci. Tech* 31(11): 79-86.

Omburo G.A., Kuo J.M., Mullins L.S. and Raushel FM. (1992) Characterization of the zinc binding site of bacterial phosphotriesterase. *J. Biol. Chem.* 267(19):13278-83.

Omburo G.A., Mullins L.S. and Raushel F.M. (1993) Structural characterization of the divalent cation sites of bacterial phosphotriesterase by 113Cd NMR spectroscopy. *Biochemistry.* 32(35):9148-55.

Orth P., Schnappinger D., Hillen W., Saenger W. and Hinrichs W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol.* 7(3):215-9.

Ougham H.J., Taylor D.G. and Trudgill P.W. (1983) Camphor revisited: involvement of a unique monooxygenase in metabolism of 2-oxo-delta 3-4,5,5-trimethylcyclopentenylacetic acid by *Pseudomonas putida. J. Bacteriol.* 153(1):140-52.

Pabo C.O. and Sauer R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053-95.

Pearson J.P., Passador L., Iglewski B.H. and Greenberg E.P. (1995) A second N-acylhomoserine lactone signal produced by *Pseudomonas aeruginosa. Proc Natl Acad Sci U S A.* 92(5):1490-4.

Peterson J.A., Lorence M.C. and Amarneh B. (1990) Putidaredoxin reductase and putidaredoxin; cloning, sequence determination, and heterologous expression of proteins. *J. Biol. Chem.* 265 (11): 6066-6073.

Petit F. And Furstoss R. (1995) Synthesis of (1S, 5R)-2,8-dioxabicyclo[3.3.0]octan-3one from its enantiomer: A subunit of Clerodane derivatives. *Synthesis* (December) 1517

Platt T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* 55:339-72.

Rani N. Lakshimi and Lalithakumari D. (1994) Degradation of methyl parathion by *Pseudomonas putida. Can. J. Microbiol.* 40; 1000-1006.

Rastogi V.K., DeFrank J.J., Cheng T.C. and Wild J.R. (1997) Enzymatic hydrolysis of Russian-VX by organophosphorus hydrolase. *Biochem. Biophys. Res. Commun.* 241(2):294-6.

Raushel F.M. (2002) Bacterial detoxification of organophosphate nerve agents. *Curr. Opin. Microbiol.* 5(3):288-95

Rheinwald J.G. Chakrabarty A.M. and Gunsalus I. C. (1973) A transmissible plasmid controlling camphor oxidation in *Pseudomonas putida. Proc. Nat. Acad. Sci. U.S.A.* 70(3): 885- 889.

Roberts S.M. and Willettts J.A. (1993) Development of the enzyme-catalysed Baeyer-Villiger Reaction as a useful technique in organic synthesis. *Chirality.* 5:334-337.

Robson N.D., Cox A.R., McGowan S.J., Bycroft B.W. and Salmond G.P. (1997) Bacterial N-acyl-homoserine-lactone-dependent signalling and its potential biotechnological applications. *Trends Biotechnol.* 15(11):458-64.

Rosenberg M. and Court D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13:319-53.

Rost B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* 266:525-39.

Rowland S.S., Speedie M.K. and Pogell B.M. (1991) Purification and characterization of a secreted recombinant phosphotriesterase (parathion hydrolase) from *Streptomyces lividans. Appl. Environ. Microbiol.* 57(2):440-4.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) *Molecular cloning: a laboratory manual 2ed.* Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.

Schumacher M.A., Miller M.C., Grkovic S., Brown M.H., Skurray R.A. and Brennan R.G. (2001) Structural mechanisms of QacR induction and multidrug recognition. *Science*. 294(5549):2158-63.

Schumacher M.A. and Brennan R.G. (2002) Structural mechanisms of multidrug recognition and regulation by bacterial multidrug transcription factors. *Mol. Microbiol.* 45(4):885-93.

Schumacher M.A., Miller M.C., Grkovic S., Brown M.H., Skurray R.A. and Brennan (2002) R.G. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.* 21(5):1210-8.

Selifonov S.A. (1992) Microbial oxidation of adamantanone by *Pseudomonas putida* carrying the camphor catabolic plasmid. *Biochem. Biophys. Re.s Commun.* 186(3):1429-36.

Sethunathan N. and Yoshida T. (1973) A *Flavobacterium* sp. that degrades diazinon and parathion. *Can. .J Microbiol.* 19(7):873-5.

Shimazu M., Mulchandani A. and Chen W. (2001) Simultaneous degradation of organophosphorus pesticides and *p*-nitrophenol by a genetically engineered *Moraxella* sp. with surface-expressed organophosphorus hydrolase. *Biotechnol. Bioeng.* 76(4):318-24.

Sokatch R.J. (1986) *The bacteria, A treatise on structure and function Volume X; The biology of Pseudomonas.* Academic Press, Inc. London.

Sonnhammer L.L. Erik, Heijne von Gunnar and Krogh Anders (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *In proceeding of the 6th Int conf. On Intelligent Systems in Molecular biology*. 1-8. Glasgow, AAAI Press.

Spiers A.J., Buckling A., Rainey P.B. (2000) The causes of *Pseudomonas* diversity. *Microbiology*. 146:2345-50.

Stanier R.Y., Palleroni N.J., Doudoroff M. (1996) The aerobic pseudomonads: a taxonomic study. *J. Gen. Microbiol.* 43(2):159-271.

Steiert G.J., Pogell M.B., Speedie K.M. and Laredo J. (1989) A gene coding for a membrane-bound hydrolase is expressed as a secreted soluble enzyme in *Streptomyces lividans*. *Bio/Technol.* 7:65-68.

Stothard P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques.* 28(6):1102-1104.

Surette M.G., Stock J.B. (1996) Role of alpha-helical coiled-coil interactions in receptor dimerization, signaling, and adaptation during bacterial chemotaxis. *J. Biol. Chem.* 271(30):17966-73.

Summers D. K. (1996) *The biology of plasmids*. Blackwell Science.

Takano E., Chakraburtty R., Nihira T., Yamada Y. And Bibb M.J. (2001)A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 41(5):1015-28.

Tanaka M. Haniu M, Yasunobu K. T. Dus K. and Gunsalus I. C. (1974) The amino acid sequence of putidaredoxin, an iron-sulfur protein from *Pseudomonas putida*. *J. Biol. Chem.* 249(12): 3689-3701.

Tatusova T.A. and Madden T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett.* 174(2):247-50

Taylor D.G. and Trudgill P.W. (1986) Camphor revisited: Studies of 2,5-diketocamphane 1,2-monooxygenase from *Pseudomonas putida* ATCC 17453. *J. Bacteriol.* 165(2): 489-497.

Ullah J.H., Walsh T.R., Taylor I.A., Emery D.C., Verma C.S., Gamblin S.J. and Spencer J. (1998) The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 A resolution. *J. Mol. Biol.* 284(1):125-36.

Unger B.P., Gunsalus I.C. and Silgar S.G. (1986) Nucleotide sequence of the *Pseudomonas putida* Cytochrome P-450cam gene and its expression in *Escherichia coli. J. Biol. Chem.* 261(3): 1158-1163.

Vallon O. (2000) New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure. *Proteins.* 38(1):95-114.

van Beilen J.B., Eggink G., Enequist H., Bos R. and Witholt B. (1992) DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alkJKL* genes of Pseudomonas oleovorans. *Mol Microbiol.* 6(21):3121-36.

van der Meer J.R., de Vos W.M., Harayama S. and Zehnder A.J. (1992) Molecular mechanisms of genetic adaptation to xenobiotic compounds. *Microbiol. Rev.* 56(4):677-94.

van der Werf M.J., Swarts H.J. and de Bont J.A. (1999) *Rhodococcus erythropolis* DCL14 contains a novel degradation pathway for limonene. *Appl. Environ. Microbiol.* 65(5):2092-102.

Vanhooke J.L., Benning M.M., Raushel F.M. and Holden H.M. (1996) Three-dimensional structure of the zinc-containing phosphotriesterase with the bound substrate analog diethyl- 4-methylbenzylphosphonate. *Biochemistry*. 35(19):6020-5.

von Heijne G. (1988) Transcending the impenetrable: how proteins come to terms with membranes. *Biochim. Biophys. Acta*. 947(2):307-33.

Wang B.S., Pabo C.O. (1999) Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *Proc. Natl. Acad. Sci. U. S. A.* 96(17):9568-73.

Waddle J. and Baldwin T.O. (1991) Individual alpha and beta subunits of bacterial luciferase exhibit bioluminescence activity. *Biochem. Biophys. Res. Commun.* 178(3):1188-93.

Willetts J.A. (1997) Structural studies and synthetic applications of Baeyer-Villiger monooxygenases. *Trends Biotech.*15: 55-62.

Wilmot C.M. and Thornton J.M. (1988) Analysis and prediction of the different types of beta-turn in proteins. *J. Mo.l Biol.* 203(1):221-32.

Yu C.A. And Gunsalas I.C. (1969) Monooxygenase; VII Camphor ketolactonase I and the role of three protein components. *J. Biol. Chem.* 24 (22): 6149-6152.

Zhongli C., Shunpeng L. and Guoping F. (2001) Isolation of methyl parathion-degrading strain M6 and cloning of the methyl parathion hydrolase gene. *Appl. Environ. Microbiol.* 67(10):4922-5.