Cloning, Expression and Characterization of Novel Lipase and Esterases from *Burkholderia multivorans* UWC10

Konanani J. Rashamuse

A thesis submitted in fulfilment of the requirements for the degree of Doctor Scientiae in the Department of Biotechnology University of the Western Cape.

November 2005

Supervisor: Prof. Don A. Cowan

Abstract

An esterase and lipase producing *Burkholderia multivorans* strain was isolated by culture enrichment strategies. A shotgun library of *Burkholderia multivorans* genomic DNA (prepared in *E. coli*/pUC18) was screened for lipase and esterase activities. Three positive recombinant clones, pTEND5, pHOLA6 and pRASH14, conferring esterolytic and lipolytic phenotypes respectively, were identified. Full-length sequencing of DNA inserts was performed using subcloning and "primer-walking" strategies.

Nucleotide sequence analysis revealed that the pRASH14 plasmid DNA consisted of two open reading frames (ORF1 and ORF2) encoding 356 and 350 amino acids, respectively. Database searches revealed that ORF1 and ORF2 were homologous to lipases and chaperones from subfamily I.2. In the pTEND5 sequence, an open reading frame consisting of 978 bp, encoding 326 amino acids, was identified. Database searches revealed that this open reading frame was homologous to family V esterases. Nucleotide sequence analysis revealed that pHOLA6, plasmid DNA consisted of 1194 bp encoding 398 amino acids and showed homology to family VIII esterases. The primary structures of LipA, EstEFH5 and EstBL from pRASH14, pTEND5 and pHOLA6, respectively, showed a classical GxSxG motif, which is conserved in many serine hydrolases. In addition, EstBL also showed a consensus SxxK motif, the serine of which acts as a catalytic nucleophile in class C β-lactames and some peptidases.

The gene encoding EstBL was expressed in *Escherichia coli* and purified to homogeneity by a combination of ammonium sulphate fractionation and hydrophobic interaction, ion exchange and size exclusion chromatographies. Thermostability and pH stability profiles showed that EstBL was stable between 20 -30 °C and pH 7.0 to 8.5, respectively. EstBL showed a preference against ρ-nitrophenyl- and β-naphthyl- esters of shorter chain length (C2-C4), while activity towards β-lactam substrates was not detectable. At 1 mM, PMSF Ag^{2+}, Cu^{2+} and Ag^{2+} inhibited EstBL activity.

EstEFH was recovered from the insoluble cellular fraction by *in vitro* refolding and partially purified using His select™ Nickel column chromatography. Substrate specificity studies showed that EstEFH5 preferred shorter acyl chain length (C2-C3) substrates. Lipase expression studies showed that the chaperone gene located downstream the lipase structural gene was required for heterologous expression.

Homology modeling studies of LipA and EstBL, predicted that these proteins adopt α/β-hydrolase fold topologies. The EstBL model structure suggested that the serine residue with SxxK motif is likely to act as a nucleophile.

Key word: *Burkholderia multivorans*, Esterases, Lipases, αβ-hydrolase fold. Gene cloning, Protein expression, Protein purification.

Declaration

I KONANANI J. RASHAMUSE declare that is my own work that has not been submitted before for any degree or examination in any other university, and that all the sources I have used for quoted have been indicated and acknowledged as complete references.

Konanani J. Rashamuse **November 2005**

Signature:.....

Acknowledgements

The author would like to express his appreciation to the following people for contribution:

- Prof. Don Cowan for his professional supervision, guidance, sound judgement, encouragement, enthusiasm and creating such a conducive environment to carry out research. Ndo livhuwa!
- Drs William Stafford and Lukas Rohr for numerous hints and advices.
- Dr Heide Goodman, Nazneen Ibrahim and Lisa Morse for their technical support
- Prof. Stephanie Burton, Department of Chemical Engineering UCT for the use of HPLC.
- Prof. Helmut Schwab, Institute of Biotechnology, Technical University of Graz for kindly providing the pMS470Δ8 expression vector.
- Dr Robin Mitra, CSIR for kindly providing ethyl ferulate.
- ARCAM family for good times, bad times and all times
- Hola7 to the boys Tshepo Tsekoa and Jozi Lako for friendship and support for the past three years
- Inter Magadidzha, Hulisani Ratshinanga and Simon Thanayani for the support throughout my postgraduate studies.
- Langanani Ramusetheli thanks for everything. I would not have made it this far without your unconditional support. You are amazing!
- Tshifhiwa Masikhwa for the support, encouragement and all the good times
- To my parents and my siblings for unconditional faith and moral support throughout my studies.
- A debt of gratitude is owed to CSIR, DAAD and DST for much need financial support.
- To the higher power for blessings

Table of contents

Chapter Two: Materials and Methods

Chapter Four: Cloning and sequencing of lipase and esterase encoding genes from *Burkholderia multivorans* UWC10

Chapter Five: Heterologous expression purification and characterization

List of Figures

List of Tables

List of Abbreviations

Å	angstrom
bp	base pair
Da	Dalton
dNTPs	Deoxynucleotide tri-phosphates
DTT	dithiothretol
EDTA	Ethylenediminetetraacetic acid
EF	Ethyl ferulate
FA	Ferulic acid
g	gram
h	hour
IPTG	Isopropyl-β-D-thiogalactoside
kb	Kilobase
K_{cat}	catalytic turnover
kDa	Kilodalton
K_M	Michaelis Menten constant
L	litre
L	litre
Lb	Luria-Bertani
M	molar
m	milli (10^{-3})
min	minute
MOPS	3-(*N*-morpholino) propanesulfonic acid
O.D	optical density
°C	degree Celsius
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PIPES	piperazine-1,2 bis [2-ethanesulfonic acid]
SDS	sodium dodcyl sulphate
U	Unit
v/v	volume per volume
Vmax	rate of enzyme-catalysed reaction at infinite substrate concentration
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
μ	micro (10^{-6})

CHAPTER ONE

Literature Review

1.1 Classification of microbial lipolytic enzymes

Microbial carboxylesterases and lipases have traditionally been classified on the basis of substrate specificities. This classification scheme presented a serious drawback because it required that all enzymes should be compared using the same or at least related substrates, and preferably under same conditions (Whitaker, 1972). However, as esterases and lipases display broad substrate specificities, it has become clear that classification of these enzymes cannot be based solely on their functional properties. The growing number of primary structures has provided a clearer picture of the evolutionary relationships between lipolytic enzymes of different origin. Currently, microbial lipolytic enzymes (including both esterases and lipases) are classified into eight families (Arpigny and Jaeger, 1999). This classification scheme is based on a comparison of amino acid sequences and some fundamental biological properties of these enzymes.

Family I consists of the so-called "true lipases". This family is further subdivided into six subfamilies. Subfamilies I.1 and I.2 are composed of lipases with N-terminal signal or leader peptides and are secreted by two Xcp-dependent pathways that typically require the participation of lipase-specific foldase or chaperone proteins. The chaperone gene, when present, is located adjacent to the lipase structural gene. Typical examples of this group include lipases from *Pseudomonas aeruginosa* (*Ps. aeruginosa*) D50587 (Chihara-Siomi *et al.*, 1992), *Burkholderia glumae* (*B. glumae*) X70354 (Frenken *et al*, 1992), *B.*

cepacia (formerly *Ps. cepacia*) M58494 (Jorgensen *et al.*, 1991), *Acinetobacter calcoaceticus* (*Ac. calcoaceticus*) BD413 (Kok *et al.*, 1995), *Ac. calcaoceticus* RAG-1 (Sullivan *et al.*, 1999), *Proteus vulgaris* (*P. vulgaris*) (Kim *et al.*, 1996) and *Vibro cholerae (V. cholerae*) (Ogierman *et al.*, 1997). Furthermore, enzymes of these subfamilies also share some prominent structural features such as a Ser, His, Glu/Asp catalytic triad. The catalytic serine is normally located within a G-x-S-x-G consensus sequence motif (where x-represents an arbitrary amino acid residue). A single disulphide bridge and aspartic residues involved in a calcium binding site are also typical of subfamilies I.1 and I.2. Subfamily I.3 includes *Ps. fluorescens* lipase D11455 (Duong, *et al.*, 1994) and *Serratia marcescens* lipase D13253 (Li *et al.*, 1995) which have a much higher molecular weight than lipases of subfamilies I.1 and I.2. Subfamily I.3 lipases lack both chaperone and the N-terminal signal peptide, and are secreted via a one ATP-binding cassette transporter system (Jaeger *et al.*, 1999; Omori and Idei, 2003). For the purpose of clarity, subfamilies I.1 to I.3 lipases will collectively be referred to throughout the text as Group I proteobacterial lipases.

Subfamilies I.4-I.6 consist of lipases from gram-positive organisms. SUBFAMILIES I.4 includes lipases from *Bacillus pumilus* (*B. pumilus*) A34992 and *Bacillus subtilis* (*B. subtilis*) M74010 (Dartois *et al.*, 1992) and are some of the smallest lipases known, with molecular masses of approximately 19 kDa. Subfamily I.5 includes enzymes from *Bacillus* species (such as *Geobacillus stearothermophilus* lipase U7878 and *Geobacillus thermocatenulatus* X95309) (Schmidt-Dannert *et al.*, 1997), with molecular masses of about 45 kDa. Subfamily I.6 comprises the Staphylococcal lipases. These high molecular

weight enzymes are secreted as precursors and are cleaved in the extracellular medium by a specific protease, yielding a mature protein of approximately 400 residues. *Staphylococcus hyicus* (X02824) lipase displays high phospholipase activity, an activity not observed with any other true lipase (Ayora *et al.*, 1994; Van Kampen *et al.*, 2001).

The enzymes of family II lack the consensus pentapeptide sequence motif (G-x-S-x-G), and contain a GDSL consensus sequence motif. *Streptomyces scabies* (*S. scabies*) esterase (M57297) is the best-studied member of this family (McQueen and Schottel, 1987), and contains a catalytic Ser-His dyad (Wei *et al.*, 1995). Family III enzymes are best represented by lipases from *S. exfolietiatus*. They show 20% amino acid sequence identity with intracellular and plasma isoforms of the human PAF-AH proteins. In contrast to family II, these enzymes contain a classical catalytic triad with an α/β hydrolase fold.

Family IV is also referred to as the HSL family, as enzymes belonging to this family show remarkable similarity to mammalian hormone-sensitive lipases (HSLs). HSLs catalyze the hydrolysis of triacylglycerols in adipose tissue and are rate limiting enzymes in the mobilization of fatty acids from stored lipids. Human HSL consists of an N-terminal domain of unknown function and a catalytic domain into which a regulatory module is inserted (Hemila *et al.*, 1994). Bacterial enzymes belonging to this family display similarity to the catalytic domain of HSL, implying that mammalian HSLs might have evolved from these bacterial Family IV enzymes. The members of this family contain a consensus GDSAG pentapeptide, corresponding to the GxSxG motif. Enzymes

in this family also contain a highly conserved HGGG motif, the function of which is not currently known.

Family V includes esterases from *Sulfolobus acidocaldarius* (AF071233), *Acetobacter pasteurianus* (AB013096) and *Ps. fluorescens* (Kim *et al.*, 1997b). Moreover, esterases in this family share significant sequence homology to non-lipolytic enzymes, such as epoxide hydrolase, dehalogenase, halo-peroxidase and calcium independent lysophospholipase A_2. They are generally small, with molecular masses ranging from 23 to 26 kDa.

Families VI and VII contain large bacterial esterases with molecular masses of around 55 kDa. They share significant amino acid sequence homology (30% identity, 40% similarity) with eukaryotic acetylcholine esterases and intestine/liver carboxylesterases (e.g., pig liver esterase). This family also includes one of the best-studied esterase, a ρ-nitrobenzyl esterase from *B. s subtilis* (Zock *et al.*, 1994).

Family VIII esterases display high homology to class C β-lactamases. These enzymes contain G-x-S-x-G and S-x-x-K motifs. Site directed mutagenesis studies of BDA from *Brevibacterium linens* (*Brevi. linens*) (Sakai *et al.*, 1999) and EstB from *B. gladioli* (Petersen *et al.*, 2001) have demonstrated that the serine of the G-x-S-x-G motif does not play a catalytic role: in contrast, the serine of the S-x-x-K motif was shown to be vital for catalysis. Known members of this family include *Arthrobacter globiformis* esterase (AAA99492) (Nishizawa *et al.*, 1995), BDA from *Brevi. linens* (BAA78097) (Sakai *et*

al., 1999), EstB from *B. gladioli* (AAF59826) (Petersen *et al.*, 2001) and Lip8 from *Ps. fluorescens* (BAD69792) (Ogino *et al.*, 2004).

1.2 Microbial Carboxylesterases: Definition, distribution and biochemical properties

Carboxylesterases (carboxylester hydrolases, EC 3.1.1.1) are defined as enzymes that catalyze the hydrolysis of ester bonds of carboxyl ester substrate molecules to form an alcohol and carboxylic acid (Gandolfi *et al.*, 2000; Bornscheuer, 2002). Carboxylesterases can also be classified according to substrate preferences. The "true" carboxylesterases are those that display maximum activities against substrates with acyl chain lengths of less than 10 carbon atoms and which are generally water soluble (Jaeger *et al.*, 1999).

Carboxylesterases are widely distributed throughout all three domains of life (Okuda, 1991). Microbial carboxylesterases have been isolated from diverse ecosystems including cold adapted environments (Suzuki *et al.*, 2002; Wei *et al.*, 2003; Kulakova *et al.*, 2004), and high temperature environments (Ikeda and Clark, 1998; Huddleston *et al.*, 1995; Manco *et al.*, 1998; Manco *et al.*, 2000; Morana *et al.*, 2002; Hotta *et al.*, 2002; Gao *et al.*, 2003). Carboxylesterases from mesophilic microbial sources represent a large proportion of those reported and have been recovered from genera such as *Pseudomonas* (Shimada *et al.*, 1993; Khalameyzer *et al.*, 1999; Kim *et al.*, 2003; Chao *et al.*, 2003), *Rhodococcus* (Rathbone *et al.*, 1997; Gudelj *et al.*, 1998), *Burkholderia* (Schlacher *et al.*,1998; Reiter *et al.*, 2000; Petersen *et al.*, 2001; Kim *et al.*, 2004a), *Bacillus* (Wood *et*

al., 1995; Niazi *et al.*, 2001; Prim *et al.*, 2001), *Streptomyces* (Von den Haar *et al.*, 1997; Gandolfi *et al.*, 2000) and *Alcaligenes* (Gledhill and Kell, 1998).

A considerable number of microbial carboxylesterases have been cloned and expressed in heterologous hosts (Table 1.1). Amongst the cloned carboxylesterases, the majority are monomeric, although dimeric and tetrameric have also been reported (Table 1.1). The subunit molecular weights of the known carboxylesterases range from 28 to 54 kDa. The isoelectric points (pI) of these enzymes are generally between 4 and 7, with an exception being Esf1 from *Ps. fluorescenes*, which has a pI of 10 (Khalameyzer *et al.*, 1999).

The physiological functions of many carboxylesterases are still unclear, however, a number of physiological roles have been proposed. For example the acetyl, cinnamoyl and feruloyl esterases that degrade hemicellulose are believed to be involved in metabolic pathways that provide access to carbon (Dalrymple *et al.*, 1996). Carboxylesterases are also known to be pathogenic factors. Some bacterial and fungal plant-pathogenic strains are known to possess cell-wall degrading esterases, which are believed to be important in invading plant tissues (McQueen and Schottel, 1987; De Vries, 1997). Some carboxylesterases have also been reported as detoxifying agents, including a *Bacillus* esterase, which inactivates the macrolide phytotoxin brefeldin, A (Wei *et al.*, 1996) and an esterase from *S. lividans* that inactivates the antibiotic fusidic acid (Von den Haar *et al.*, 1997).

Table 1.1: Biochemical properties of recombinant microbial carboxylesterases

Origin	T_{opt} (°C)	pH_{opt} (pI)	Native(subunit) MW(kDa)	Substrate specificity (Best substrate)	References
Acinetobacter No.6 (AelH) [e]	30	9.0 (nr)	28 (28)	ρNP (ρNP-acetate)	Suzuki *et al.*, 2002
Acinetobacter No.6 (EstA8) [e]	50	7.8 (nr)	54 (52)	ρNP (ρNP-caprylate)	Wei *et al.*, 2003
Aeropyrum pernix K1 [e]	90	8.0 (nr)	63 (63)	ρNP (ρNP-caprylate)	Gao *et al.*, 2003
Alicyclobacillus acidocaldarius [e]	70	nr	34	ρNP (ρNP-hexanoate)	Manco *et al.*, 1998
Aspergillus awamori (ACEA) [P]	40	6.0	31 (31)	α-naphthyl esters (α-naphthyl acetate)	Koseki *et al.*,1997
Aspergillus niger (FAEA) [P]	60	5.0 (3.3)	(36)	Methyl ferulate	Jude *et al.*, 2001
Bacillus sp. BP-7 [e]	45	7.5 (5.1)	(53.4)	ρNP (ρNP-butyrate)	Prim *et al.*, 2001
Bacillus subtilis [e]	55	8.5	(32)	Naproxen	Quax and Broekhuizen, 1994
Burkholderia gladioli (EstB) [e]	30	7.0 (4.5)	42 (80)	Cephalosporin (Cephalosporin C)	Petersen *et al.*, 2001
Geobacillus stearothermophilus (Est55) [e]	60	9.0 (nr)	(34)	ρNP (ρNP-butyrate)	Ewis *et al.*, 2004
Geobacillus stearothermophilus (Est30) [e]	70	9.0 (nr)	(30)	ρNP (ρNP-caprylate)	Ewis *et al.*, 2004
Lactococcus lactis (EstA) [L]	nr	nr	(29.6)	ρNP (ρNP-butanoate)	Fernandez *et al.*, 2000
Orpinimyces sp. PC-2 [e]	(30)	7.0 (nr)	(35)	Xylose tetra acetate	Blum *et al.*, 1999
Pseudomonas aeruginosa PAO1 [e]	55	9.0 (5.2)	19 (28)	α-naphyl ester (α-naphyl acetate)	Pesaresi *et al.*, 2005
Pseudomonas aeruginosa LST03 (LIP8) [e]	30	7.0 (nr)	(41)	(Triacetin and methyl acetate)	Ogino *et al.*, 2004
Pseudomonas fluorescens (EstF1) [e]	43	7.5 (10.9)	35	Lactones (δ-valerolactone)	Khalameyzer *et al.*, 1999
Pseudomonas putida (EST) [e]		7.0 (nr)	61 (31)	(R,S)-acetylthioisobutyric	Itoh *et al.*, 2001
Pseudomonas citronellolis (EstA) [e]	40	6.5 (9.5)	(40)	ρNP (ρNP-caproate)	Chao *et al.*, 2003
Pseudomonas sp. KCTC10122BP [e]	30	8.5 (nr)	41 (41)	ρNP (ρNP-butyrate)	Kim *et al.*, 2003
Psychrobacter sp. ANT300 (PsyEst [e])	37	nr	(43)	ρNP (ρNP-hexanoate)	Kulakova *et al.*, 2004
Pyrobaculum calidifontis [e]	90	7.0 (nr)	98 (34)	ρNP (ρNP-caproate)	Hotta *et al.*, 2002
Pyrococcus furiosus [e]	100	nr	nr	4-Methylumbelliferyl-acetate	Ikeda and Clark, 1998
Rodococcus sp. H1(Her) [e]	nr	8.5 (nr)	137 (38)	6-acetylmorphine	Rathbone *et al.*, 1997
Sulfolobus solfataricus (EstA) [e]	80	7.0	114 (32)	ρNP (ρNP-valerate)	Morana *et al.*, 2002

ρNP =para-nitrophenyl esters; nr = not reported; [e] = expressed in *E. coli*; [L]=*Lactococcus lactis*; [P]=*Pichia pastoris*

1.3 Catalytic Mechanism

All known carboxylesterases are serine hydrolases (Bornscheuer, 2002) and their hydrolytic mechanisms resemble those of lipases and serine proteases (Menge, 1990). These enzymes contain a catalytic triad that usually consists of a catalytic serine embedded within a semi-conserved pentapeptide (G-x-S-x-G) motif in the active site (x- denotes any amino acid). In some cases, however, the catalytic serine is located either within the GDSL (McQueen and Schottel, 1987) or the SxxK motifs (Petersen *et al.*, 2001). The ester hydrolysis is mediated by a nucleophilic attack of the active serine on a carbonyl group of the substrate, in a charge-relay system with the two other residues, one basic and one acidic (Dodson and Wlodawer, 1998).

The reaction mechanism of ester hydrolysis by carboxylesterases involves four steps (Fig. 1.1). In the first step (step 1), the hydroxyl oxygen of the active serine attacks the carbonyl carbon atom of the ester bond; a reaction that yields a tetrahedral intermediate stabilized by the catalytic His and Asp residues. The Asp residue forms a hydrogen bond with the adjacent imidazole ring of the His residue. The imadazole ring of the His residue becomes protonated and positively charged. This interaction orientates the imidazole group in the triad and increases its pK_a, favouring proton transfer (Dodson and Wlodawer, 1998).

Once the tetrahedral intermediate is formed (step 2), it is then stabilized by two hydrogen bonds which are formed with the amide bonds of residues sited in the oxyanion hole. This hydrogen bonding is believed to be important, in allowing a closer approach than the van

der Waals distance which favours proton transfer. The alcohol moiety is then released, followed by the formation of the acyl-enzyme complex. This is followed by an attack at the carbon atom of acyl-enzyme complex by the hydroxyl group of the water molecule which yields a second tetrahedral intermediate (step 3). Finally, the acyl component is released and the active enzyme regenerated (step 4).

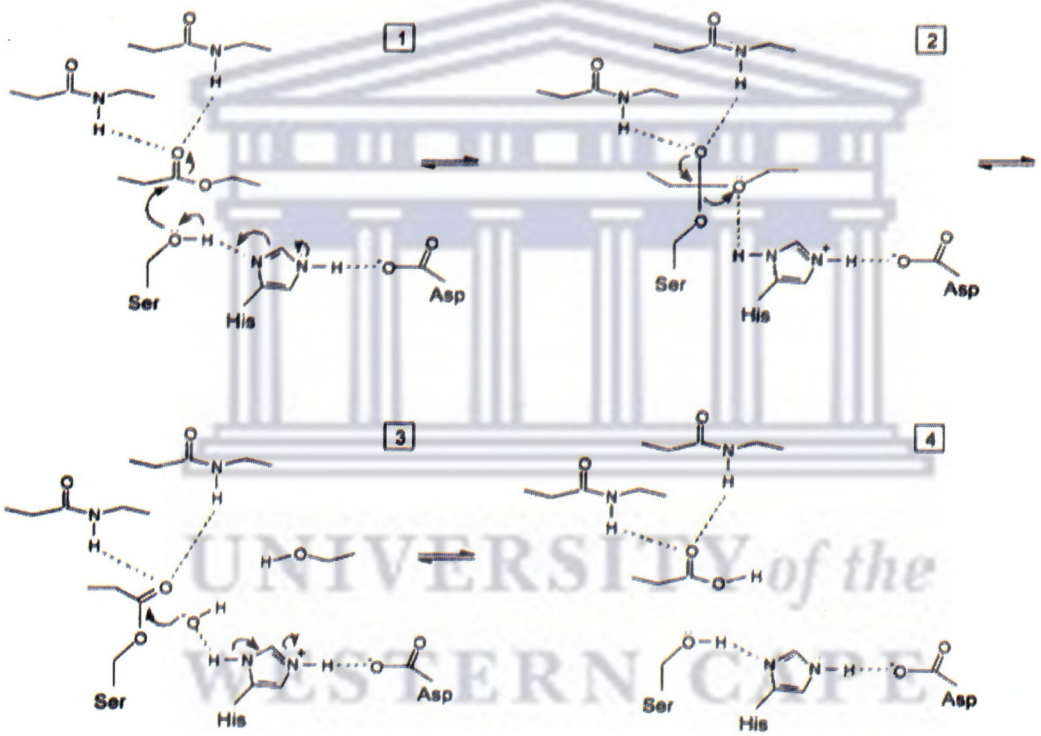


Figure 1.1: Nucleophilic attack on the ester bond by carboxylesterase catalytic triad (Taken from Jeager *et al.* 1999).

1.4 Three-dimensional structures of microbial carboxylesterases

A number of microbial carboxylesterase crystal structures have been reported. These include EsEST from *S. scabies* (Wei *et al.*, 1995), an esterase from *Ps. fluorescens* (Kim *et al.*, 1997b), an esterase from *Alcaligenes* sp. (Bourne *et al.*, 2000), AFEST from *Archaeoglobus fulgidus* (De Simone *et al.*, 2001), EstB from *B. gladioli* (Wagner *et al.*, 2002), EstA from *Aspergillus niger* (Bourne *et al.*, 2004), EST2 from *Alicyclobacillus acidocaldarius* (De Simone *et al.*, 2004) and PdaA from *B. subtilis* (Blair *et al.*, 2004).

The three-dimensional structures revealed that esterases share some common features. All reported esterases adopt a similar core topology, known as the α/β hydrolase fold (Ollis *et al.*, 1992; Nardini and Dijkstra, 1999). This topology, which shows a definite order of α-helices and β-sheets (Fig. 1.2), is shared by other proteins of diverse origin and substrate specificity. The number of protein recognized as members of this superfamily now exceeds 5200, making it one of the most common core structure found in nature (Hotelier *et al.*, 2004). The α/β hydrolase fold provides a scaffold for the catalytic triad (Ser, Asp and His) and the consensus G-x-S-x-G flanking region (Dodson and Wlodawer, 1998). Three dimensional structures of family VI esterases including a brefeldin esterase from *B. subtilis* (Wei *et al.*, 1999) and those from themophilic hosts *Alicyclobacillus acidocaldarius* (De Simone *et al.*, 2000) and *Archaeoglobus fulgidus* (De Simone *et al.*, 2001) revealed that these enzymes harbour multiple α-helix at their N-terminals that comprise cap structures located above their active sites.

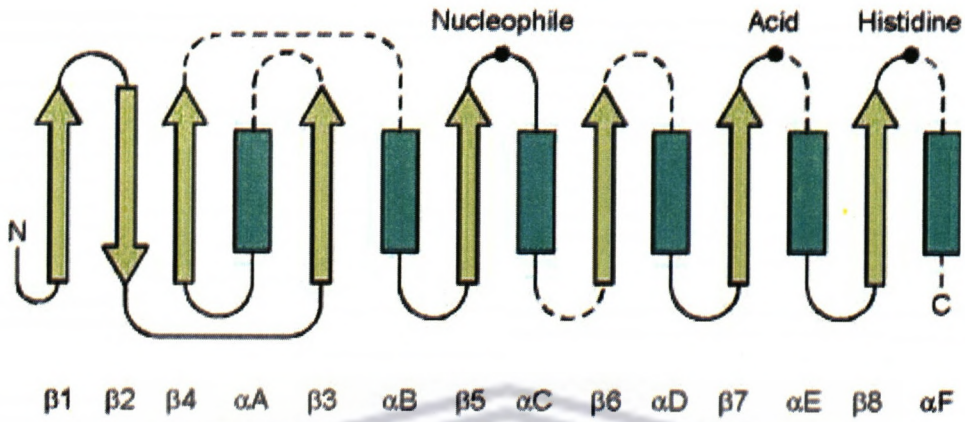


Figure 1.2: Schematic representation of the canonical α/β hydrolase fold (Taken from Nardini *et al.*, 2000).

1.5 Application of microbial carboxylesterases

Microbial carboxylesterases are a diverse group of enzymes with high potential for industrial applications. There are number of features that make esterases potential industrial biocatalysts, including their ability to catalyze a wide range of substrates and high activity and stability in organic media (Jones, 1990; Drauz and Waldmann, 1995). Furthermore, esterases do not require cofactors for their catalytic activity (Bornscheuer, 2002), eliminating the need to recycle co-factors. The use of esterases as industrial biocatalysts has also been reinforced by their ability to carry out chemo-, regio- and stereo-selective reactions (Lagarde *et al.*, 2002).

In addition of catalyzing the hydrolysis of short chain fatty acid esters, carboxylesterases catalyze a wide range of technologically relevant substrates. For instance, an esterase from *B. subtilis* has been used to stereo-specifically resolve (*R, S*) - naproxen to *S*-naproxen, an important non-steroidal anti-flammatory drug (Quax and Broekhuizen, 1994). The reaction yielded *(S)*-naproxen with a 99% enantiomeric excess (ee) at an overall yield of 95 % (Quax and Broekhuizen, 1994). *Arthrobacter globiformis* esterase displayed efficient kinetic resolution in the synthesis of (+)-*trans* (1R, 3R) chrysanthemic acid, an important precursor of pyrethrin insecticides (Nishizawa *et al.*, 1995). This esterase catalyzed the formation of the desired enantiomer with 99% ee at 38% conversion.

Esterases that catalyze the conversion and resolution of tertiary alcohols have also been reported. Both *B. gladioli* EP6 esterase (Schlacher *et al.*, 1998) and *Rhodococcus* RR1 esterase (Gudelj *et al.*, 1998) have been shown to hydrolyze the tertiary alcohol (*R*)-linalyl acetate to (*S*)-(+)-linalool and (*R*)-(-)-linalyl acetate. Tertiary alcohols are generally poor substrates for microbial esterases, probably because of the bulky nature of these compounds causing steric hindrance in active site binding. Analyses of the primary structures of esterases and lipases that show activity against tertiary alcohols, such as *Candida antartica* lipase A and *B. subtilis* esterase, has identified the presence of a unique GGGX motif (Henke *et al.*, 2002). This motif was found located in the oxyanion hole binding pocket of these enzymes and is believed to determine the activity towards tertiary alcohols.

A *Bacillus* sp. esterase has been reported to catalyze the hydrolysis of dimethylphthalate esters, which are industrial important chemicals in the manufacturing of plastics, insect repellents, synthetic fibers and cosmetics (Niazi *et al.*, 2001). A *Rhodococcus* strain H1 esterase capable of hydrolyzing heroin (diacetylmorphine) has been reported (Rathbone *et al.*, 1997). This esterase deacetylates both the C-3 and C-4 groups of heroin, forming morphine. It has been suggested (Holt *et al.*, 1996) that the esterase catalysing heroin metabolism in bacteria could be used as a biosensor in conjunction with bacterial luciferase for the detection of nano-gram quantities of heroin. Two cocaine hydrolyzing esterases, MB11L esterase from *Ps. maltophilia* (Britt *et al.*, 1992) and CocE esterase from *Rhodococcus* sp. strain MB1 (Bresler *et al.*, 2000) have also been reported.

An application of considerable interest is the release of ferulic acid from plant cell wall polysaccharides by ferulic acid (feruloyl) esterase. An *Aspergillus niger* ferulic acid esterase (Fae III) has been shown to release ferulic acid from wheat bran (Faulds and Williamson, 1995). Ferulic acid can be enzymatically converted to vanillin, which is an important flavour compound (Falconnier *et al.*, 1994; Lesage-Messesen, 1996; Gasson, 1998).

1.6 Microbial lipases: definition and properties

Lipases (triacylglycerol acylhydrolases, EC 3.1.1.3) are enzymes that catalyze hydrolysis of triacylglycerides with subsequent release of diacylglycerols, monoacylglycerols, free fatty acids and glycerol (Fig. 1.3) (Macrae, 1989). Lipases can also be classified according to substrate preferences. Lipases have been defined as carboxylesterases that

display maximum activities against substrates with acyl chain lengths of greater than ten carbon atoms that are generally water-insoluble (Jaeger *et al.* 1999). In contrast to esterases, the physiological roles of many lipases are well understood and generally involve of key roles in the biological turnover of lipids (Godtfredsen, 1990). Lipases are also required as digestive enzymes to facilitate the transfer of lipid from one organism to another and the deposition and mobilization of fats that are used as an energy reservoir within organisms (Borgstrom and Brockman, 1984). These enzymes are also involved in the metabolism of intracellular lipids and therefore in the structure and biogenesis of biological membranes (Duong *et al.*, 1997; Villeneuve *et al.*, 2000).

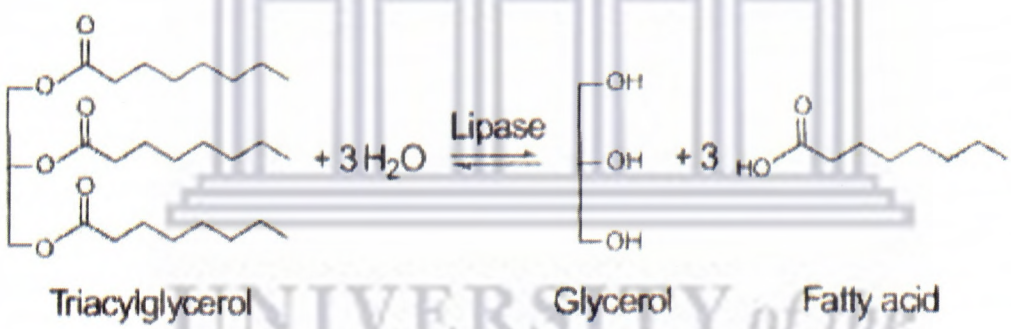


Figure 1.3: Hydrolysis of triacylglceride to glycerol and fatty acids by lipase (Taken from Jaeger and Reetz, 1998).

Since the first report on isolation of lipase producing microorganisms a century ago (Eijkmann, 1901), a large number of lipases have been reported from variety of microbial sources (Table 1.2). Microbial lipases have been isolated from diverse habitats including industrial wastes, vegetable oil processing factories, dairies (Jensen, 1983), oil contaminated with soils (Sztajer *et al.*, 1988), compost heaps (Wang *et al* 1995), coal tips

Table 1.2: Biochemical properties of native and recombinant bacterial lipases

Origin	T_{opt} (°C)	pH_{opt}(pI)	Native(subunit) MW(kDa)	Substrate specificity (Best substrate)	References
Acinetobacter sp. RAG-1	55	9.0	33	ρNP (ρNP-caprylate)	Snellman *et al.*, 2002
Bacillus megaterium[R]	45	7.0	22	ρNP (ρNP-butyrate)	Ruiz *et al.*, 2002
Bacillus sp. strain 398	65	7.0(5.1)	50	ρNP (ρNP-caproate)	Kim *et al.*, 1994
Bacillus subtilis 168	35	9.0	19	Fatty acids (C8)	Lesuisse *et al.*, 1993
Burkholderia sp.	75	7.5	30	Triacylglycerides (C16)	Rathi *et al.*, 2000
Geobacillus[R]	40	7.0	34	nr	Li and Zhang, 2005
Penicillium cyclopium	40	7.0	37	nr	Chahinian *et al.*, 2000
Proteus vulgaris	nr	10(7.3)	31	ρNP	Kim *et al.*, 1996
Pseudomonas aeruginosa EF2	50	9.0(4.9)	29	Fatty acids (C18)	Gilbert *et al.*, 1991
Pseudomonas fluorescens MC50	40	8.0	55	Tricylglycerols	Brune and Gotz, 1992
Pseudomonas fragi 22.39B	65	9.0(6.9)	33	Triacylglycerols	Brune and Gotz, 1992
Pseudomonas mendocina 3121-1	50	7.5	62	ρNP (ρNP-butyrate)	Surinenaite *et al.*, 2002
Pseudomonas sp. KWI-56	60	7.0(5.0)	33	Triacylglycerides (C10-C14)	Brune and Gotz, 1992
Pseudomonas sp. strain KB7004[R]	35	8.0	50	ρNP (ρNP-caprate))	Rashid *et al.*, 2001
Pseudomonas sp. strain B11	45	8.0	34	Pnp (ρNP-butyrate)	Choo *et al.*, 1998
Serratia marcescens	37	8.0	52	nr	Abdou, 2003
Staphylococcus epidermidis[R]	nr	7.0	77	Triglycerols	Simons *et al.*, 1998
Staphylococcus haemolyticus[R]	28	8.5(9.7)	45	ρNP (ρNP-butyrate)	Oh *et al.*, 1999
Staphylococus hyicus[L]	nr	nr	86	nr	Drought *et al.*, 2000
Streptococcus sp. N1[R]	37	8.4	32	Tributyrin	Tripathi *et al.*, 2004
Vibrio vulnificus[R]	nr	nr	31	ρNP (ρNP-myristate)	Su *et al.*, 2004

ρNP =para-nitrophenyl esters; nr = not reported; [(R)] = recombinant enzyme expressed in *E. coli*; [(L)] = recombinant enzyme *Lactococcus lactis*;

(Sharm *et al.*, 2001), cold adapted environments (Fell *et al.*, 1991a and b; Tan *et al.*, 1996; Choo *et al.*, 1998), hot springs (Dominquez *et al.*, 2005; Li and Zhang, 2005) and soda lakes (Vargas *et al.*, 2004).

1.7 Research focuses on microbial lipases

Interest in lipases research over the past decades has focused primarily on three aspects. The first aspect is related to the lipase paradigm (i.e., understanding the molecular basis of lipase catalytic function). This particular aspect of lipase research involves structural characterization, elucidation of lipase mechanisms of action, cloning and sequencing of lipase genes and biochemical characterization of these enzymes (Kazlauskas and Bornscheuer, 1998).

The second aspect has focused on the virulence properties of these enzymes. There is evidence indicating the role of extracellular lipases as virulent factors (Stehr *et al.*, 2003), leading a major focus on human pathogenic bacteria. The colonization and persistence on human skin by *Propionibacterium* has been linked with the increase in lipase activity (Gribbon *et al.*, 1993). An extracellular lipase of *Ps. aeruginosa* has been reported as an important virulence factor which acts in synergy with bacterial phospholipase C to degrade phospholipids from lung surfactants, thus promoting the invasion by this opportunistic pathogen (Jeager *et al.*, 1992). Furthermore, *Staphylococcus epidemidis*, a human skin commensal is known to be opportunistic pathogen. During the infection process, the bacterium secretes two lipases, which are believed to play a role in

supporting the growth and colonization through cleavage of sebum-derived triacylglycerols (Sterh *et al.*, 2003).

The third aspect involves the application of lipases in biotechnology. Lipases have been used extensively in organic synthesis for the production of optically active compounds (Buchholz *et al.*, 2005). The increase in applications of lipases in organic chemistry has been attributed to the following factors: lipases are highly active in a broad range of non-aqueous solvents, they often exhibit excellent stereoselectivity, they accept a broad range of esters other than triglycerides, and lipases accept nucleophiles other than water (Buchholz *et al.*, 2005). In addition to the hydrolysis and synthesis of esters, lipases have been shown to catalyze transesterification reactions, including the exchange of acyl radicals between an ester and acid, (acidolysis), an ester and another ester (interesterification) and between an ester and an alcohol (alcoholysis) (Malcata *et al.*, 1992).

1.8 Three dimensional structures of lipases

A considerable number of bacterial lipase structures have been solved. These include lipases from *B. glumae* (Noble *et al.*, 1993), *B. cepacia* (Kim *et al.*, 1997a), *Ps. aeruginosa* (Nardini *et al.*, 2000) and *B. subtilis* (Van Pouderoyen *et al.*, 2001). X-ray crystallographic and NMR studies have revealed that lipases share some common features. All lipases adopt a similar α/β hydrolase topology core (Dodson and Wlodawer, 1998; Nardini and Dijkstra, 1999). The interior topology of the α/β hydrolase fold is composed largely of parallel β-pleated strands (a minimum of five for lipases) separated

by stretches of α-helix, together forming a superhelically-twisted pleated sheet. Helical peptide sections packed on both faces of this sheet form much of the outer part of the protein (Villeneuve *et al.*, 2000). The lipases from *B. glumae* and *B. cepacia* have six parallel β strands in the central β-sheet of the α/β hydrolase fold, corresponding to strands β3 to β8 of the canonical fold (Fig. 1.2) (Noble *et al.*, 1993; Kim *et al.*, 1997a; Lang *et al.*, 1998).

Lipases are serine hydrolases and their catalytic mechanism resembles that of esterases, typically involving the catalytic triad residues, Ser-His-Asp (Section 1.3). The nucleophilic serine is embedded within a conserved G-x-S-x-G consensus motif. Mutational studies have shown that modification and replacement of the serine residue resulted in a loss of lipolytic activities, emphasizing the importance of this residue in catalysis (Davis *et al.*, 1990; Fell *et al.*, 1991a). In addition, three-dimensional structures of lipases have also revealed that the topological location of the serine residue is conserved. This residue is located at the top of a tight bend (also referred to as a nucleophilic elbow) found in all lipase structures reported to date (Nardini and Dijkstra, 1999). This nucleophilic elbow has been shown to only form when the amino acids at the -2 and +2 positions relative to the catalytic serine have small side chains, explaining predominance of glycine residues at this location (Villeneuve *et al.*, 2000).

X-ray crystallographic and NMR studies have also revealed another important structural feature, which has not been observed in any of the other known serine hydrolases-the lid domain. Previously, this domain was used to differentiate lipases from esterases (Jeager

et al., 1999). However, this criterion was proved not to be sufficient since it was later shown that some lipases do not contain the lid - domain. The role of the lid domain is discussed in details under Section 1.10.

1.9 Structural basis of *Burkholderia cepacia* lipase chiral selectivity- A case study

B. cepacia lipase is one of the most widely used enzymes for the enantiomeric resolution of esters of secondary alcohols (Theil, 1995). Detailed kinetics studies revealed that this enzyme has an enantiomeric preference for R_C over S_C compounds (Lang *et al.*, 1998). A structural view of this enzyme's enantio - selectivity towards these compounds was provided by the X-ray structure of *B. cepacia* lipase (BCL) in complex with inhibitor analog of medium alkyl chain lengths (tributyl and triocytl), R_C-(R_p,S_p)-1,2-dioctylcarbamyl-gylcerol-3-*O*-ρ-nitrophenyl octylphosphonate (TC8) (Lang *et al.*,1998).

Inhibitor binding studies revealed four binding pockets within the BCL structure: an oxyanion hole and three pockets that accommodate the *sn*-1, *sn*-2, and *sn*-3 fatty acid chains (Fig. 1.4). A well-defined hydrophobic groove was shown to be the binding site for the *sn*-3 fatty acid chain of the substrate. The *sn*-2 chain-binding site was found to be divided into small hydrophilic patch at the bottom of a cleft where the ester bond region is bound, and a larger hydrophobic patch towards the surface where the hydrophobic part of the *sn*-2 fatty acid chain is bound. Further structural analyses also revealed that the *sn*-2 chain was separated from the *sn*-1 acyl chain by residues from a calcium-binding loop. The small *sn*-1 binding site was slightly hydrophobic and was found to provide few interactions with the inhibitor (Lang *et al.*, 1998; Jaeger *et al.*, 1999).

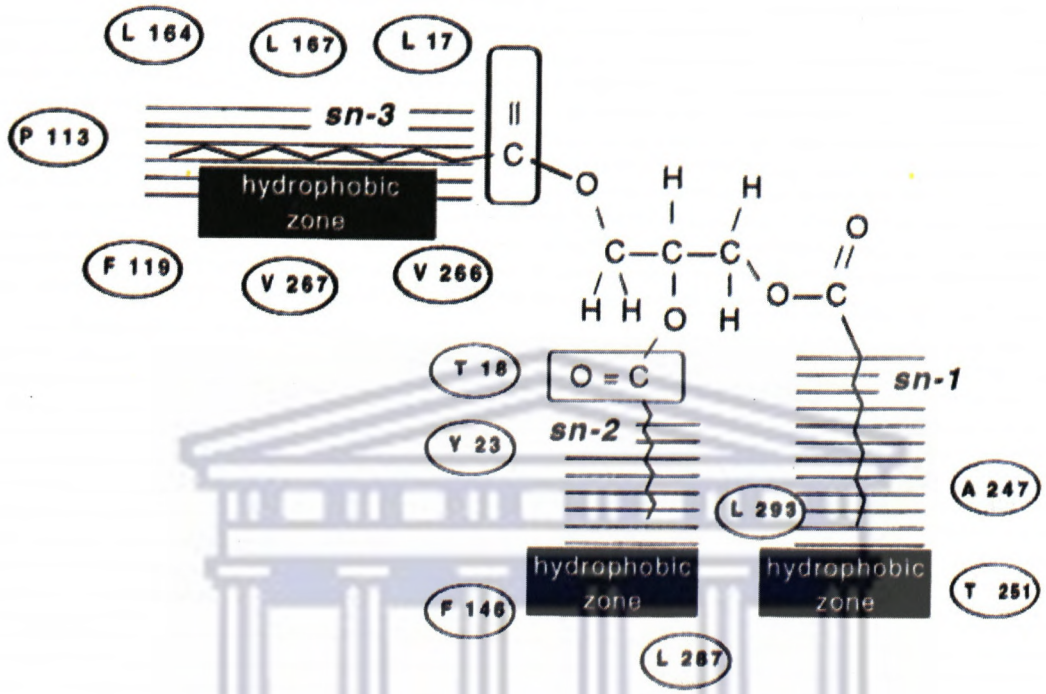


Figure 1.4: Active site of *B. cepacia* lipase. The binding pockets for the *sn*-1, *sn*-2, and *sn*-3 moieties of the lipid substrate are indicated. Also indicated are the residues lining these binding pockets (Jaeger *et al.*, 1999).

Inhibition kinetics revealed that BCL showed sevenfold preference for the (*R*)-inhibitor over the (*S*)-inhibitor, which corresponded to the *sn*-3. In order to provide evidence of the structural determinants of the enantiomeric selectivity of BCL, the authors modelled the S_C-trioctyl compound into the active site by substituent exchange at the C2 position of the glycerol moiety (Lang *et al.*, 1998). Modelling studies revealed that binding of the acyl chain to the primary hydroxyl group of the glycerol moiety clashed with the hydrophobic side chains of Leu287 and Ile290. It was postulated that to prevent this unfavourable interaction, either the amino acid side chains must move to reduce the steric conflict or

the substrate must undergo a conformational change, most probably via a rotation about a single C-C bond. The presence of such unfavourable interactions was attributed to differences in the environments of acyl binding pockets. The environment of the *sn*-1 and *sn*-2 fatty acyl chains near the stereocenter of the R_C compound was shown to be hydrophobic while that of R_C compound (*sn*-2 and *sn*-3 pocket) was found to be more hydrophilic.

Further information on the stereoselectivity of BCL has also been provided by protein engineering studies. Site directed mutagenesis studies have revealed the importance of various amino acids for the stereoselectivity of *B. cepacia* lipase (Hirose *et al.*, 1995). The combination of three mutations, V266L, L287I and F221L successfully changed the enantioselectivity from R_C to S_C compounds. The structural location of these residues revealed that V66 was located at the entrance of the acyl pocket (*sn*-3 pocket), Leu287 was at the beginning of the *sn*-2 pocket, whereas Ile290 was found located at the surface of the enzyme, approximately 20 Å from the inhibitor. Although the change in enantioselectivity following V266L and L287I mutation can be attributed to the size and width of the *sn-2* and *sn-3* pockets respectively, the influence of the F221L mutation on the enzymes's stereospecificity was unclear (Hirose *et al.*, 1995).

Previous studies have suggested that the chemical nature and/or physical state of the substrate determined the stereoselectivity of BCL (Zandonella *et al.*, 1997; Stadler *et al.*, 1995). Site directed mutagenesis (Hirose *et al.*, 1995) and structural studies (Lang *et al.*, 1998) on BCL and *Ps. aeruginosa* lipase (Nardini *et al.*, 2000) further suggested that the

differences in size and hydrophobicity/hydrophilicity ratio of the *sn*-2 and *sn*-3 binding sites determined the enantio-and regio-preferences.

1.10 Interfacial activation

The concept of interfacial activation has evolved form the observation that the activity of lipases is enhanced on insoluble substrates compared with the aqueous-soluble substrates (Verger, 1997). The structural basis of interfacial activation was elucidated in early reports of three-dimensional structures of lipases (Bradly *et al.*, 1990). Structural analysis revealed that during the process of interfacial activation, an α-helical structure (termed the "lid" domain) acts as a lid or flap covering the active site of the enzyme, making it inaccessible to solvent and substrate. Displacement of the lid exposes the catalytic site, converting the enzyme from the inactive (closed) to the active (open) conformation. The inner surface of the lid, which is exposed on opening, is sufficiently hydrophobic to facilitate association of the enzyme with a lipid interface. Thus, the exposure of the catalytic residues is accompanied by a marked increase in the non-polarity of the surrounding surface (Verger, 1997).

However, not all lipases display the concept of interfacial activation. Lipases from *B. glumae* (Noble *et al.*, 1993), *Candida Antarctica,* B (Uppenberg *et al.*, 1990) and *B. subtilis* (Van Pouderpyen *et al.*, 2001), led to a realization that the presence of lid-like structure does not necessarily correlate with interfacial activation.

1.11 Applications of the microbial lipases

At present, hydrolytic enzymes occupy a major share in industrial enzyme market, which in 2000 was estimated at around 1.5 billion U.S dollars (McCoy, 2000). Microbial proteases remain the dominant group in industrial applications mainly due to their extensive use in the detergent and dairy industries (Kirk *et al.* 2000). Multifaceted properties of microbial lipases have prompted the increase usage of these enzymes in a wide array of industrial applications. A considerable number of excellent reviews are available, describing the applications and potential applications of these enzymes at an industrial scale (Macrae *et al.*, 1989; Reetz and Jeager, 1998; Faber, 2000; Villeneuve *et al.*, 2000; Sharm *et al.*, 2001; Gupta *et al.*, 2004; Buchholz *et al.*, 2005). As it is not possible to summarize all these applications, this section gives a brief overview of some of the major application areas.

1.11.1 Applications in organic synthesis

In 2000, lipases were reported to account for about 20% of all biotransformation processes (Faber, 2000). In addition, lipases are probably the most frequently used hydrolases in organic synthesis. This stems from the fact that, in addition of catalyzing hydrolysis reactions, lipases catalyze various reverse reactions (esterification, transesterification and aminolysis) in organic solvents with high stereo- and region-specificity (Malcata *et al.*, 1992). Despite the substantial use of lipases in research and development processes, the number of industrial processes based on lipase catalysis is very limited. Some of the successful examples in this area include a DMS developed lipase based process for the production of the Captopril intermediate, (*R*)-3-chloro-2-

methyl propionate. This process makes use of *Candida cylindracea* lipase which preferentially hydrolyzed the (*S*)–enantiomer to 98% ee with 64% conversion (Buchholz *et al.*, 2005).

Another successful lipase based industrial process is the highly enantio-selective acylation of (*R, S*)-1-phenethyl amine in a process established by BASF (Buchholz *et al.*, 2005). In this process the (*R*)–amine was separated from the (*S*)-amine by distillation or extraction, and the free (*R*)–amine was released through the basic hydrolysis. A lipase from *Burkhoderia* sp. was then used to resolve a variety of aryl alkyl amines and amino alcohols (up to multi-tone scales in some cases) based on its broad substrate specificity.

1.11.2 Applications in the food industry

The physical properties and the nutritional and sensory values of triacylglycerides are influenced by three features: the position of the fatty acids, the glycerol backbone, the chain length of fatty acids and degree of saturation (Jeager and Reetz, 1998). In a commercial lipase-based technology developed by Quest-Loders Croklaan, an immobilized lipase from *R. miehei* was used in a transesterification reaction to replace palmitic acid in palm oil with stearic acid to give the desired stearic-oleic-stearic triglyceride (Jeager and Reetz, 1998). In another study, Pabai *et al.* (1995) used a lipase catalyzed transesterification reaction to decrease the degree of saturation of butter fatty acids. Furthermore, immobilized lipases from *Candida antarctica*, *Candida cylindracea*, *Humicola lanuginosa*, *Pseudomonas* sp. and *Geotrichum candidum* have been used in an

esterification of phenols to synthesise lipophilic antioxidants to be used in commercial sunflower oil (Pandey *et al.*, 1999).

1.11.3 Applications in detergent industry

One of the major areas of lipases application includes the use as additives in industrial laundry and household detergents to remove lipids stains from fabrics. Some of the successfully developed lipases for this application include Lipolase™, a recombinant lipase developed by Novo Nordisk from the fungus *Thermomyces lanuginosus* and over-expressed in *Aspergillus oryzae* (McCoy, 2000). Other examples include two bacterial lipases, Lumafast™ and Lipomax™ from *Ps. mendocina* and *Ps. alcaligene,* respectively which were developed by Genencor International (Jeager and Reetz, 1998).

1.11.4 Other applications

Lipases have also been used in the pulp and paper industries. A fungal lipase from *Candida rugosa* has been successfully used to remove the pitch (hydrophobic components of the wood including triacylglycerides and waxes) from pulp produced during paper processing. This technology has been commercialized by Nippon Paper Industry in Japan and has achieved up to 90% of triglyceride hydrolysis (Jeager and Reetz, 1998). Lipases have also been used in areas of bioremediation. A lipase from *A. oryzae* has been used to degrade hair wastes for the production of cysteine. Furthermore, lipases have been used to remove biofilm deposits from cooling water systems, to aid the

manufacture of liquid soap and to the upgrading of low value waste fats (Pandey *et al.*, 1999).

1.12 Heterologous lipase genes expression

Attempts to produce large amounts of active lipases in the cytoplasm of well characterized expression host (*E. coli*) have been marked with both success and failure. Lipases especially from the genus *Bacillus*, have been successfully over-expressed in active soluble form in *E. coli*'s cytoplasm (Dartois *et al.*, 1992 and 1994; Nthangeni *et al.*, 2001). In contrast, over-expression of lipase genes from genera such *Pseudomonas* and *Burkholderia* in *E. coli* have so far failed, always resulting in the formation of inactive protein aggregates or inclusion bodies (Svendsen *et al.*, 1995; Rosenau and Jaeger, 2000). These unsuccessful attempts have been attributed to a number of factors, including the lack of appropriate secretion machinery such as Xcp-like protein complexes and toxicity of some lipases to host strains. Although active lipases can be recovered from these protein aggregates by tedious *in vitro* denaturation and renaturation procedures, low yields have always been cited as a major problem (Quyen *et al.*, 1999 and 2004).

1.13 Role of activator proteins

Cloning and heterologous expression studies of extracellular lipase genes from Gram-negative genera such as *Burkholderia, Pseudomonas*, *Vibrio* and *Acinetobacter* revealed the requirement of an activator gene, which is necessary for lipase activity (Wohlfarth *et*

al., 1992, Frenken *et al.*, 1993, Amand *et al.*, 1994). The activator gene has also been termed Lif (lipase specific foldase) and Chap (Chaperone) (Wohlfarth *et al.*, 1992, Frenken *et al.*, 1992; Aamand *et al.*, 1994, El Khattabi *et al.*, 1999; Teo *et al.*, 2003; Su *et al.*, 2004). Currently, Lifs are classified into four families (I-IV) based on the primary structures. Family I comprises of Lifs from *Ps. aeruginosa*, *Ps. mendocina*, *Ps. wisconsinesis*, and *Ps. alcaligenes*. Family II comprises Lifs from *B. cepacia*, *B. glumae*, *Ps. fragi*, *X. fastidosa* and *R. metallidurans*. Family III and IV comprise of Lifs from *Acinetobacter* and *Vibrio* species respectively (Rosenau *et al.*, 2004).

1.13.1 Physiological roles

The physiological roles and mechanism of Lifs actions have recently been reviewed by Rosenau *et al.* (2004). It has been shown that Lifs are highly specific to their cognate lipases. Lifs and lipases originating from closely related organisms can substitute for each other, albeit with limited efficiency (Rosenau *et al.*, 2004). When the Lif from *Ps. aeruginosa* was co-expressed with lipase from *Ps. alcaligenes*, enzymatically active lipase was obtained (El Khattabi *et al.*, 1999). The co-expression of Lif from *B. glumae* with the lipases from *Ps. aeruginosa* did not result in the functional folding of lipase, irrespective of expression in either *Ps. aeruginosa* or *B. glumae* (Rosenau *et al.*, 2004).

To date, the lipase:Lif ratio needed full activation of newly synthesized lipase is not yet clear. The *in vivo* studies based co-immuno-precipitation (Hobson *et al.*, 1995) and chemical cross-linking (Shibata *et al.*, 1998), experimental approaches have suggested 1:1 ratio, while *in vitro* refolding experiments have suggested a 1:4 lipase:Lif ratio (Ihara

et al., 1995; El Khattabi *et al.*, 1999). The *in vitro* refolding studies have also shown that the formation and reactivation of the denatured lipase by its cognate chaperone is a calcium dependent process (Shibata *et al.* 1998).

1.13.2 Molecular mechanism of Lif actions

In living cells there are two classes of folding modulators, which prevent non-specific side reactions during protein folding (Ellis, 1999). The first class includes the molecular chaperones known as DnaK-DnaJ-GrpE and the GroEL-GroES, which function by preventing off-pathway reactions such as aggregation rather than by providing steric information for newly synthesized proteins to fold correctly. The second class includes proteins folding catalysts or foldases which, in contrast to molecular chaperones, catalyze the folding process by lowering a high-energy barrier in the folding pathway. Since these foldases contribute essential steric information to the mature domain of the protein, they have been termed steric chaperones (Ellis, 1998). Recent experimental evidence of *B. glumae* Lif suggests that Lifs lower energetic barriers during cognate lipase folding and are therefore steric chaperones (El Khattabi *et al.*, 2000).

1.14 The Genus *Burkholderia*

The increasing interest in the genus *Burkholderia* may be attributed to two principal causes. Firstly, the majority of *Burkholderia* species are known to be plant pathogens causing diseases such as soft rot in *Allium cepa* Burkholder (1950). Some *Burkholderia* species have also emerged as opportunistic human pathogens, particularly in individuals

with compromised immunity (LiPuma *et al.*, 1999). Secondly, members of this genus have some role in biotechnology particularly in the areas of biocatalysis, bioremediation and as biocontrol agents (Section 1.14.2).

1.14.1 Taxonomy

A genus *Burkholderia* was first described by Burkholder (1950), and belongs to the β-division of the phylum *Proteobacteria*. Early taxonomic studies resulted in some species, now assigned to the genus *Burkholderia*, being misclassified as *Pseudomonas* species. These include the *Ps. cepacia* complex and six other related species belonging to rRNA group II of the genus *Pseudomonas (Ps. solanacearum*, *Ps. picketti*, *Ps. gladioli*, *Ps. mallei*, *Ps. pseudomallei*, and *Ps. caryophylli*). These species were reassigned to genus *Burkholderia* in 1992 (Yabuuchi *et al.*, 1992). At about this time, a marked heterogeneity amongst *B. cepacia* strains from different ecological niches was also observed (Larsen *et al.*, 1993; Ursing *et al.*, 1995), leading to the assembly of *B. cepacia* strains into a group currently known as the *Burkholderia cepacia* complex (Bcc). The Bcc consists of about nine genomovars[a] (genomovar I –IX) (Coenye *et al.*, 2001).

Members of this complex share moderate levels of DNA-DNA homology (30 to 50%), but high 16S rRNA sequence similarity (98 to 99%). Furthermore, they all possess an unusually complex and plastic genome consisting of two to four replicons (chromosomes), which harbour several insertion sequences that are believed to increase the frequency of genetic mutation and recombination (Ramette *et al.*, 2005).

[a]Ursing *et al.* (1995) coined a term genomovar to denote group of strains which are phenotypically similar but genotypically distinct

Recent advancements in molecular identification approaches have allowed the re-classification of some Bcc genomovars into different species. These approaches include the development of species and/or genomovar-specific primers coupled with restriction fragment length polymorphism (RFLP) analysis. Heterogeneities within the 16S and 23S rDNAs, and the *recA* gene, have resulted in the designation of species names (LiPuma *et al.*, 1999; Coenye *et al.*, 2001; Seo and Tsuchiya, 2005; Ramette *et al.*, 2005). Genomovar I has been named *B. cepacia*; genomovar II, *B. multivorans*; genomovar III, *B. cenocepacia*; genomovar IV, *B. stabilis*; genomovar V, *B. vietnamiensis*; genomovar VI, *B. dolosa*; genomovar VII, *B. ambifarai*; genomovar VIII, *B. anthina*, and genomovar IX, *B. pyrrocinia.*

1.14.2 Application in biotechnology

There are reports of the role of the *Burkholderia* species in area of bioremediation, including the degradation of high molecular weight aromatic polycyclic aromatic hydrocarbons by *B. cepacia* (Juhasz *et al.*, 1996), the degradation of the trichloroethylene by *Burkholderia* isolates (Leahy *et al.*, 1996) and the degradation of hydrocarbons by *Burkholderia* isolates from groundwater in fixed-film bioreactors (Massol-Deya *et al.*, 1997). Members of *Burkholderia* genus are known to produce some important industrial enzymes including *B. cepacia* lipase, one of the most widely enzyme in the enantiomeric resolution of secondary alcohol esters. For example, a *B. cepacia* lipase has been used for both enatioselective acetylation of 1-phenoxy-2-hydroxybutane (Luic *et al.*, 2001) and the enantiomeric resolution of racemates, such as (*R, S*)-2-naphthyl cyanohydrin acetate (Kiruma *et al.*, 2002) and (R, S)-ethyl 3-phenylbutyrate (Koga *et al.*, 2003). A novel *B.*

cepacia esterase gene expressed in *E. coli* is capable of hydrolyzing methyl (*R, S*)-β-acetylmercaptoisobutyrate to (*R*)-β-acetylmercaptoisobutyric acid hydroxylase (Kim *et al.*, 2004a).

There are also reports describing the roles of *Burkholderia* species as effective biological control agents against foliar (Knudsen and Spurr, 1987; Huengens and Parke, 2000; Coenye *et al.*, 2001) and plant pathogens (Janisiewicz and Roitmann, 1988; Ramette *et al.*, 2005). A *Burkholderia* isolate was shown to secrete a secondary metabolite, pyrrolnitrin, which inhibits plant pathogens such as *Rhizoctonia solani* (Hwang *et al.*, 2002). Although, *Burkholderia* species offer an alternative as pesticides in agriculture, their commercial application has been hampered by the possible risks they pose to human health.

1.15 Aims and objectives

The initial aim was to screen bacterial sources for their ability to convert ethyl ferulate to ferulic acid, and to access lipase or esterase genes by screening at the molecular level. The overall objective of the object was to identify, characterize and possibly purify enzyme(s) that could be used for the development of biocatalytic process for modification of ferulic acid derivatives.

The specific objectives of this study were therefore:

1. To isolate, screen and identify novel bacterial isolates with ethyl ferulate hydrolyzing activity

2. To clone and sequence the genes encoding esterolytic and lipolytic activities from the selected isolate
3. To heterologously express the gene(s) in the suitable host expression system,
4. To purify and characterize the enzyme(s)

CHAPTER TWO

Materials and Methods

2.1 Materials

All reagents used in this study were of analytical or molecular biology grade. Unless stated otherwise reagents were purchased from Sigma Chemical Co. (Germany) or Merck (Germany). Restriction enzymes, DNA ligase and Shrimp alkaline phosphatase (SAP) were from Fermentas Life Sciences (Vilnius, Lithuania). Oligonucleotides used in this study were synthesized by Inqaba Biotech (South Africa) and were designed with an aid of PerlPrimer v1.1a software and OLIGO-calculator available on InqabaBiotech website (www.Inqababiotech.co.za). Organic solvents were from Kimix (South Africa).

Small-scale cultures were performed on an orbital shaking incubator from Vision Scientific Co. Ltd. Agar plates were incubated in Gallenkamp series incubators. Centrifugation was achieved using Beckman J2-21 centrifuge (with JA-14 or JA-20 rotors) or Eppendorf (5810R and 5415D) centrifuge. Absorbance readings were measured using a Biomate spectrophotometer from Thermo-electron Corporation. Media, Buffers, and glassware were autoclaved at 1.0 kg cm^{-2}, 121 ^{o}C for 20 min using Austester-G autoclave. The pH of the solutions were measured using Crison pH meter BasiC 20. Distilled and double distilled (ddH_2O) water were purified with reverse osmosis RiOs^{TM} purication system (MILLIPORE).

2.2 Media

Strains (*B. multivorans* and *E. coli*) were routinely cultured in nutrient rich broth or nutrient agar (Difco) according to manufacturer's instructions. Culture media used for special applications during this study are described below. For the preparation of the agar plates, 1.5 % (w/v) bacteriological agar was added. For M9 minimal based medium, purified agar (AGAR) was used (Bio Basic Inc).

M9 minimal medium (Sambrook and Russell, 2001)

5x M9 Salts (Autoclaved separately)

Constituent	**L^{-1}**
$Na_2HPO_4.7H_2O$	64 g
KH_2PO_4	15 g
NaCl	2.5 g
NH_4Cl	5.0 g

Complete M9 minimal medium (the following components were added aseptically)

Constituent	**L^{-1}**
5x M9 salts	200 ml
1 M $MgSO_4$	2 ml
1 M $CaCl_2$	0.1ml
Carbon source (20%)	(Appropriate amount)

1 M $MgSO_4$, 1 M $CaCl_2$, and the carbon source stock solution (20%) were filter sterilized.

M9-EF (Ethyl ferulate) minimal medium

Constituent	L^{-1}
5x M9 salts	200 ml
1 M $MgSO_4$	2.0 ml
1 M $CaCl_2$	0.1 ml
Ethyl ferulate	(0.1-0.5%)

LB Medium (Luria-Bertani Medium)

Constituent	L^{-1}
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

The pH was adjusted to pH 7.0 with 5 N NaOH

SOB medium

Constituent	L^{-1}
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
KCl (250 mM)	10 ml

The pH was adjusted to 7.0 before autoclaving. After autoclaving the broth was cooled to ~50 °C and the following filter sterilized solution (L^{-1}) was aseptically added:

$MgCl_2$ (2 M)	5ml

SOC Medium

Constituent	L^{-1}
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
KCl (250 mM)	10 ml

The pH was adjusted to 7.0 before autoclaving. After autoclaving the broth was cooled to ~50 °C and the following filter sterilized solutions were aseptically added:

$MgCl_2$ (2M)	5 ml
Glucose (1M)	20 ml

2.3 Buffers

Table 2.1: Buffers and solution preparations

Buffer	Components	pH
6x Agarose loading buffer	30% (v/v) glycerol; 0.25% (w/v); bromophenol blue; 15% (w/v) ficoll (Type 400)	
Alkaline solution I	50 mM glucose; 25 mM Tris-HCl (pH 8.0); 10 mM EDTA (pH 8.0)	
Alkaline solution II	0.2 N NaOH; 1% (w/v) SDS	
Alkaline solution III	5 M potassium acetate; 11.5 ml glacial acetic acid; 28.5 dH_2O	
Gradient Buffer 1	500 mM NaCl; 20 mM Tris; 8 M urea 20% (v/v) glycerol	**7.4**
Gradient Buffer 2	500 mM NaCl; 20 mM Tris; 20% (v/v) glycerol	**7.4**
Imidazole elution Buffer	500 mM NaCl; 20 mM Tris; 20% (v/v) glycerol; 250 mM imidazole	**7.4**
Inoue transformation buffer	45 mM $MnCl_2.4H_2O$; 15 mM $CaCl_2.2H_2O$; 250 mM KCl; 10 mM PIPES (pH 6.7)	**6.7**
Phosphate buffer saline	137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 2 mM KH_2PO4	**7.6**
1M Potassium phosphate buffer	717 ml 1 M K_2HPO_4; 283 ml 1 M KH_2PO_4	**7.2**
10x SDS PAGE electrode Buffer	0.25 M Tris-HCl; 2 M glycine; 1% (w/v) SDS	**8.3**
2x SDS PAGE sample buffer	0.125 M Tris-HCl (pH 6.8); 20% (v/v) glycerol; 4% (w/v) SDS; 0.2% (w/v) bromophenol blue; 200 mM DTT	**8.0**
Coomassie staining solution I (CSMI)	10% (v/v) acetic acid; 0.125% (w/v) coomassie brilliant blue G; 25% (v/v) isopropanol	

Coomassie staining solution II (CSMII)	10% (v/v) acetic acid; 0.003% (w/v) coomassie brilliant blue G; 10% (v/v) isopropanol	
Coomassie staining solution III (CSMIII)	10% (v/v) acetic acid; 0.003% (w/v) coomassie brilliant blue G	
Refolding Buffer1	100 mM NaH_2PO_4; 100 mM Tris; 8 M urea	**8.0**
Refolding Buffer2	100 mM NaH_2PO_4; 100 mM Tris; 8 M urea	**7.0**
SDS PAGE destaining solution	10% (v/v) acetic acid; 1% (v/v) glycerol	
TE buffer	10 mM Tris-HCl (pH 8.0); 1 mM EDTA (pH 8.0)	**8.0**
0.5x TBE	45 mM Tris-borate; 1mM EDTA	

2.4 Strains and vectors

Table 2.2: Strains and vectors used in this study

Strains/Plasmids	Genotype or relevant characteristics (Markers)	Source or Reference
Burkholderia multivorans UWC10	Esterase-lipase producer	This study
E. coli (DH5α)	[*supE44 ΔlacU169 (φ80 lacZΔM15) hsd*R17 *rel*A1 *gyr*A96 *thi-1 rec*A1]	UWC culture collection
***E. coli* BL21[DE3]**	[(r_B-m_B-*omp*TF$^-$) *[lon]* *hsd*S_B] with DE3 a λ prophage carrying the T7 RNA polymerase.	Novagen
pUC18	High copy number cloning vector. Insertion into multiple cloning sites disrupts lacZα reading frame. Recombinants are identified by blue white screening. (Ap^r)	(Yanisch-Perron *et al.* 1985)
pTZ57R/T	High copy number TA cloning vector. Insertion into multiple cloning sites disrupts lacZα reading frame. Recombinants are identified by blue white screening. (Ap^r)	Fermentas Life Sciences

pMOS*Blue*	High copy number blunt-end cloning vector. Insertion into multiple cloning sites disrupts lacZα reading frame. Recombinants are identified by blue white screening. (Ap^r, Tet^r)	Amersham Biosciences
pET22b (+)	High level expression vector. Expression of the target genes is driven by T7lac promoter upon induction with IPTG. The vector is designed to allow fusion of the target proteins to pelB leader peptide. (Ap^r)	Novagen
pMS470Δ8	High level expression vector. Expression of the target genes is driven by P_{tac} promoter on induction with IPTG. (Ap^r)	Balzer *et al.* (1992)
pMAL–p2X	High level expression vector. Expression of the target genes is driven by P_{tac} promoter on induction with IPTG. The vector is designed to allow fusion of the target proteins to *E. coli malE* gene sequence to facilitate export to the cytoplasm. (Ap^r)	New England Biolabs Inc
pBurk16S	pTZ57R/T derivative carrying 1.5 kb amplified 16S RNA gene fragment. (Ap^r)	This study
pRASH14	pUC18 derivative carrying 7.5 kb fragment encoding lipase activity. (Ap^r)	This study
pTEND5	pUC18 derivative carrying 2.1 kb fragment encoding ethyl ferulate hydrolyzing esterase activity. (Ap^r)	This study
pHOLA6	pUC18 derivative carrying 4.5 kb fragment encoding esterase activity. (Ap^r)	This study
pMOSLipAB	pMOS*Blue* derivative carrying a full length 2.2 kb lipase-chaperone fragment. (Ap^r, Tet^r)	This study
$pMOSLip_{signal}$	pMOS*Blue* derivative carrying a truncated	This study

	2.1 kb lipase chaperone fragment without N-terminal signal sequence. (Ap^r, Tet^r)	
pMOSEFH5	pMOS*Blue* derivative carrying a full length 978 bp EstEFH5 fragment. (Ap^r, Tet^r)	This study
pMOSEFH6	pMOS*Blue* derivative carrying a truncated 948 bp EstEFH5 esterase fragment. (Ap^r, Tet^r)	This study
pMOSEstBL	pMOS*Blue* derivative carrying a full length 1.18 kb EstBL fragment. (Ap^r, Tet^r)	This study
pETLip	pET21a (+) derivative carrying a full length 2.2 kb lipase-chaperone fragment. (Ap^r)	This study
pETLip$_{singal}$	pET21a (+) derivative carrying 2.1 kb lipase-chaperone fragment without leader signal. (Ap^r)	This study
pETEFH5	pET22b (+) derivative carrying a full length 978 bp EstEFH esterase fragment. (Ap^r)	This study
pETEFH5$_{tag}$	pET22b (+) derivative carrying a full length 978 bp EstEFH esterase fragment fused to 6x His-tag. (Ap^r)	This study
pETEFH6	pET22b (+) derivative carrying a truncated 948 pb EstEFH esterase fragment. (Ap^r)	This study
pMALEFH5	pMAL–p2X derivative carrying a full length 978 bp EstEFH fragment. (Ap^r)	This study
pETESTBL	pET22b(+) derivative carrying a full length 1.18 kb EstBL fragment. (Ap^r)	This study
pMSESTBL	pMS470Δ8 derivative carrying a full length 1.18 kb EstBL fragment. (Ap^r)	This study

(Ap^r =Ampicillin resistance; Tet^r = Tetracycline resistance)

2.5 Microbial isolation and biotransformation methods

2.5.1 Selective enrichment

Samples used for isolation of ethyl ferulate hydrolyzing microorganisms were maize silage collected from a maize processing farm in Stellenbosch (South Africa). Maize silage samples (1.0 g) were resuspended in 10 ml of sterile milli-Q water containing 2% (v/v) Tween 20. Samples were vortexed for 5 min to dislodge microorganisms and allowed to stand for 1 h at room temperature before aliquots (100 µl) of serially ($10\text{-}10^{-5}$) diluted samples were aseptically spread onto M9-EF minimal medium plates containing various concentrations of ethyl ferulate (0.1-0.5%) as a sole carbon source. Filter sterilized cyclohexamide 0.1% (w/v) was added to the medium to prevent fungal growth. The plates were incubated aerobically at a range of temperatures (25, 30, 37, 45, 50 °C) for up to seven days. Colonies were repeatedly streaked on the same minimal nutrient medium until pure cultures were obtained.

2.5.2 Biomass production

Selected bacterial isolates were routinely cultured in 500 ml Erlenmeyer flasks containing 100 ml M9-EF minimal medium, or nutrient broth with or without ethyl ferulate (EF) as required. Samples were grown aerobically with shaking and growth was monitored by measuring the change in optical density (OD) at 600 nm using sterile medium as a blank.

2.5.3 Preparation of extracellular fractions

After the growth in the liquid media, cultures were harvested by centrifugation (5000 x g, 5 min, at 4 °C). Supernatants were recovered (whole cells were kept for preparation of total cell lysate), concentrated using 15 ml protein spin columns with a 10 kDa cut-off membrane (Centriprep YM-10, Millipore) and used as a source of crude extracellular enzymes.

2.5.4 Preparation of cell lysate (soluble) fraction

Whole cells obtained as described (Section 2.5.3) were washed twice with sterile ddH_2O and resuspended at a density of 0.5 g l^{-1} (wet weight) in (3-(*N*-morpholino) propanesulfonic acid) MOPS buffer (50 mM, pH 7.0). Cells were disrupted by sonication using Sonoplus HD-070 (Bandelin, Germany) set at: 5x cycle burst, 50% max. amplitude for 2 min, at 30 sec interval, in an ice bath and cells debris was removed by centrifugation (16 000 x g, 30 min, at 4 °C). Supernatants were used as the source of crude intracellular fraction.

2.5.5 Biotransformation of EF in liquid cultures

Pure colonies from positive isolates were grown overnight in nutrient broth, washed twice by repeated centrifugation (5000 x g, 5 min at 20 °C) and resuspended in sterile ddH_2O. Inocula (10% v/v) of overnight cultures were used to inoculate 100 ml of either M9-EF minimal medium or nutrient broth containing 0.1% (v/v) EF substrate solution in 500 ml Erlenmeyer flasks. Cultures were grown aerobically at 30 °C with shaking (150

rpm). Samples (1 ml) were aseptically withdrawn at defined time intervals, centrifuged (16000 x g, 5 min, at room temperature) and the released ferulic acid (FA) in the culture broth monitored by Thin Layer Chromatography (TLC) and High Performance Liquid Chromatography (HPLC). A control flask without inoculum was included as a negative control to monitor background hydrolysis of the substrate under these conditions. Results are presented as mean values of triplicate assays.

2.5.6 Biotransformation of EF by resting cells

Cultures were grown aerobically as described previously (Section 2.5.5), until they reach mid stationary growth phase (OD_{600nm} = 1.2), and then harvested by centrifugation (5000 x g, 5 min, 4 °C). The pellets were washed twice by repeated centrifugation (5000 x g, 5 min at 20 °C) with sterile ddH_2O, and resuspended in MOPS-NaOH buffer (50 mM, pH 7.0) at a density of 0.5 g l^{-1} (wet weight). Ethyl ferulate (0.2% v/v) was added to the resting cells then incubated at 30 °C with gently shaking at 100 rpm. Samples (1 ml) were aseptically withdrawn at defined time intervals, centrifuged (16000 x g, 5 min, at room temperature) and the supernatants were analyzed for the release of FA by TLC and HPLC. A negative control flask without cells was included under identical conditions to monitor background hydrolysis of the substrate. Results are presented as mean values of triplicate assays.

2.6 Analytical Methods

2.6.1 Qualitative plate screening assays

2.6.1.1 Ferulic acid esterase agar plate assay

Ethyl ferulate agar plates were prepared essentially as described by Donaghy *et al.* (1998). The medium (100 ml) containing 1.5% (w/v) agar in distilled water was autoclaved then allowed to cool to ± 50 °C. While stirring rapidly, 0.4% (v/v) of filter sterilized EF in dimethyl sulphoxide (DMSO) was added. The medium turns cloudy because of the EF precipitating in the bulk aqueous phase. The plates were allowed to solidify and dry (3 h, 25 °C). Wells (10 mm in diameter) were punched into the plates using the back of sterile pipette tip and cell lysate was added. A zone of clearance is evidence of ferulic acid esterase activity. Fungal ferulic acid esterase (Biocatalysts, England) was used as a positive control.

2.6.1.2 Tributyrin agar plate assay

Tributyrin agar plates were prepared as described by Ro *et al.* (2004). To prepare agar plates, 10 ml of tributyrin and 90 ml of a 1% (w/v) Gum Arabic solution were mixed by sonication (5x cycle burst, 50% max. amplitude for 2 min, at 30 sec). The white cloudy solution was mixed with 900 ml of MOPS buffer (50 mM, pH 7.0) containing 1.5 % (w/v) agar. After autoclaving, 25 ml aliquots of the tributyrin agar solution were poured into Petri dishes and allowed to solidify at room temperature before punching wells. Following incubation in the presence of the test solution at 30 °C overnight, esterase or lipase activity is observed as a zone of clearance around the wells.

2.6.1.3 Olive oil-Rhodamine B lipase plate assay

Olive oil-Rhodamine B plates were prepared using modified method of Kouker and Jaeger, (1987). Olive oil (10 ml) and 90 ml of a 1% (w/v) Gum Arabic solution were mixed by sonication (5x cycle burst, 50% max. amplitude for 30 sec). The white cloudy solution was mixed with 900 ml of MOPS-NaOH buffer (50 mM, pH 7.0) containing 1.5 % (w/v) agar. After autoclaving, the following filter sterilized solutions were aseptically added before pouring the plates: 1 ml of 0.5% (w/v) Rhodamine B dye and 0.5 ml of 0.5 % (w/v) of $CaCl_2$. Samples (100 µl) of the extracellular or cell free fraction were added into the wells and incubated at 30 °C overnight. Lipase activity is observed as a fluorescent halo under long range UV light (300-350 nm).

2.6.1.4 Screening of esterase positive *E. coli* transformants

E. coli (DH5α) cells transformed with recombinant plasmids were plated on LB agar plates supplemented with 0.1 mM isopropyl-β-D-thiogalactoside (IPTG), ampicillin (100 µg ml^{1}), 1% (v/v) tributryrin and 0.1% (w/v) Gum Arabic. Plates were incubated at 25 °C for up to 7 days. Clones showing halos were replicated onto fresh plates, before growth in liquid culture for glycerol stocks and plasmid isolation.

2.6.1.5 Screening of lipase positive *E. coli* transformants

E. coli (DH5α) cells, transformed with recombinant plasmids were plated on LB agar plates supplemented with 0.1 mM isopropyl-β-D-thiogalactoside (IPTG), ampicillin (100 µg ml^{-1}), 0.05% (w/v) Rhodamine B, 1% (v/v) olive oil, 0.1% (w/v) gum arabic and 0.5

% (w/v) $CaCl_2$. Plates were incubated at 30 °C for up to 7 days. Clones showing fluorescent halos were replicated onto fresh plates, before growth in liquid culture for glycerol stocks and plasmid isolation.

2.6.2 Quantitative assays

2.6.2.1 Esterase assay I with ρ-nitrophenyl acetate

For routine enzyme assay, esterase activity was determined as essentially described by Petersen *et al.* (2001). The reactions were performed in 1 ml volumes containing Tris-HCl (50 mM. pH 7.5), 1 mM ρ-nitrophenyl acetate as the substrate and an appropriate amount of enzyme extract. Substrate stock solution (100 mM ρ-nitrophenyl acetate in methanol) was prepared fresh. The reaction was carried out at 25 °C. The reaction rates were monitored by measuring the release of ρ-nitrophenol at 405 nm in 1 cm path length cuvettes. The background hydrolysis of the substrate was deducted using reference samples without enzyme under identical conditions. The molar co-efficient of extinction under these conditions was 14800 M^{-1} cm^{-1}. One unit (1U) of the esterase activity was defined as the amount of esterase catalyzing the release of 1 µmol of ρ-nitrophenol per min from ρ-nitrophenyl acetate at 25 °C. The specific activity was defined as units per mg of protein (U mg^{-1}). A calibration curve of ρ-nitrophenol under these conditions is show in Appendix 2.I.

2.6.2.2 Esterase assay II with β-naphthyl acetate

Esterase assay against β-naphthyl esters was performed by measuring the release of β-naphthol essentially as described by Prim *et al.* (2001). The assay mixtures containing 0.5 ml Tris-HCl (100 mM, 7.5), 1.75 ml dH_2O and 0.15 ml substrate solution (2 mM β-naphthyl acetate in methanol) were pre-incubated at 30 °C for 2 min. The reaction was started by adding appropriate amount of diluted enzyme extract and incubated at 30 °C for 5 min. The reaction was terminated by adding 0.5 ml of Fast Garnet GBC dye solution (1 mg ml^{-1} in 10% SDS w/v) and incubated further at room temperature for 20 min. The background hydrolysis of the substrate was deducted using reference samples without enzyme under otherwise identical conditions. The molar co-efficient of extinction under these conditions was 10100 M^{-1} cm^{-1}. One unit (1U) of the esterase activity was defined as the amount of esterase catalyzing the release of 1 μmol of β-naphthol per min from β-naphthyl acetate at 30 °C. The specific activity was defined as units per mg of protein (U mg^{-1}). A calibration curve of β-naphthol under these conditions is show in Appendix 2.II.

2.6.2.3 Lipase assay with ρ-nitrophenyl palmitate

Lipase assay was performed using the method of Gupta *et al.* (2002). Substrate solution A (7.9 mM ρ-nitrophenyl palmitate in isopropanol) and solution B [Tris-HCl buffer (50 mM, pH 7.5) containing 0.1% (w/v) sodium cholate, 0.1% (w/v) Gum Arabic, 2% (v/v) Triton x-100] were mixed at a ratio of 1:9 to produce solution C (with the final substrate concentration of 0.79 mM). The reaction was carried out with a 2.45 ml volume of solution C and 0.05 ml of diluted enzyme extract. The reaction rates were monitored as

described in section 2.6.2.1. The molar co-efficient of extinction under these conditions was 2700 $M^{-1}cm^{-1}$ (Gupta *et al.*, 2002).

2.6.2.4 Ferulic acid esterase (FAE) assay

FAE activity assay was performed using method of Andersen *et al.* (2002). Reactions were carried out in 1 ml volumes containing MOPS-NaOH buffer (50 mM. pH 6.5), 1 mM ethyl ferulate as substrate and appropriate amount of enzyme extract. Substrate stock solution (50 mM ethyl ferulate) was prepared in a DMSO-water solution (2:1). The reaction was started by addition of 10 µl diluted enzyme extract at 30 °C. After 20 minutes incubation the reaction was stopped by adding 100 µl of 10% (v/v) acetic acid solution. Samples (500 µl) were transferred to HPLC vials and analyzed by HPLC as described in Section 2.6.4. The release of FA was quantified by calculating the area of the FA peak at 315 nm. The background hydrolysis of the substrate was deducted using reference samples without enzyme under otherwise identical conditions. One unit of the enzyme activity was defined as the amount of enzyme required to release 1µmol of FA min^{-1} from ethyl ferulate substrate at 30 °C. A calibration curve of the FA standards determined under these conditions is shown in the Appendix 2.III.

2.6.2.5 ß-lactamase assay

ß-lactamase activity was determined by a direct spectrophotometric method of Avison *et al.* (2000) in 1 cm light-path cuvettes with readings recorded every 2 s for 3 min at the wavelength for optimal absorbance of the ß-lactam substrate. *B. cereus* lactamase (Calbiochem, Germany) was used as a positive control. Hydrolysis of the ß-lactam ring was measured by the decrease of optical density of the cephalotine, cephalosporin C and ampicillin solution at 255, 265 and 233 nm respectively. Substrate solutions were prepared fresh in 10 mM phosphate buffer (pH 7.0) to final concentrations of 100 µM. Background readings of each antibiotic were performed prior to each assay, ensure that any decrease in substrate was solely the result of adding the enzyme extracts. The molar co-efficient of extinctions for cephalotine, cephalosporin C, ampicillin under these conditions were 7374, 2500 and 656$M^{-1}cm^{-1}$ respectively (Avison *et al.*, 2000). One enzyme unit (U) was defined as the amount of enzyme that released 1 µmol of antibiotic per ml per min under the given assay conditions. Specific activity was expressed as enzyme units per mg of protein.

2.6.2.6 Protein Determination

Protein concentrations were determined by the method of Bradford (1976) using bovine serum albumin (BSA) as a standard. The assay was carried out as follows: 5 µl of sample plus 95 µl dH_2O were added to 1.5 ml Eppendorf tubes followed by addition of 900 µl Bradford reagent. The mixture was incubated for 5 min at room temperature before measuring the absorbance at 595 nm. A complete assay mixture containing sterile dH_2O

instead of the protein samples served as a blank. The calibration curve is shown in the Appendix 2.V.

2.6.3 Thin layer chromatography (TLC) analysis

After biotransformation of EF by selected isolates, supernatants were acidified to pH<2 with 10 N HCl, extracted twice with diethyl ether, then dried over anhydrous Na_2SO_4 and concentrated by evaporation under N_2:CO_2 (80:20). Extracts were dissolved in a minimal volume of methanol and chromatographed on 5 x 10 cm silica gel F_{20} TLC plates (Merck), together with commercial ferulic acid as a positive control. The mobile phase was chloroform: methanol: formic acid (85:15:1 v/v/v). The spots were visualized at 254 nm and the R_f values were measured.

2.6.4 High performance liquid chromatography (HPLC)

Supernatants were analyzed with a LaChrom HPLC system equipped with UV detection and an 80 position auto-injection sampler. Separation was achieved on a Wakosil II C18 reverse phase column with a gaurd column. The detector wavelength was set at 284 nm. The mobile phase was water: methanol: acetic acid (80:20:2.5 v/v/v) at a flow rate of 1 ml min^{-1}.

2.6.5 Agarose gel electrophoresis

DNA fragments and plasmids were separated in 0.8-2.5 % (w/v) agarose gels prepared in 0.5 x TBE buffer containing 0.5 μg ml^{-1} ethidium bromide solution (Sambrook and Russell, 2001). Prior to electrophoresis, samples were mixed with 6x agarose loading buffer. *Pst*I digested λDNA was used as a molecular weight marker. Electrophoresis was performed in 0.5x TBE buffer at voltage range of 20-100 V using a GNA 100 gel electrophoresis apparatus (Amersham Bioscience (Uppsala Sweden)). Ethidium bromide stained agarose gels were visualized under ultraviolet light, 300 nm and photographed with a digital imaging system (AlphaImager 2000, Alpha Innotech, San Leadro, CA).

2.6.6 Denaturing SDS-polyacrylamide gel electrophoresis

Denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmli (1970) using a Mighty small™ SE 280 vertical Slab unit (Hoefer Inc, USA) with 1 mm or 1.5 mm gels containing 10-15% acrylamide (Table 2.3 and 2.4). Protein samples (8 μl) were prepared by suspending in SDS-PAGE loading buffer (2 μl) and were denatured by boiling in sealed tubes for 5 min, followed by centrifugation (16000 x g, 2 min, at room temperature). The samples were loaded onto the gel and electrophoresed at room temperature under constant voltage 60 V until the dye front migrated into the separating gel, followed electrophoresis at 120 V. Pre-stained SDS-PAGE protein molecular weights markers were from Fermentas Life Science (Vilnius, Lithuania): 116 kDa β-galactosidase, 66.2 kDa bovune serum albumin, 45 kDa Lactate dehydrogenase, 35 kDa REase Bsp981, 18.4 β-lactoglubin and 14.1 kDa lysozyme.

Table 2.3: SDS-PAGE resolving gel constituents

Separation Gels in 0.375 M Tris-HCl pH 8.8				
% Acrylamide	7%	10%	12%	**15%**
Constituents	Volume (ml)	Volume (ml)	Volume (ml)	**Volume (ml)**
dH_2O	7.65	6.15	5.1	**3.6**
1.5 M Tris-HCl (pH 8.8)	3.75	3.75	3.75	**3.75**
20% (w/v) SDS	0.075	0.075	0.075	**0.075**
Acrylamide(Bis-acrylamide (30%/0.8% w/v)	3.45	4.95	6	**7.5**
10% (w/v) ammonium persulphate	0.075	0.075	0.075	**0.075**
TEMED (Tetramethylethylenediamine)	0.0075	0.0075	0.0075	**0.0075**
Total	15	15	15	**15**

Table 2.4: SDS-PAGE stacking gel constituents

Stacking Gels: 0.4% in 0.125 M Tris-HCl pH 6.8	
Constituents	**Volume(ml)**
dH_2O	**3.07**
0.5 M Tris-HCl (pH 6.8)	**1.25**
20% (w/v) SDS	**0.025**
Acrylamide (Bis-acrylamide (30%/0.8% w/v)	**0.67**
10% (w/v) ammonium persulphate	**0.025**
TEMED (Tetramethylethylenediamine)	**0.05**
Total	**5.05**

2.6.7 Native polyacrylamide gel electrophoresis

Non-denaturing (native) polyacrylamide gel electrophoresis was performed under identical conditions except that samples were not boiled and SDS was omitted from all electrophoresis solutions.

2.6.8 Staining and Destaining of PAGE gels

Following electrophoresis, gels were stained with three Coomassie Brilliant Blue staining (CBS) solutions as follows: Gels were firstly soaked in 50 ml CSMI, heated for 30 sec in a microwave, followed by incubation at room temperature with shaking for 20 min. The CSMI solution was then discarded, the gels rinsed in dH_2O, followed by soaking in CSMII for 20 min with shaking after brief heating for 30 sec in the microwave. Staining with CSMIII was performed using the same procedure. Destaining was carried out overnight by soaking in SDS-PAGE destaining solution.

2.6.9 Esterase activity staining

Purified enzyme was electrophoresed on native-PAGE gels as previously described (Section 2.6.7) with an exception that samples were eletrophoresed under constant voltage of 30 V. Following electrophoresis, esterase activity was detected *in situ* by incubating the gel at 25 °C in a 20 ml substrate solution containing: Tris-HCl buffer (20 mM, pH 7.5), 2 ml of substrate stock solution (1% [w/v] α-naphthyl acetate in acetone) and 500 µl of fast blue B stock dye solution (2% [w/v]) in dH_2O). The reaction was allowed to proceed for 5 min at room temperature before stopped by soaking the gel in a

10 % (v/v) glacial acetic acid solution. Esterase active was detected as a dark red band (Gudelj *et al.*, 2002).

2.7 Preparation of *E. coli* competent cells and transformation procedures

2.7.1 Chemical competent cells

Chemically competent cells were prepared by method of Inoue *et al.* (1990). A single, fresh *E. coli* colony (2-3 mm in diameter) was used to inoculate 20 ml of SOB medium in a 250 ml Erlenmeyer flask, followed by growth for 8 h at 37 °C with vigorous shaking (250 rpm). The starter culture (10 ml) was used subsequently to inoculate 250 ml volumes of SOB medium in 2 L Erlenmeyer-baffled flasks. Cultures were then grown at 18 °C with gentle shaking (120 rpm) until OD_{600nm} reached ~0.55. Cultures were chilled on ice bath for 10 min, followed by centrifugation (3000 x g, 10 min, 4 °C). The cell pellets were resuspended by gently swirling in 100 ml ice-cold Inoue buffer, and then harvested by centrifugation (3000 x g, 10 min, 4 °C). The washing step was repeated twice. The final pellets were resuspended in 5 ml ice-cold Inoue buffer containing 1% (v/v) DMSO at a final resuspension density of 2 x 10^{11} cells/ml. Aliquots (100 µl) were transferred into pre-chilled sterile 0.5 ml Eppendorf tubes before snap-freezed by immersing in the liquid nitrogen. The tubes containing competent cells were then stored at -80 °C.

2.7.2 Eletrocompetent cells

Electrocompetent cells were prepared as described by Sambrook and Russell (2001). A single fresh colony was used to inoculate 5 ml of LB broth and grown overnight at 37 °C with shaking at 150 rpm. The overnight starter culture was used to subsequently inoculate 500 ml of LB broth in 2 L Erlenmeyer flask, followed by growth under the same conditions until the culture OD_{600nm} reached ~0.6. Cultures were chilled on ice for 10 min, followed by centrifugation (3000 x g, 10 min, 4 °C). The harvested cell pellets were washed twice in 500 ml ice-cold 10 % (v/v) glycerol. The washed cell pellets were resuspended in 5 ml volume at a final resuspension density of 2 x 10^{11} cells/ml. Aliquots (100 µl) were transferred in pre-chilled sterile 0.5 ml Eppendorf tubes before snap-freezed by immersing in the liquid nitrogen. The tubes containing competent cells were then stored at -80 °C.

2.7.3 Transformation of electrocompetent cells

Electrocompetent cells were first thawed on ice for 10 min before electroporation. Cell suspensions (50 µl) were gently mixed with 1 µl (~1 ng) of ligation mixture before transfering into 2 mm electroporation cuvettes. Cells were shocked once in the electroporator (Gene pulse, Bio-Rad Laboratories, Hercules, CA) set at 25 µF capacitance, 1.8 kV voltage, 200 Ω resistance. Following electroporation, 950 µl of SOC medium was immediately added then transferred in 1.5 ml Eppendorf tubes for incubation with shaking (100 rpm) at 37 °C for 1-1.5 h.

2.7.4 Transformation of $CaCl_2$ treated chemical competent cells

Cell suspensions (50 µl) together with up to 5 µl (~5 ng) of ligation mixture were gently mixed in 1.5 ml Eppendorf tubes followed by incubation on ice for 20 min. The mixtures were heat shocked (42 °C, 50 sec), followed by cold shock on ice for 2 min. SOC medium (950 µl) was immediately added and the suspension was incubated with shaking (100 rpm) at 37 °C for 1-1.5 h.

Transformed cells were plated on appropriate agar media and grown overnight at 37 °C. In both transformation procedures the efficiency of the cells were checked by transforming 10 pg of the supercoiled plasmid DNA. Negative transformation control reactions did not contain either the plasmid DNA or the ligation mixture (i.e. cells only).

2.8 DNA isolation, purification and quantification

2.8.1 DNA isolation

Genomic DNA of *B. multivorans* UWC10 was isolated by modified procedure of Marmur (1963), from late-stationary growth phase (OD_{600nm}= 2.0). Cultures were grown in 100 ml nutrient broth with shaking (150 rpm) at 30 °C in 500 ml baffled Erlenmeyer flasks. The cells were harvested by centrifugation (5000 x g, 10 min, 10 °C), washed twice in TE buffer (pH 8.0), then resuspended in 9 ml of TE buffer (pH 8.0) containing 0.1 g l^{-1} lysozyme solution, followed by shaking incubation (37 °C, 150 rpm, for 2 h). 0.5 ml of 10% (w/v) SDS and 0.1 ml of 20 mg ml^{-1} proteinase K solution were then added, and the suspensions were incubated for a further 2 h at 37 °C. Following lysis, 1.8 ml of 5

M NaCl and 1.5 ml of 10% (w/v) acetyltrimethylammonium bromide (CTAB) in 0.7 M NaCl were added and the suspensions were incubated for 30 min at 65 °C.

An equal volume of a chloroform:isoamyl alcohol (24:1) mixture was added to the suspensions and gently mixed by inversion at room temperature for 10 min and followed by centrifugation (10 000 x g, 20 min, at 25 °C) to separate the phases. After two extractions with chloroform:isoamylalcohol (24:1) mixture, an equal volume of Tris-HCl (0.5 M, pH 8.0) buffered phenol was added and the mixtures were gently mixed by inversion followed by centrifugation (10 000 x g, 20 min, at 25 °C) to achieve phase separation. The phenol: chloroform: isoamyl alcohol (25:24:1) step was repeated twice followed by an additional chloroform:isoamyl alcohol (24:1) mixture extraction. Nucleic acids precipitation was achieved by adding 2x volumes of the cold (-20 °C) ethanol. Precipitated high molecular weight nucleic acids were recovered by spooling with a glass rod, followed by rinsing in 70 % (v/v) ethanol. After evaporation of 70% ethanol the nucleic acids were resuspended in 5 ml TE buffer.

2.8.2 Small scale plasmid DNA preparation

Escherichia coli (*E. coli*) transformants harboring plasmid DNA were inoculated into 5 ml LB/ampicillin (100 μg ml^{-1}) broth and grown for 12 h at 37 °C with shaking at 200 rpm. Cultures (1 ml) were harvested by centrifugation (16 000 x g, 1 min, at room temperature), and the supernatants discarded. The cell pellets were resuspended in 100 μl alkaline solution I, briefly vortexed and by incubated for 5 min at room temperature. The alkaline solution II (100 μl) was then added, gently mixed and the suspensions were

incubated on ice for 2 min, followed by addition of the alkaline solution III (100 µl). The suspensions were gently mixed until a precipitate was observed and then centrifuged (16000 x g, 10 min, at room temperature). Supernatants were collected and the nucleic acids precipitated by adding 1/10 volume of 3 M potassium acetate (pH 5.2) and 2x volume of cold ethanol (-20 °C). The precipitated nucleic acids were collected by centrifugation (16000 x g, 20 min, 4 °C), and washed with 70 % (v/v) ethanol. Ethanol was removed by brief centrifugation (16000 x g, 1 min, 4 °C) before the pellets were air dried and resuspended in TE buffer (pH 8.0).

2.8.3 RNAse digestion

To remove RNA co-purified during DNA isolation, 100 µl of 20 mg ml^{-1} ribonuclease A (RNAse) was added to 5 ml nucleic acid solution followed by incubation for 2 h at 37 °C water bath. The impurities were removed by two extractions with phenol: chloroform: isoamyl alcohol (25:24:1), followed by ethanol precipitation as described in Section 2.8.5. The recovered DNA was dissolved in 1 ml TE buffer (pH 8.0).

2.8.4. Determination of DNA purity and concentration

DNA purity and concentrations were determined by measuring absorbance at 280 nm, 260 nm and 230 nm, according to the method of Sambrook and Russell (2001) where one absorbance unit at 260 nm is equal to 50 µg dsDNA/ml. A DNA solution was considered pure if the A_{260nm} to A_{230nm} ratio was between 1.8 and 2.3 and A_{260nm} to A_{280nm} ratio was between 1.5 to 2.0 (Marmur, 1963).

2.8.5 Precipitation of the DNA

Unless stated otherwise, routine DNA precipitation was carried out by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and two volumes of (-20 °C) ethanol followed by incubation at -20 °C for 20 min before centrifugation (16 000 x g; 20 min 4 °C). The precipitated DNA was washed with (-20 °C) 70% (v/v) ethanol, followed by 70% ethanol evaporation before resuspending in an appropriate volume of TE buffer (pH 8.0).

2.9 Library construction

2.9.1 Restriction endonuclease digestion of DNA

Cleavage of the DNA with restriction endonucleases was performed at the manufacturers recommended temperatures according to the method of Sambrook and Russell (2001). Restriction reactions were stopped and prepared for agarose gel electrophoresis by adding 1/10 volume of 6x DNA agarose loading dye. Alternatively, if the restricted DNA was to be further used directly for enzymatic reactions, the digestion was either stopped by heating (65 °C, 10 min) or by gel purification.

2.9.2 Partial digestion of genomic DNA

Partial endonuclease restriction digestion of genomic DNA was carried out using the four base cutter *Sau*3A1. Genomic DNA (1 µg) in 20 µl of 1X *Sau*3A1 buffer was incubated with different amounts of *Sau*3A1 (1.0 - 0.00964U) at 37 °C for 1 h. The reaction was stopped by addition of $1/10^{th}$ volume of 6x agarose loading dye. Resultant fragments

were fractionated on 0.8% agarose gel electrophoresis for 12 h at 20 V and 2-8 kb fragments were recovered from the gel slices and purified.

2.9.3 Dephosphorylation of DNA

Removal of phosphate residues from the 5′-termini of the restricted plasmid DNA prior to cloning reactions was performed according of Sambrook and Russell (2001), using Shrimp Alkaline Phosphatase (SAP). Plasmid DNA (containing 1-2 pmoles of 5′ ends) in 50 µl of 1X SAP buffer was incubated with 2 U SAP at 37 °C for 2 h. SAP was heat-inactivated by incubation at 65 °C for 20 min. DNA was purified before being used for ligation.

2.9.4 Ligation of DNA

Ligation reactions were carried out essentially as described by Sambrook and Russell (2001) using T4 DNA ligase. DNA up to 1 µg (quantified by agarose gel electrophoresis), at an Insert (I): Vector (V) molar ratio of approximately 3:1 in 20 µl of 1X ligation buffer, was incubated overnight at 16 °C with 1 U T4 DNA ligase. Ligation reactions were either used directly to transform chemically competent *E. coli* cells or were first purified before transforming *E. coli* electrocompetent cells.

2.10 Polymerase chain reaction (PCR)

In order to amplify target DNA, 0.2 ml thin walled tubes were used in a GeneAmp PCR system 2700 (Applied Biosystem) or Eppendorf Mastercycler gradient thermocycler

equipped with a heated lid. Unless otherwise stated, the standard 50 µl-PCR reaction contained the following reagents: 1x PCR buffer, DNA template (20 ng plasmid or 50 ng chromosomal DNA), 0.5 µM of the upstream and downstream primers, 200 µM dNTPs mixture (dATP, dCTP, dGTP and dTTP), 1.25 U of *Taq* DNA polymerase. Reactions were made up to 50 µl with sterile ddH_2O. PCR additives such as 1.0 % (v/v) DMSO and 200 mM betaine were added as required and the final reaction volumes were adjusted appropriately. *Pfu* DNA polymerase was used instead of *Taq* DNA polymerase to amplify PCR products with increased fidelity, when necessary.

DNA templates for colony PCR were prepared by resuspending colonies (2-3mm in diameter) in 50 µl TE buffer (pH 8.0). The cells were disrupted by boiling (100 °C, 5 min) followed by centrifugation (16000 x g, 5 min at room temperature). Supernatants (10 µl) were used directly as a source of the template DNA in a standard 50 µl PCR reaction.

Thermocycling conditions: 94 °C for 4 min (20-30) cycles: 94 °C for 30 sec, X °C, for 45 sec, 72 °C for 2 min and 1 cycle: 72 °C for 10 min (X denotes relevant annealing temperature which was chosen 5 °C below the assumed primer melting temperatures calculated using the following formula (Tm= [no. of GC] x 4 + [no. of AT] x 2 °C).

2.11 TA cloning

The TA insert cloning of the PCR product was carried using InST/Aclone ™ PCR product cloning kit (Fermentas Life Sciences). The pTZ57TR/T plasmid was supplied as

*Eco*321 linearised vector treated with terminal deoxynucleotidyl transferase to create 3'-ddT overhangs at both ends, to facilitate the cloning of *Taq* polymerase amplified PCR products. The ligation reactions were carried out following manufacturer's instructions.

2.12 Blunt-end cloning

PCR products amplified with *Pfu* DNA polymerase were blunt-end cloned in pMOSB*lue* vector using pMOS*Blue* Blunt Ended Cloning Kit (Amersham Biosciences). The pMOSB*lue* vector was supplied as a blunt-end linearised vector cut with *Eco*RV. The PCR products are converted into blunt-ended phosphorylated products in a reaction containing 0.5 mM DTT, 1x pk buffer and 1 µl pk enzyme mix reaction. The resulting blund-ended PCR products were subsequently ligated into the pMOSB*lue* vector before used to transformation chemically competent cells.

2.13 Protein expression

Unless otherwise stated *E. coli* BL21 (DE3) was used as an expression host. Expression constructs were used to transform electrocompetent *E. coli* BL21 cells followed by plating overnight in LB agar plates containing either ampicillin or carbenicillin (100 µg ml^{-1}). Fresh colonies were used to inoculate 5 ml of LB broth supplemented with ampicillin or carbenicillin (100 µg ml^{-1}) and grown until OD_{600nm}=1.0. Cultures (1 ml) were harvested by centrifugation (16000 x g, 1 min at 25 °C) and the cell pellets washed twice (to remove secreted β-lactamase) with 1 ml of LB broth containing 100 µg ml^{-1} antibiotics, followed by resuspension in the same volume of LB broth. The resuspended

pellets (1 ml) were used to inoculate 500 ml of LB containing carbenicillin (100 μg ml^{-1}). Expression was performed following either of the two procedures:

Procedure 1: Cultures were grown at 37 °C with vigorous shaking (250 rpm) until the OD_{600nm} was between 0.5-0.6. Expression of the recombinant protein was induced by IPTG at stated concentrations and under incubation conditions as specified in the relevant results sections. Cultures were further incubated for 3-4 h after induction.

Procedure 2: After washing and resuspension of the overnight cultures, expression was carried out at the point of inoculation. Cultures (1ml) were used to inoculate 500 ml containing appropriate antibiotic (100 μg ml^{-1}). At this point, expression of the recombinant protein was induced with 0.1 mM IPTG followed by shaking incubation (150 rpm) at 18 °C for 12 h.

2. 13. 1 Preparation of the cell free extracts

Cell pellets harvested by centrifugation (3500 x g, 20 min, 4 °C) from the cultures expressing the recombinant protein were washed twice in phosphate buffer saline and resuspended in 20 ml Tris-HCl (50 mM, pH 8.0). Resuspended cells were disrupted by sonication using Sonoplus HD-070 (Bandelin, Germany) set at: 5x cycle burst for 5 min, 50% max. amplitude, 30 sec interval, in an ice bath. The cells debris was removed by centrifugation (16 000 x g, 30 minutes at 4 °C). Supernatants were used as a source of intracellular soluble proteins.

2.14 Protein Purification

All chromatographic experiments were performed at room temperature on FPLC system (P-5002 peristaltic pumps, GM-1 gradient mixer, UV-MII monitor, Frac-100 fraction collector, LCC-501 plus controller) operated with the FPLC director Unicorn 4.1 software (Amersham Biosciences, Uppsala, Sweden). The conductivity of sodium chloride and ammonium sulphate during the course of purification was determined by measuring the refractive index using the integrated Abbe refractometer. Prior to use, all buffers were filtered through nitrocellulose acetate membrane filters (pore size 0.2 μm, Micron Separation Inc) and degassed. When appropriate, protein samples were concentrated by membrane filtration (Centriprep YM-10, cut-off, 10 kDa, Millipore) before dialysis. Buffer exchanges between chromatographic steps were achieved by dialysis using dialysis tubing (cut-off 10 kDa, Merck).

2. 14. 1 Ammonium sulphate fractionation

Solid ammonium sulphate was added with rapid stirring to the soluble cell free extract to achieve 20% saturation followed by incubation overnight on ice at 4 °C. Precipitated proteins were removed by centrifugation (20 000 x g, 30 min, 4 °C) and supernatants used for hydrophobic interaction chromatography.

2.14.2 Hydrophobic interaction (HIC)

Supernatants from the ammonium sulphate fractionation step were loaded onto a HiLoad^{TM} 16/10 Phenyl Sepharose XK16 hydrophobic interaction column (Amersham Biosciences) pre-equilibrated with Tris-HCl buffer (20 mM, pH 7.5) containing 0.6 M ammonium sulphate (Buffer A). Unbound proteins were eluted with 5 column volumes of buffer A. Proteins bound to the hydrophobic matrix were eluted over 5 column volumes at a flow rate of 2 ml min^{-1} with a linear gradient of decreasing ammonium sulphate concentration (0.6-0 M) generated with Tris-HCl buffer (20 mM, pH 7.5) (Buffer B). Fractions (2 ml) were collected and fractions exhibiting activity were pooled, concentrated and dialysed against Tris-HCl buffer (20 mM, pH 7.5).

2.14.3 Ion exchange chromatography

The enzyme solution obtained from HIC was loaded onto a HiLoad^{TM} 26/10 Q-Sepharose XK26 ion exchange column which was pre-equilibrated with Tris-HCl buffer (20 mM, pH 7.5) (Buffer B). The unbound proteins were washed with 5 column volume of buffer B. Proteins bound to the ion-exchange resin were eluted with a linear gradient of increasing sodium chloride concentration (0-0.5 M) generated with Tris-HCl buffer (20 mM, pH 7.5) (Buffer C) at a flow rate of 2 ml min^{-1}. Fractions (2 ml) were collected and fractions exhibiting activity were pooled, concentrated and dialyzed against Tris-HCl buffer (20 mM, pH 7.5).

2.14.4 Gel exclusion chromatography

Concentrated active fractions eluted from the ion-exchange column were loaded onto a Superdex 75 GL 10/300 size exclusion column pre-equilibrated and eluted with Tris-HCl buffer (20 mM, pH 7.5) containing 150 mM NaCl at a flow rate of 0.5 ml min^{-1}. Fractions (0.5 ml) were collected and fractions exhibiting activity were pooled, concentrated and dialyzed against Tris-HCl buffer (20 mM, pH 7.5).

2.14.5 Preparation of inclusion bodies

Cultures expressing the recombinant protein were harvested and disrupted by sonication as previously described (Section 2.13.1). The cells debris was recovered as a cell free extracts then washed twice with Tris-HCL buffer (50 mM, pH 8.0) containing 1% Triton x100, with centrifugation (15 000 x g, 20 min, 4 °C) and the supernant discarded.

2.14.6 Denaturing, refolding and purification of inclusion bodies

The insoluble inclusion bodies obtained from Section 2.14.5 were resuspended in refolding buffer 1 at a density of 0.2 g ml^{-1} followed by incubation at 25 °C for 1 h while stirring. The clear supernatants (5 ml) obtained after centrifugation (20 000 x g; 20 min, 20 °C) were loaded onto a (1.3 cm x 3 cm) His select ™ Nickel gel column from Sigma Chemical Co. (Germany). The unbound proteins were eluted with 5 column volumes of refolding buffer 2 at a flow rate of 0.5 ml min^{-1}. Bound proteins were refolded over 5 column volumes at a flow rate of 0.5 ml min^{-1} with a linear gradient of a decreasing urea concentration (8-0 M) generated with gradient buffer 1. The refolded bound proteins

were eluted with a linear gradient of increasing imidazole concentration (0-250 mM) generated with imidazole elution buffer at a flow rate of 1 ml min^{-1}. The imidazole was removed by dialysis and the protein fraction was concentrated by membrane filtration (Centriprep YM-10, cutt-off, 10 kDa, Millipore) before analyzed on the SDS-PAGE. When required, protease inhibitor cocktail set III (AEBSF, Aprotinin, Bestatin, E-64, Leupeptin and Pepstanin A) were added at specified concentrations as described under relevant results sections.

2.15 Biochemical characterization methods

2.15.1 pH stability

The following buffers were used to assess pH stability of the purified esterase: Sodium phosphate (50 mM, pH 6.0-7.0), Tris-HCl (50 mM, pH 7.5-8.5), Carbonate/Bicarbonate buffer (50 mM, pH 9.-10). To determine the pH stability of the purified esterase, the enzyme was incubated in buffers at specified pH values (6.0-9.0) for 30 min at 25 °C. The reaction was initiated by addition of 1 mM ρ–nitrophenyl acetate substrate solution. The reaction rates were measured by monitoring the release of ρ-nitrophenol at 405 nm in 1cm path length cuvette. Because the molar extinction coefficient of ρ-nitrophenol is pH-dependent, calibration curves were prepared over the pH range studied. The molar coefficients at the respective pH were as follows (pH 6=4100 $M^{-1}cm^{-1}$; pH 6.5=8500 $M^{-1}cm^{-1}$; pH 7.0=12600 $M^{-1}cm^{-1}$; pH 7.5=14800 $M^{-1}cm^{-1}$; pH 8.0=17000 $M^{-1}cm^{-1}$; pH 8.5=18400 $M^{-1}cm^{-1}$; pH 9.0=18900 $M^{-1}cm^{-1}$).

2.15.2 Thermostability

Thermostability of the purified esterase was determined over the temperature range of 10-70 °C. Enzyme samples were incubated for 30 min at various temperature ranges in Tris-HCl (50 mM. pH 7.5) and the residual activity determined using standard assay.

2.15.3 Substrate specificity

To determine the effect of substrate chain length on esterase activity, Enzyme activity determined using the standard assay in a 1 ml reaction volume containing Tris-HCl (50 mM. pH 7.5) and 1 mM of specified ρ- nitrophenyl esters of various chain lengths: ρ-nitrophenyl acetate (C2), ρ-nitrophenyl propionate (C3), ρ-nitrophenyl butyrate (C4), ρ-nitrophenyl caproate (C6), ρ-nitrophenyl caprylate (C8), ρ-nitrophenyl decanoate (C10), ρ-nitrophenyl myristate (C14), and ρ-nitrophenyl palmitate (C16) (Substrate stock solutions were prepared fresh in methanol. Esterase activity was also determined against β-naphthyl esters of various chain lengths (C2-C12) under assay conditions described in Section 2.6.2.2.

Enzyme kinetics using ρ-nitrophenyl esters were determined using ρ-nitrophenyl substrates (C2-C4) at various concentration ranges 0.1 to 10 mM. When necessary, 0.04% (v/v) Triton^{TM} x100 was added in the reaction mixture to increase solubility of the substrates.

2.15. 4 Effect of various compounds on esterase activity

Enzyme samples were incubated at 25 °C for 30 min in Tris-HCl (50 mM pH 7.5) containing one of the following additives: detergents (1% v/v final concentration), inhibitors (0.01-1 mM final concentration), metals and transition metals (1 mM final concentration). The residual activity was measured using the standard assay. The activity of the enzyme under identical conditions but with no additives was taken as 100%.

2.16 DNA sequencing

The nucleotide sequences were determined with an automated DNA sequencer model 373A and a dideoxy chain termination procedure with fluorescein-labeled primers (Perkin Elmer Applied Biosystems) available at the University of Cape Town (South Africa). All PCR products and cloned DNA fragments for sequencing were purified from solution or agarose gels using the GFX plasmid DNA and Gel Band Purification kit (Amersham Biosciences).

2.17 Homology Modelling

The threading program FUGUE (Shi *et al.*, 2001) and GenTHREADER (Jone, 1999) were used for the identification of structural homologues and structure predictions. Alignments outputs from FUGUE and GenTHREADER were subsequently used for model building using the external program MODELLER (Sali, 1995). Structural modelling of *B. multivorans* LipA and EstBL were based on the structure of *Chromobacterium viscosum*; PDB code: 1CVL (Lang *et al.*, 1996) and *B. gladioli* PDB

code: 1CI8 (Wagner *et al.*, 2002), respectively. These were the most suitable templates based on the high degree of sequence conservation; 78 % and 52% identity to 1CVL and 1CI8, respectively. Stereochemical analysis of the structures was performed using RAMPAGE (Lovell *et al.*, 2001). All models were viewed and presented as figures using PyMOL (Delano, 2002).

2.18 Computational techniques

Nucleotide and amino acid sequences were obtained from the GenBank, DDBJ, EMBL, SWISSPROT and Protein data bank (PDB) databases. Homology searches were performed on publicly accessible sequence databases by using the basic local alignment search tool (BLAST) (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al.*, 1997). Lipase and esterase homology searches were performed on the following databases: BRENDA (http://www.brenda.uni-koeln.de/), ESTHER (http://bioweb.ensam.inra.fr/ESTHER) and Lipase Engineering Database (http://www.led.uni-stuttgart.de/).

The sequences generated in this study were analyzed with Gene construction kit 2 (Textco BioSoftware, Inc.), Bioedit (Hall, 1990) and DNAMAN software programs. Multiple sequence alignments were performed with CLUSTALX (Thompson *et al.*, 1997). A neighbour joining tree was constructed from Kimura distances analysis using Treecon 1.3 (Van de Peer and De Wachter, 1997) and Phylowin programs. Leader sequences were predicted using the signalP3 software available from (http://www.cbs.dtu.dk/services/SignalP/). Protein families, domains and functional sites were deduced using Prosite (http://www.ebi.ac.uk/interpro/). Topology and secondary

structure predictions were performed with PROF, PROFsec and PHD respectively (Rost, 1996) available on (http://cubic.bioc.columbia.edu). Theoretical pI and MW values were predicted using the EXPASY web site (Http://www.expasy.ch) (Appel *et al.*, 1994).

CHAPTER THREE

Isolation, screening, and identification of ethyl ferulate hydrolyzing bacteria

3.1 Introduction

A number of approaches have been developed to access novel enzyme genes from the environment. These include culture-enrichments (Wahler and Reymond, 2001), culture independent (also known as Metagenomic approaches) (Rondon *et al.*, 2000), PCR-based sequence-independent (Bell *et al.*, 2002) and genome sequencing (Cowan, 2000; Kim *et al.*, 2004b; Ro *et al.*, 2004) approaches. The advantages and disadvantages of these technologies have been reviewed elsewhere (Lorenz *et al.*, 2002; Cowan *et al.*, 2005).

A major fraction of the commercially available enzymes have been obtained by classical culture enrichment approaches, involving the cultivation of microorganisms and the subsequent screening of the pure strains for the desired catalytic activity (Wahler and Reymond, 2001). This is typically achieved by simply including the substrate of interest in the growth medium as a sole carbon source.

In the present study, a classical enrichment approach was used for isolation of bacteria which encoded ethyl ferulate hydrolyzing (i.e. ferulic acid esterase) activity (Fig. 3.1).

The term ethyl ferulate hydrolyzing (EFH) activity will be used interchangeably throughout the text to refer to ferulic acid esterase (FAE) activity.

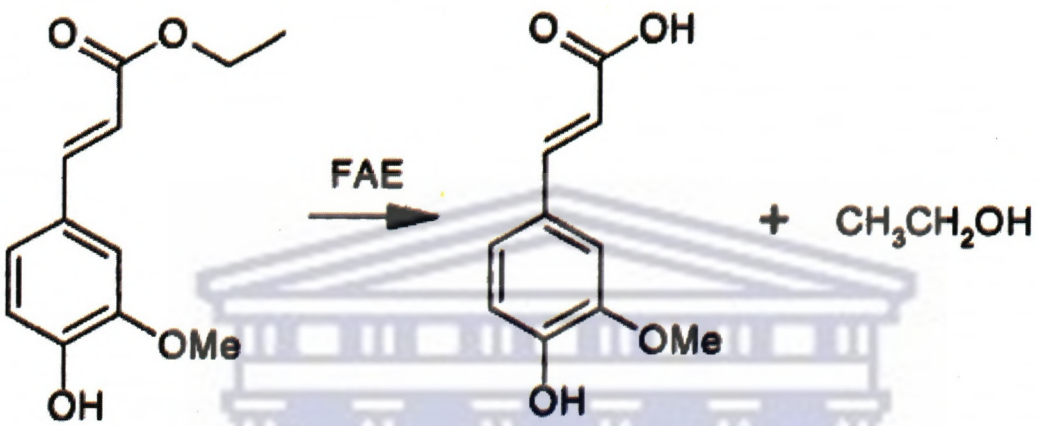


Figure 3.1: Reaction scheme showing the hydrolysis of ethyl ferulate to ferulic acid and ethanol by ferulic acid esterase (FAE). (Taken from Andersen *et al.*, 2002)

3.2 Isolation strategies

Silage samples (from a maize processing farm in Stellensbosch, South Africa) were used to isolate ethyl ferulate hydrolyzing bacteria. Primary screening was performed by spreading aqueous suspensions of these samples on M9-EF minimal nutrient medium plates containing different concentrations of ethyl ferulate (EF) as a sole carbon source. Samples were incubated aerobically at various temperatures (30-50 °C) for up to eight days. Colonies were observed only on media containing 0.1% (v/v) EF, after four days of incubation at 30 °C. Colonies were picked and restreaked onto the same minimal nutrient medium plates. The plating cycle was repeated until pure cultures were obtained, based on the light microscopic observations (data not shown). Rapid growth in the presence of

the EF as a sole carbon source was taken as a putative evidence for the presence of ethyl ferulate hydrolyzing (EFH) enzymes.

3.2 Screening for ethyl ferulate hydrolyzing enzymes

Thirteen pure bacterial isolates, selected for their ability to grow on EF as a sole carbon source, were further examined for their ability to produce EF-hydrolyzing enzyme(s). In order to confirm the presence of EFH enzyme(s) in these isolates, a qualitative agar screening assay (Donaghy *et al.* 1998) was used. The principle of the assay is that at high concentration, EF has very low solubility in aqueous solution resulting in a "cloudy haze". A zone of clearance is produced as the result of the enzymatic de-esterification of the ethyl ferulate substrate. The appearance of a zone of clearance around the culture inoculum indicates the presence of EF-hydrolyzing esterase(s).

All thirteen selected isolates were cultured in liquid M9-EF minimal nutrient medium (containing 0.1% v/v EF as a sole carbon source), and whole-cell cultures were used for the determination of EFH activity on a plate screening assay. All of the thirteen isolates showed a zone of clearance (Table 3.1). One isolate (designated UWC 10) consistently showed high EFH activity as measured by the zone of clearance (Fig. 3.2). It was noted that this isolate always showed a zone of clearance around the colony margins, an observation which was not noted with the other isolates. All thirteen isolates were further qualitatively assayed for non-specific esterase and lipase activities using tributyrin and olive oil plate assays, respectively. While non–specific esterase activity was observed in all isolates, some of the isolates tested negative for lipase activity (Table 3.1). Isolate

(UWC10) was chosen for further studies based on high EFH, non-specific esterase and lipase activities.

Table 3.1: A summary of qualitative data from plate assays of the thirteen bacterial isolates.

Isolates	Ethyl ferulate assay (EF hydrolyzing activity)	Tributyrin assay (Esterase activity)	Olive oil + Rhodamine B (Lipase activity)
UWC1	++	++	+++
UWC2	++	+++	_
UWC3	++	+++	+++
UWC4	++	++	+++
UWC5	++	++	+++
UWC6	++	+++	_
UWC7	++	+++	+++
UWC8	++	++	-
UWC9	++	+++	+++
UWC10	+++	+++	+++
UWC11	++	+++	-
UWC12	++	+++	_
UWC13	++	+++	_

(- =No zone of clearance; ++ = Zone of clearance < 5 mm; +++ >10 mm. For the lipase assay, the zone of clearance was observed by fluorescent under long range UV (350nm).

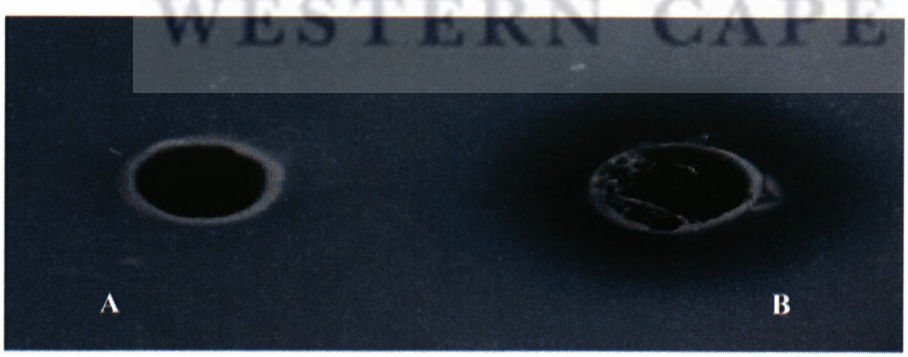


Figure 3.2: Ferulic acid esterase plate assay showing zones of clearance. (A) Negative control, UWC10 heat killed by boiling at 100 °C for 10 min) and (B) UWC10 culture.

3.3 Enzyme localization studies

In order to investigate the localization of EFH activity in UWC10, a culture was grown in a liquid minimal medium until mid-stationary growth phase (OD_{600nm}=1.2) before harvested. While no EFH activity was detected in the culture broth fraction, the cell lysates after sonication showed EFH activity (Fig. 3.3). The absence of EFH activity in the culture broth strongly suggested that EFH activity could be located either in the periplasm or the cytoplasm.

Early screening assays suggested that isolate UWC10 possessed both non-specific esterase and lipase activities. In order to differentiate and localize these activities, quantitative assays were performed using ρ-nitrophenyl acetate and ρ-nitrophenyl palmitate as a substrates for esterase and lipase, respectively (Table 3.2). Most of the lipase activity (95.3%) was found in the culture broth, while only minor (4.7%) lipolytic activity was detected in the intracellular fraction. In contrast, most of the esterolytic activity was found in the intracellular activity (76%), with only 24% detectable in the culture broth. Furthermore, no EFH activity could be detected in the culture broth, further confirming the screening assay observations which indicated that EFH activity of isolate UWC10 was intracellular.

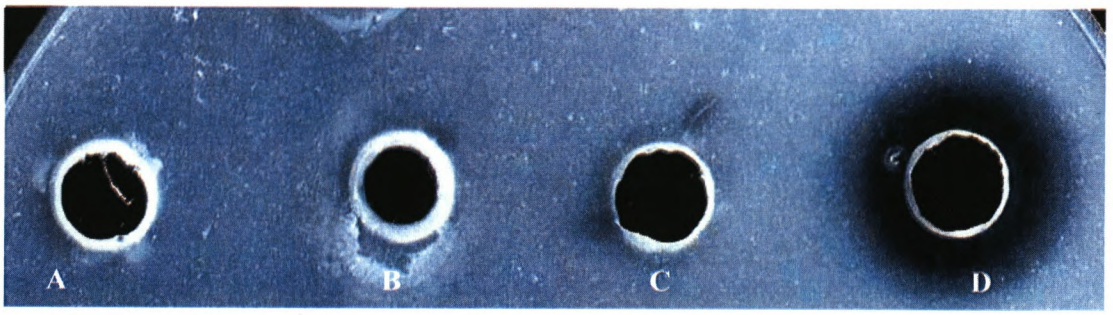


Figure 3.3: Ferulic acid esterase plate assay showing activity of isolate UWC10 culture fractions (A) (heat killed) extracellular fraction, (B) (heat killed) soluble fraction, (C) extracellular fraction, and (D) soluble fraction. Each well contained approximately 100 μg ml^{-1} protein.

Table 3.2: Localization of esterolytic and lipolytic activities. Data are mean values of triplicate assays.

Assay	Extracellular Specific Activity (U mg^{-1})	Intracellular Specific Activity (U mg^{-1})
EFH	0	0.75 ± 1.2
Non-specific esterase	11.2 ± 3.5	35.4 ± 8.3
Non-specific lipase	72.3 ± 2.2	3.6 ± 1.1

3.4 Effect of nutrient conditions on the esterolytic and lipolytic activities

The regulation of production EFH, esterolytic and lipolytic activities were investigated under different nutrient conditions (Table 3.3). No significant differences in EFH, esterase and lipase specific activities were observed in cultures grown in nutrient broth in the presence and absence of EF and FA as potential inducers. The same trend was noted when isolate UWC10 was cultured in minimal medium with glucose as a sole carbon

source. These findings gave an overall indication that under the conditions investigated the EFH, esterase, and lipase specific activities were probably not under the influence of catabolite repression and that both esterolytic and lipolytic activities were constitutively expressed.

Table 3.3: Effect of carbon source on the EFH, esterase and lipase specific activities

A 10 % (v/v) inoculum of the UWC10 grown in the nutrient broth (OD_{600nm} = 1.2) was used to inoculate all flasks. Cultures were grown in 100 ml volumes (see materials and methods) to mid-stationary phase before harvested for enzyme activity and protein determination. Data are mean values of triplicate assays.

Media	[b]EFH specific activity ($U\ mg^{-1}$)	[b]Esterase specific activity ($U\ mg^{-1}$)	[a]Lipase specific activity ($U\ mg^{-1}$)
Nutrient Broth	0.48 ± 0.4	12 ± 4.2	51 ± 3.5
Nutrient broth + EF (0.1%)	0.52 ± 0.1	13 ± 2.1	49 ± 5.1
Nutrient broth + FA (1%)	0.50 ± 0.1	11 ± 6.9	50.2 ± 2.0
M9 + glucose (0.4%)	0.50 ± 0.7	12 ± 1.2	52 ± 4. 9
M9 + EF (0.1%)	0.49 ± 0.3	12.9 ± 1.1	50 ± 3.7
M9 + EF (0.1%) + glucose (20 %)	0.49 ± 0.5	11.5 ± 1.0	50 ± 2.0
M9 + FA (1%)	0.50 ± 0.1	11 ± 1.3	53 ± 2.5
M9 + FA (1%) + glucose (20%)	0.49 ± 0.1	11.6 ± 1.4	53 ± 3.3

([a] = Extracellular fractions after concentrating; [b]=Intracellular fractions after sonication)

3.4 Growth and activity profiles

Growth and activity profiles were determined in nutrient broth cultures (Fig. 3.4). The decision to use nutrient broth was based on the fact that the EFH enzyme(s) was shown to

be constitutively expressed (Section 3.3). EFH activity was detectable from early exponential growth phase to mid-stationary phase (Fig. 3.4). EFH specific activity was observed to plateau at the transition from exponential to stationary-phase, while during the stationary growth phase the specific EFH activity showed a decreasing trend. Addition of EF and FA to nutrient broth did not enhance the EFH specific activity (data not shown), consistent with early observations which indicated that EFH activity of the UWC10 is not induced by these substrates.

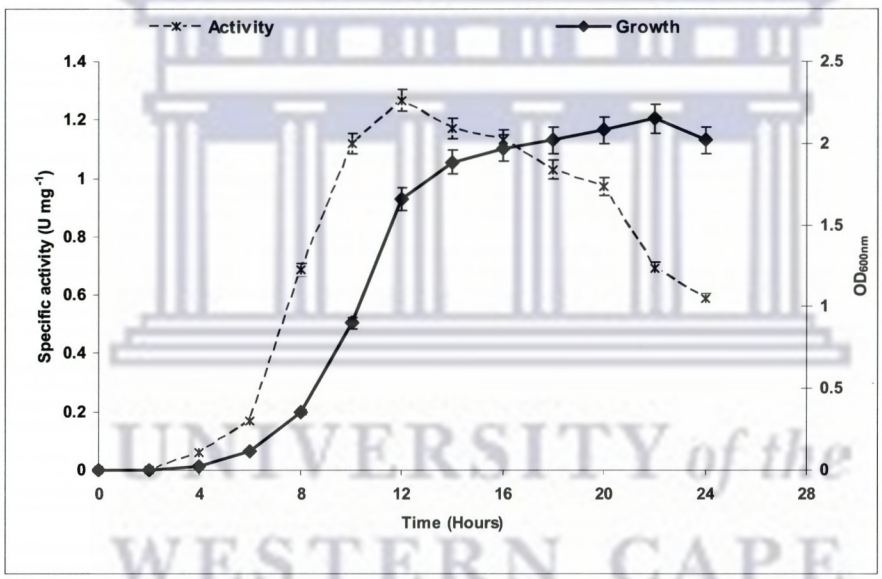


Figure 3.4: Growth and EFH specific activity profiles for the UWC10 isolate. Samples were removed aseptically at given time intervals for growth monitoring at OD_{600mn}, EFH activity assays and determination of protein content. The EFH and protein determination was performed on the cell free extracts of the sonicated fractions. Specific activity was expressed as U mg^{-1} and data are mean values of triplicate assays.

3.5 Esterolysis of EF in liquid culture

Ethyl ferulate deesterification was monitored during the growth of UWC10 in M9-EF minimal medium containing 0.1% (v/v) EF. The liquid medium was inoculated with 10 % (v/v) of a stationary phase culture pre-grown in nutrient medium. Ethyl ferulate hydrolyzing activity was confirmed by the production of ferulic acid in the culture broth. As expected, negligible FA was detected at time zero (Fig. 3.5). The amount of FA produced increased to 3.4% and 76.4% after 12 h and 24 h of growth, respectively. After 36 h and 48 h growth the level of FA declined to 48.4% and 25%, respectively. The decrease in FA concentration during growth implies that FA was being further metabolized. The control experiment without inoculum indicated that EF was very stable under these conditions.

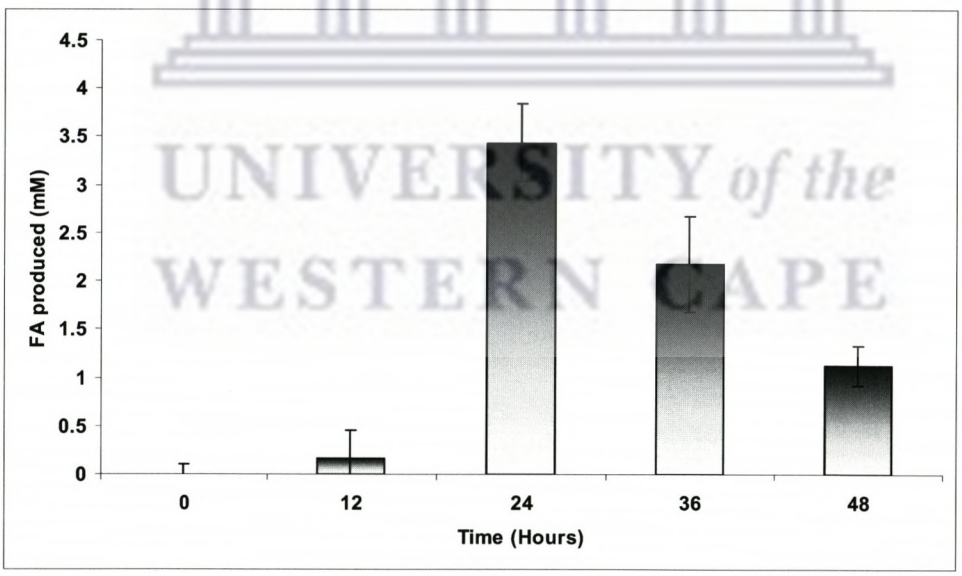


Figure 3.5: Ferulic acid produced in ethyl ferulate containing minimal medium. Data are mean values of triplicate assays. The initial EF concentration was 0.1% (w/v) (4.5 mM).

In order to further confirm these observations, samples of the culture supernatant were analyzed by Thin Layer Chromatography (Fig. 3.6). A single spot (R_f =0.85) was observed at time zero (Fig. 3.6, lane 2). After 12 h, three spots were observed: based on the R_f values, these were identified as EF (R_f =0.85), ferulic acid (R_f = 0.54) and ρ-coumaric acid (R_f = 0.41). After 24 h of growth the substrate was completely utilized as indicated by the absence of EF (Fig. 3.6, lanes 4 and 5). However, both ferulic acid and ρ-coumaric acid were still detectable. Although both ferulic acid and ρ-coumaric acid were still observed after 48 h of growth, an additional spot (R_f =0.76) was also observed. Based on the likely metabolic products of FA, this spot was tentatively assigned as one of the vanillin derived compounds: i.e., vanillin, vanillic acid or vinyl guaiacol (Fig. 3.6, Lane 6).

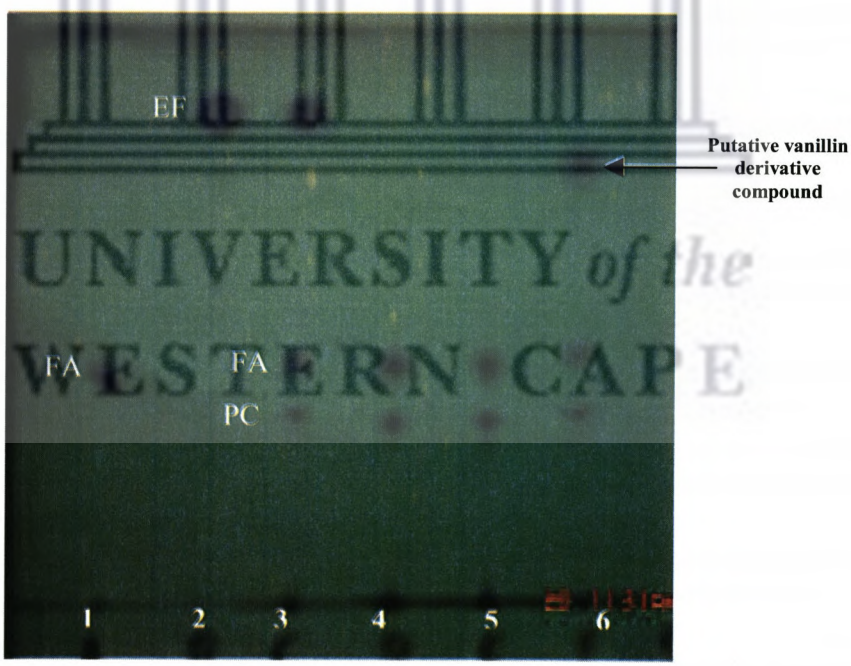


Figure 3.6: Thin layer chromatoplate showing ethyl ferulate (0.1% v/v) biotransformation products. Lane 1: ferulic acid (1μg μl^{-1}), Lane 2: 0 h, Lane 3:12 h, Lane 4: 24 h, Lane 5: 36 h, and Lane 6: 48 h. The chromatogram was developed in chloroform:methanol:formic acid (85:15:1:v/v/v) as a mobile phase and was visualized under UV light at 366 nm. FA = ferulic acid, EF=ethyl ferulate, PC=ρ-coumaric acid.

3.6 Biotransformation of EF by resting cells

Ethyl ferulate deesterification was also monitored using resting cells of isolate UWC10, previously cultured in nutrient broth (Fig. 3.7). Maximum FA titres (56.5%) were produced after a 6 h incubation while after 12 h only 3.7% of the theoretical yield was detectable. While both the accumulation and subsequent loss of FA were more rapid than on growing cultures, the trend was essentially identical to that observed in growing cultures (Section 3.5).

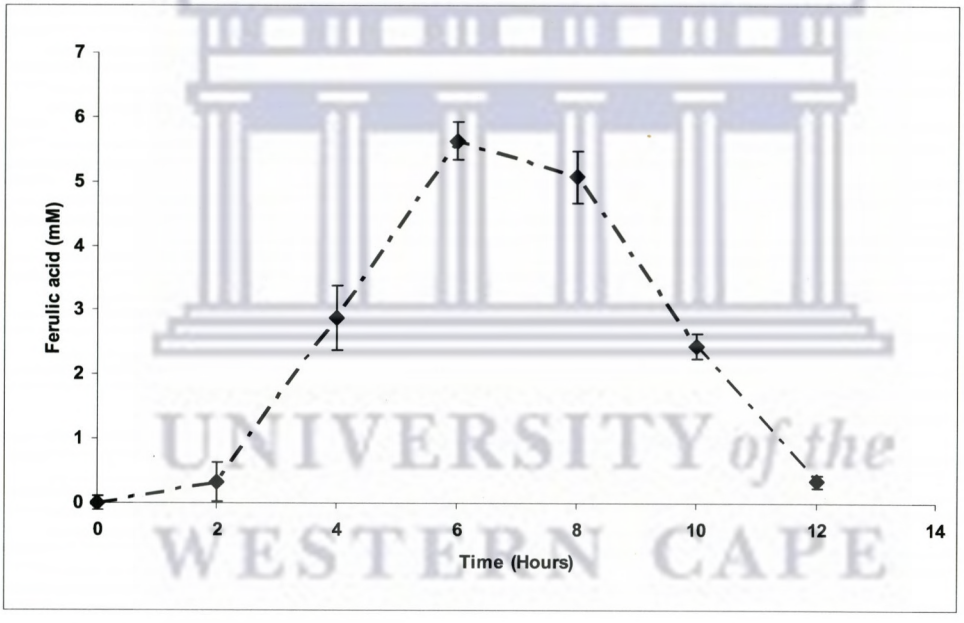


Figure 3.7: Time course showing ferulic acid production by isolate UWC10 resting cells. Data are mean values of triplicate assays. The initial EF concentration was 0.2% (w/v) (9.0 mM).

Ethyl ferulate biotransformation products were analyzed by TLC (Fig. 3.8). Complete utilization of the substrate was observed after 6 h (Fig. 3.8), corresponding to the maximum FA production as indicated by quantitative assays (Fig. 3.7). The appearance of new spot corresponding to a putative vanillin-related derivative was also observed after 12 h. The new spot had the same R_f value (0.76) as the one observed during the biotransformation studies in liquid medium.

In order to identity this spot, chromatoplates were developed and sprayed with ferric chloride and ethanolic-vanillin reagents since vanillin and its derivatives respond differently to these spray reagents. When the chromatoplate was sprayed with ferric chloride the spot gave no reaction, while with ethanolic-vanillin spray the spot could be visualized as a violet color (data not shown). These observations suggested that the spot could be vinyl guaiacol since vanillin does not react with the ethanolic-vanillin spray. Furthermore, when vanillin was included as a standard, it exhibited an R_f value of 0.68 (data not shown). These findings led to tentative conclusion that the additional spot might be vinyl guaiacol.

The mechanism of bioconversion of FA to vinyl guaiacol is known to be via non-oxidative decarboxylation of the FA side chain, a reaction mediated by decarboxylase enzyme(s) (Crawford and Olson, 1978; Nazareth and Mavinkurve, 1986; Rosazza *et al.*, 1995). In order to provide evidence of such a mechanism, tests were performed for ferulic acid decarboxylase activity. Ferulic acid decarboxylase activity is determined by the disappearance of a peak at 285 nm and the coincident formation of a new peak at 258 nm

(Karmakar *et al.*, 2000). The ferulic acid decarboxylating activity of isolate UWC10 resting cells pre-grown in the nutrient broth was studied over a 6 h period (Fig. 3.9). Ferulic acid decarboxylase activity could be detected within the first hour of incubation. Based on these findings it was concluded that the ferulic acid produced during biotransformation is being further decaboxylated to vinyl guaiacol.

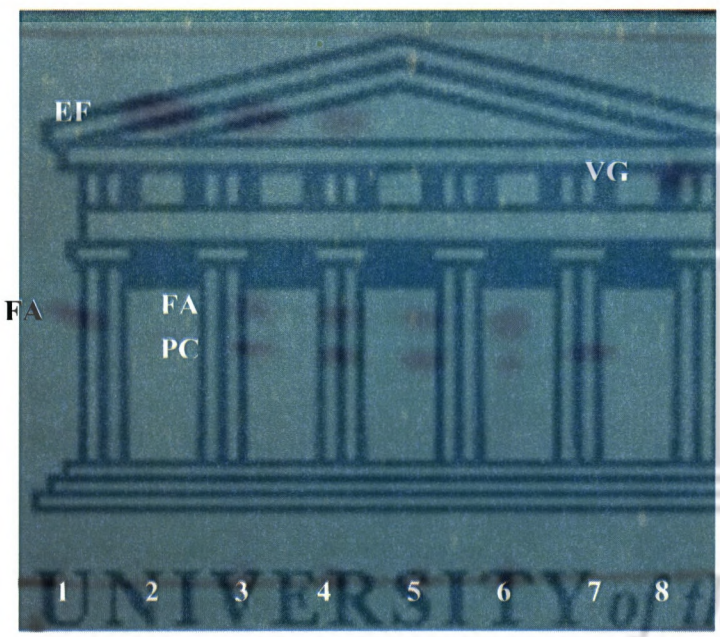


Figure 3.8: Thin layer chromatoplate showing ethyl ferulate 0.2 % (v/v) biotransformation products by UWC10 resting cells. Lane 1: ferulic acid (1µg μl^{-1}), Lane 2: 0h, Lane 3:2 h, Lane 4: 4 h, Lane 5: 6 h, and Lane 6: 8 h, lane 7: 10 h, Lane 8: 12 h. The chromatogram was developed in chloroform: methanol:formic acid (85:15:1 v/v/v) as a mobile phase and was visualized under UV light at λ = 366 nm. FA = ferulic acid (EF=ethyl ferulate, PC=ρ-coumaric acid, VG= vinyl guaiacol.

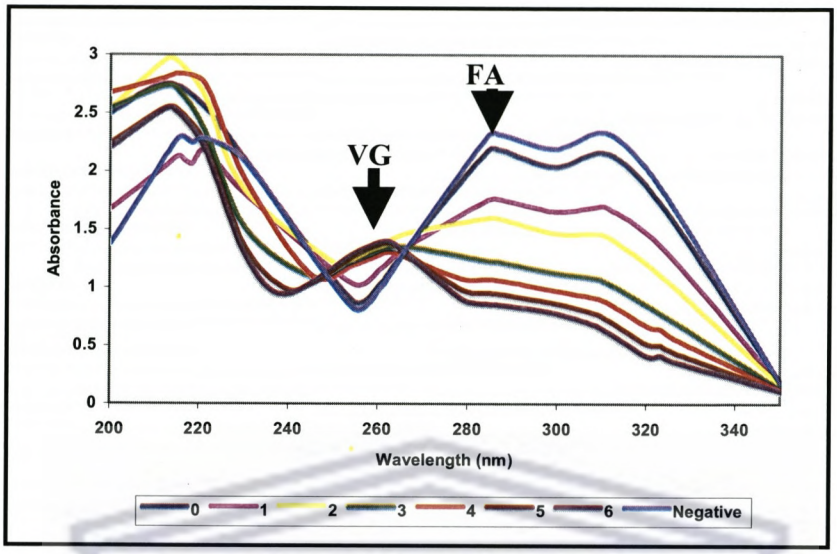


Figure 3.9: UV scan showing absorbance changes on incubation of ferulic acid with UWC10 resting cells. Data are average values of duplicate assays. FA concentration was 1 % (v/v) final.

Figure 3.10 shows the HPLC spectra of the EF biotransformation products with UWC10 resting cells. Control experiments showed that ethyl ferulate had a retention time of 14 min, while FA showed a retention time of around 27 min. At time zero, only an EF peak can be observed. After 4 h incubation, peaks corresponding to both EF and FA could be observed. However, no putative ρ-coumaric was observed, as indicated by TLC analysis. After 8 and 12h a distinct peak corresponding to vinyl guaiacol (retention time, 7 min) confirming the initial TLC observation that FA was being further metabolized to vinyl guaiacol. However, after 24 h of incubation no distinct peaks could be observed, indicating that vinyl guaiacol was being further degraded (data not shown).

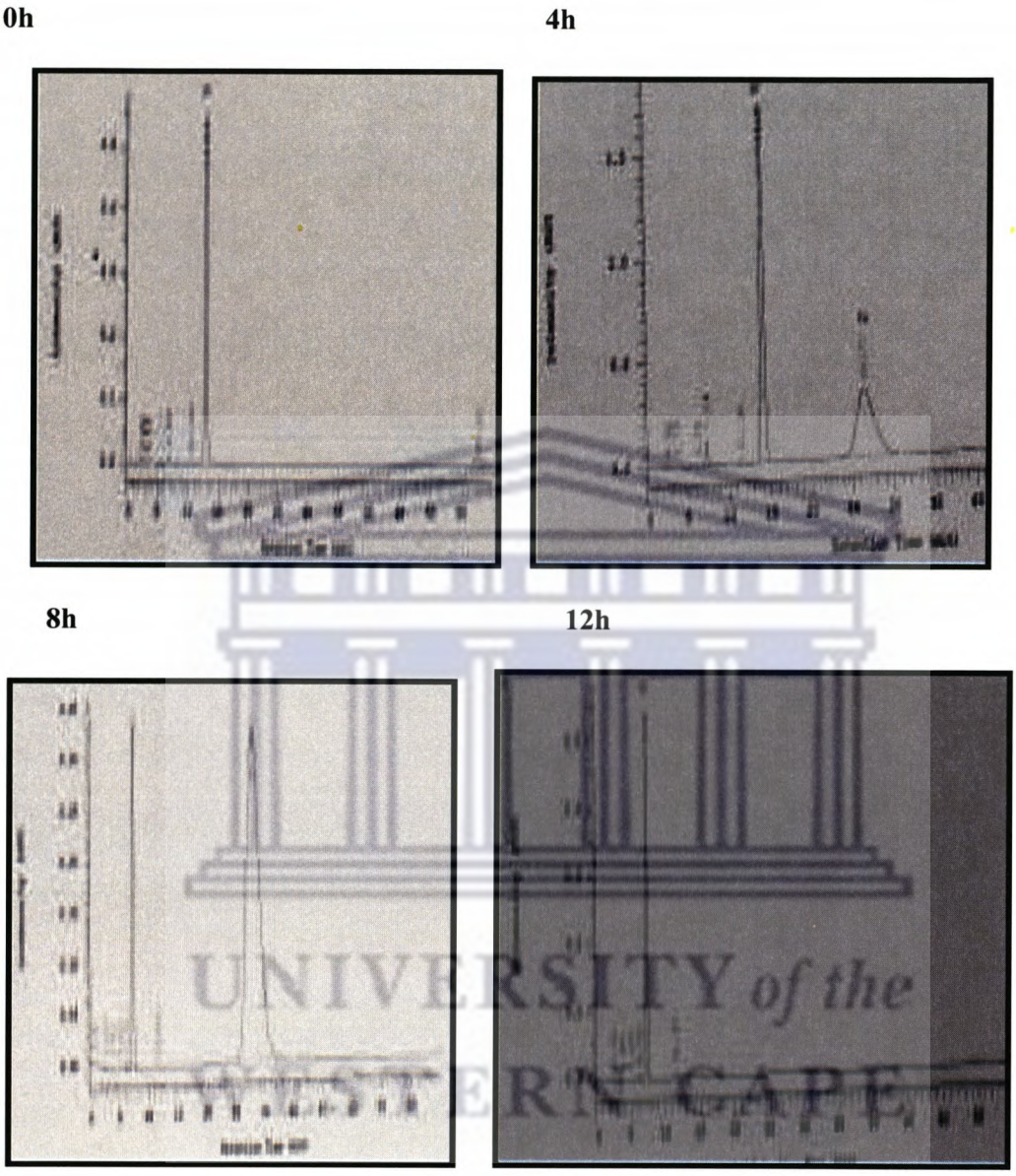


Figure 3.10: HPLC spectra showing EF biotransformation products generated by UWC10 resting cells.

3.7 Molecular Identification of isolate UWC10

3.7.1 16S rRNA gene amplification

In order to determine the identity of isolate UWC10, phylogenetic analysis was carried out using the 16S rRNA gene sequence as a molecular marker. Genomic DNA from strain UWC10 was used as a template for PCR amplification using the universal primer pair E9F and U1510R (Table 3.4). Electrophoresis of the PCR products showed the expected 1.5 kb band (Fig. 3.11). The PCR product was gel purified and ligated into a TA cloning vector, pTZ57R/T, then used to transform chemically competent *E. coli* (DH5α) cells.

Two representative recombinant clones from the 16S rRNA gene library were chosen (based upon α-complementation using blue/white selection) and screened for the correct insert by colony PCR using the vector-derived M13F and M13R primer pair (Table 3.4). Following colony PCR amplification and agarose gel electrophoresis, an insert of the expected size was obtained from both colonies (data not shown). Amplified Ribosomal DNA Restriction Analysis (ARDRA) of these PCR products revealed that the two clones harboured the same insert (Fig. 3.12). Plasmid DNA from one of the two colonies (designated pBurk16S) was isolated, purified and sequenced in both directions using the following primers: M13R, M31F, F2, R2, F4 and R4 (Table 3.4).

Table 3.4: Species-specific and nucleotide sequencing PCR primers used in this study.

Primer name	Target	Sequence (5′→3′)	References
E9F	7-26[a]	GAGTTTGATCCTGGCTCAG	Farrelly *et al.*, 1995
U150UR	1490-1512[a]	GGTTACCTTGTTACGACTT	Farrelly *et al.*, 1995
M13F	598-615[b]	GTAAAACGACGGCCAGT	Yanisch-Perron *et al.*, 1985
M13R	734-751[b]	CAGGAAACAGCTATGAC	Yanisch-Perron *et al.*, 1985
$F2^{Seq}$	338-357[a]	ACTCCTACGGGAGGCAGCAG	(Lane, 1991)
$R2^{Seq}$	806-785[a]	GGACTACCIGGGTATCTAATCC	Lane, 1991
$F4^{Seq}$	785-806[a]	GGATTAGATACCCIGGTAGTCC	Lane, 1991
$R4^{Seq}$	1242-1229[a]	CCATTGTAGIACGTGTGIAGCCC	Lane, 1991
G1	153-133[c]	GCCATGGATACTCCAAAAGGA	Whitby *et al.*, 1998
G2	939-958[c]	TCGGAATCCTGCTGAGAGGC	Whitby *et al.*, 1998
SPR3	969-985[d]	TCGAAAGAGAACCGGGCG	Whitby *et al.*, 2000
SPR4	939-985[e]	TCGAAAGAGAACCGATA	Whitby *et al.*, 2000
BCR1(F)	2-20[f]	TGACCGCCGAGAAGAGCAA	Mahenthiralingam *et al.*, 2000
BCR2(R)	1044-1024[f]	CTCTTCTTCGTCCATCGCCTC	Mahenthiralingam *et al.*, 2000

[a] = Position in relation to *E. coli* 16S rRNA gene (GenBank Accession number J10859)

[b] =Position in relation to pTZ57R/T (Fermentas life sciences)

[c] = Position in relation to *B. cepacia* 16S rRNA gene (GenBank Accession number X16368)

[d]= Position in relation to *B. multivorans* 16S rRNA gene (GenBank Accession number Y18703)

[e]= position in relation to *B. stabilis* 16S rRNA gene (GenBank Accession number DQ118268)

[f]= Position in relation to *B. multivorans rec*A gene (GenBank Accession number U70431)

^{Seq}= sequencing primers

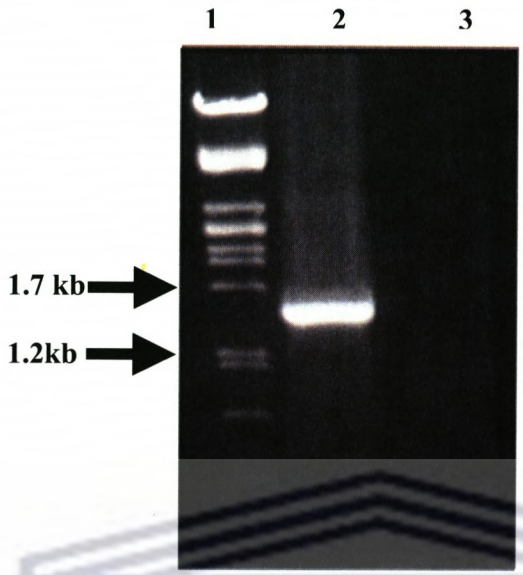


Figure 3.11: Agarose (1%) gel electrophoregram showing amplification of the 16S rRNA gene using the E9F and U1510R primer pair. Lane 1: *Pst*I-digested λDNA markers; Lane 2: 16S rRNA gene fragment; Lane 3: Negative control without DNA template.

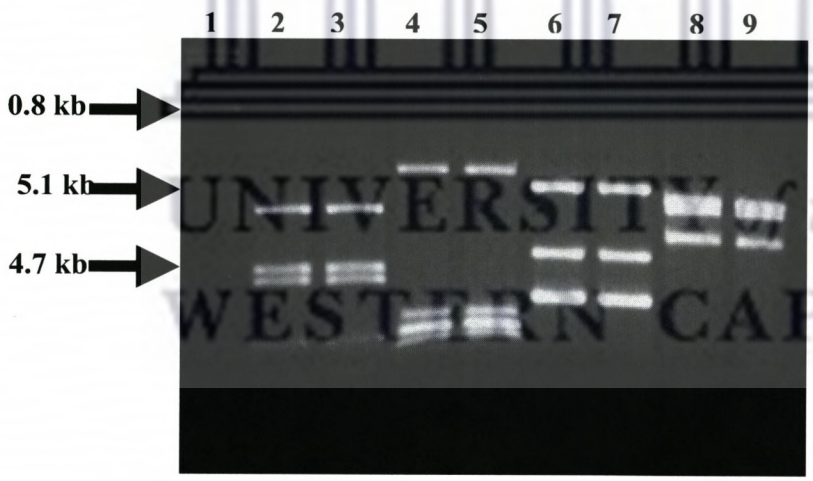


Figure 3.12: Agarose (2.5%) gel electrophoregram showing amplified ribosomal DNA restriction analysis (ARDRA) of the amplified M13 PCR product from two representative clones. Lane1: *Pst*I-digested λDNA marker; Lane 2 and 3: *Hae*III; Lane 4 and 5: *Alu*I; Lane 6 and 7: *Dde*I and Lane 8 and 9: *Rsa*I

3.7.2 Homology searches

A BLAST search with 1498 bp sequence revealed that the 16S rRNA gene of UWC10 had a very high degree of nucleotide sequence identity to the genus *Burkholderia* (Table 3.5). For further comparison of the UWC10 16S rRNA gene sequence with sequences from other *Burkholderia* species, a phylogenetic analysis was performed using Neighboring-Joining Distance analysis (Fig. 3.13). For the purpose of rooting the phylogenetic tree, the *Ralstonia eutropha* (AB007995) 16S rDNA sequence was included as an out-group. A phylogram analysis based on the 16S rRNA sequences of *Burkholderia* species strongly suggested that UWC10 strain is closely related to *B. multivorans*, a member of *B. cepacia* complex.

Table 3.5: BLAST (n) search data for the UWC10 16S rRNA gene.

Strains	**Accession number**	**Score (Bits)**	**Identity (%)**	**E_{value}**
Burkholderia multivorans	AF097531	2927	1489/1494 (99%)	0.00
Burkholderia multivorans	BMU18703	2919	1476/1478 (99%)	0.00
Burkholderia vietnamiensis	AF097534	2896	1485/1494 (99%)	0.00
Burkholderia multivorans	AB092606	2882	1466/1771 (99%)	0.00
Burkholderia cepacia	AB110089	2872	1479/1490 (99%)	0.00
Burkholderia cepacia	AY741359	2870	1478/1489 (99%)	0.00
Burkholderia cepacia	AY741345	2866	1464/1471 (99%)	0.00
Burkholderia sp.	AY769904	2866	1464/1471 (99%)	0.00
Burkholderia sp	AY973819	2866	1464/1471 (99%)	0.00
Burkholderia sp.	AF219125	2866	1464/1471 (99%)	0.00

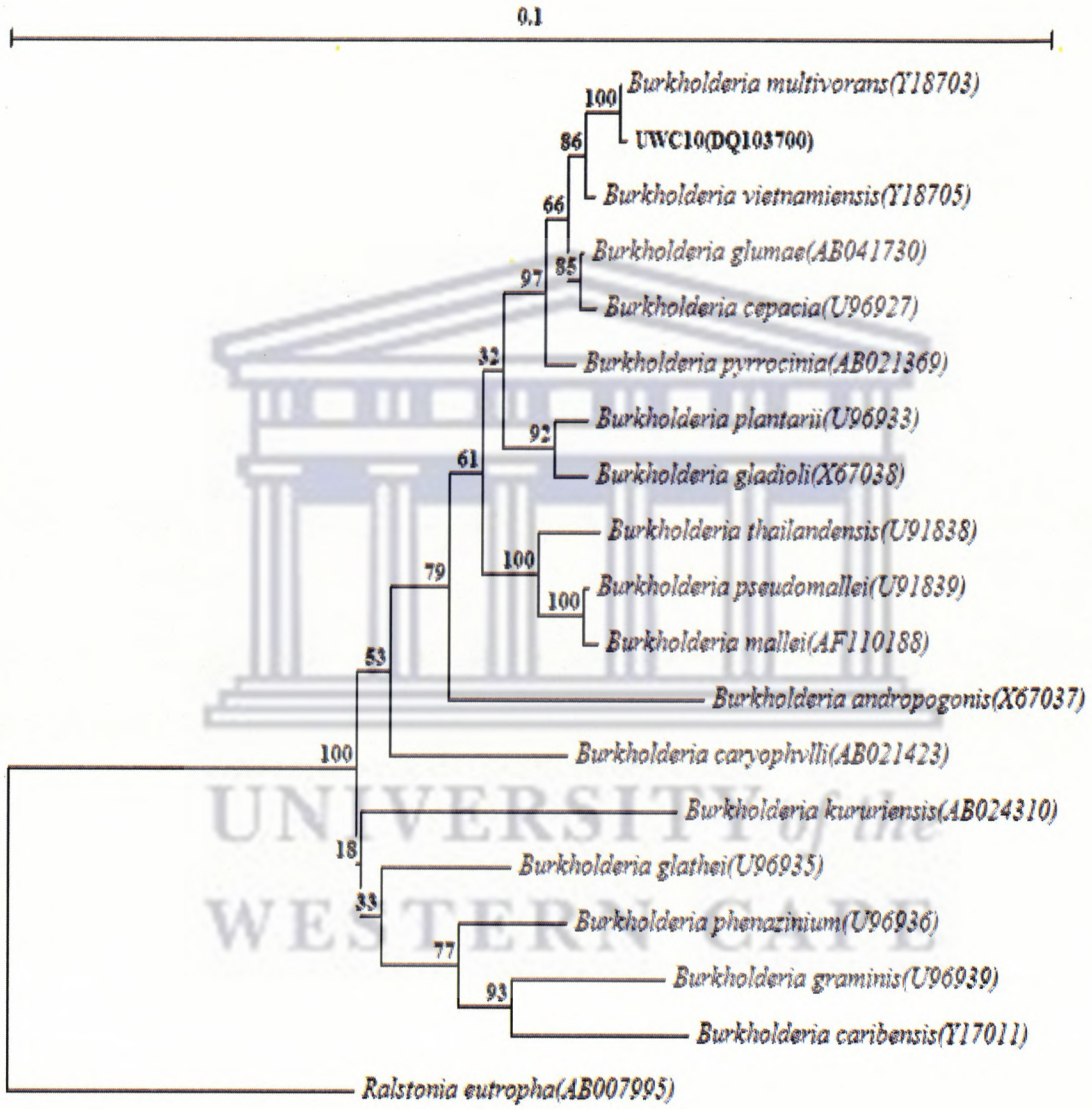


Figure 3.13: Phylogenetic tree showing the position of UWC10 strain. The scale bar represents 0.1 substitutions per nucleotide position.

3.7.3 Identification of UWC10 by species-specific primers

Since 16S rRNA gene analysis revealed very high levels of nucleotide sequence identity (99%) to a number of *Burkholderia* species, further experiments were required to establish the exact identity of the UWC10 strain. Species–specific primers, G1/G2 (Whitby *et al.*, 1998) and SPR3/SPR4 (Whitby *et al.*, 2000), based on the 16S and 23S rRNA gene sequences, have been developed to differentiate between the members *Burkholderia cepacia* complex. The primer pair G1/G2 targets a region of heterogeneity within the 16S-23S spacer region, while the SPR3/SPR4 pair targets a region of heterogeneity located within the 5′ terminus of the 16S rRNA gene.

Combination of primer pairs (G1/G2, G1/SPR3, and G1/SPR4) were evaluated for their ability to amplify UWC10 genomic DNA under previously published PCR conditions (Whitby *et al.*, 1998; LiPuma et al., 1999; Whitby *et al.*, 2000). The universal primer pair E9F and U1510R was included as a control to test PCR efficiency. All primer pair combinations were expected to give 1.3 kb PCR amplicons (Whitby *et al.*, 1998; Whitby *et al.*, 2000). However, PCR reactions under published conditions did not give PCR amplicons with any of the three primer pairs. After a series of PCR optimization steps involving variations of the annealing temperature (50-70 °C), UWC10 genomic DNA was amplified with the G1/SPR3 primer pair at an annealing temperature of 58 °C (Fig. 3.14). No amplification was achieved with the G1/G2 and G1/SPR4 primer pairs under the conditions investigated. Based on the PCR algorithm for identification of *Burkholderia* species (Fig. 3.15), the absence of a PCR product with the G1/G2 primer pair suggested that the UWC10 was neither *B. cepacia* nor *B. concepacia*, while the absence of a PCR

product with G1/SPR4 suggested that UWC10 was not *B. stabilis*. The absence of the PCR amplicon with G1/SPR3 indicated that the organism was not *B. vietnamiensis*, while amplification with G1/SPR3 supported the conclusion that strain UWC10 was *B. multivorans*.

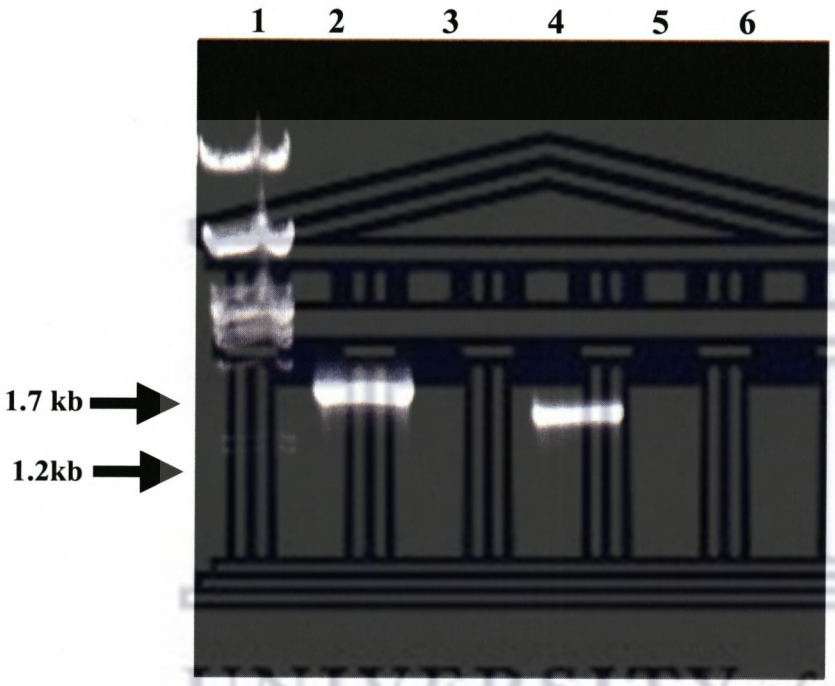


Figure 3.14: Agarose gel (1%) electrophoretogram showing PCR products using universal and species-specific primer pairs: Lane 1: *Pst*I-digested λDNA markers, Lane 2: E9F/U1510R (control), Lane 3: G1/G2, Lane 4: G1/SPR3, Lane 5:G1/SPR4 and Lane 6: G1/SPR3 with no genomic DNA.

3.7.4 Identification of the UWC10 based on *recA* gene analysis

PCR assays based on the *recA* gene have been recently developed for identification of members of the *Burkholderia cepacia* complex (Mahenthiralingam *et al.*, 2000; Payne *et al.*, 2005). Within the *Burkholderia cepacia* complex, species show low sequence

similarities between *recA* genes (94 to 95%) compared to the 16S rRNA genes (98 to 99%) (Ramette *et al.*, 2005).

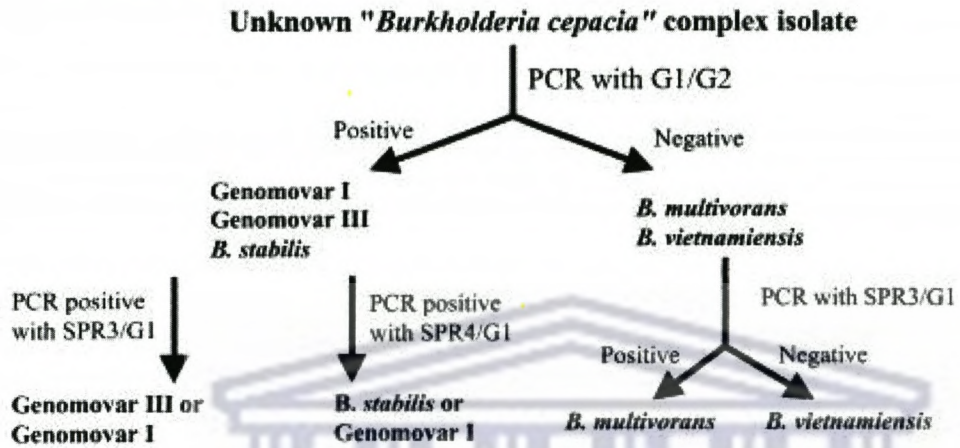


Figure 3.15: PCR algorithm to identify the species and genomovars of the *B. cepacia* complex (Whitby *et al.*, 2000).

The *recA* gene sequences are only available for some of the *Burkholderia* species, including *B. cepacia*, *B. concepacia*, *B. multivorans*, *B. stabilis*, *B. vietnamiensis*, and *B. ambifarai* (Coenye *et al.*, 2001). In addition to the *recA* gene-derived species-specific primers, Mahenthiralingam *et al.* (2000) have also developed a *recA* gene based RFLP approach, enabling the recognition of multiple types within each of the Bcc. In order to determine conclusively whether UWC10 belong to the species *B. multivorans*, a primer pair BCR1/BCR2 (Table 3.4), specifically targeting the *recA* gene of *B. multivorans* (Mahenthiralingam *et al.*, 2000), was evaluated. After a series of PCR optimization steps involving variations of the annealing temperature (50-70 °C), UWC10 genomic DNA was amplified with the BCR1/BCR2 primer pair at an annealing temperature of 58 °C (Fig. 3.16). The predicted PCR product of 1043 bp was observed. The ability of the

BCR1/BCR2 primer pair to produce a strong signal under these conditions strongly suggests that UWC10 is *B. multivorans*.

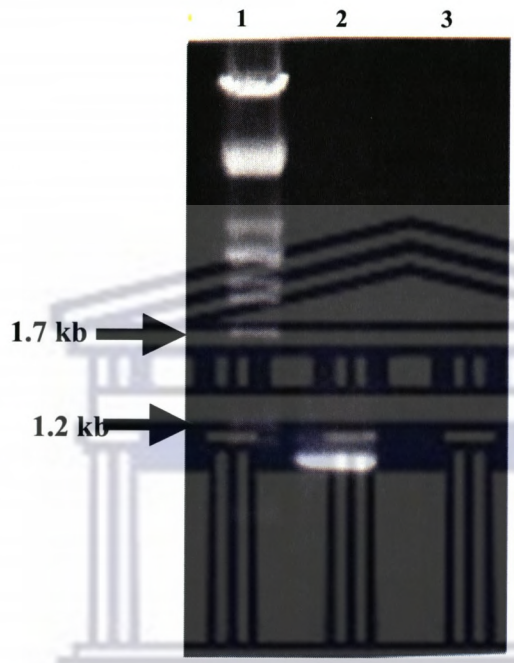


Figure 3.16: Agarose (1%) gel electrophoregram showing amplification of *recA* gene from strain UWC10. Lane 1 *Pst*I-digested λDNA markers, Lane 2: BCR1 and BCR2 primer pair, Lane 3: Negative control without the UWC10 genomic DNA template.

RFLP analyses of the *rec*A gene have been used previously to discriminate amongst the members of the Bcc. An RFLP analysis of the UWC10 *recA* gene with *Alu*I and *Hae*III was performed (Fig. 3.17). The RFLP pattern of the UWC10 *recA* gene digested with *Alu*I generated four distinct bands (Fig. 3.17, lane 2). This pattern was found to be identical to that of the *in silico* digestion profile of the *recA* gene sequence of *B. multivorans* (accession number AF143775) (Fig. 3.18). In addition, *Hae*III RFLP analysis generated a *B. multivorans* signature pattern termed "C", which has previously been

observed with 20 other *B. multivorans* (Mahenthiralingam *et al.*, 2000). A *Dde*I 16S ARDRA pattern (Fig. 3.12, Lanes 6 and 7) was identical to that of *B. multivorans* pattern "D" previously described by Seo and Tsuchiya (2005).

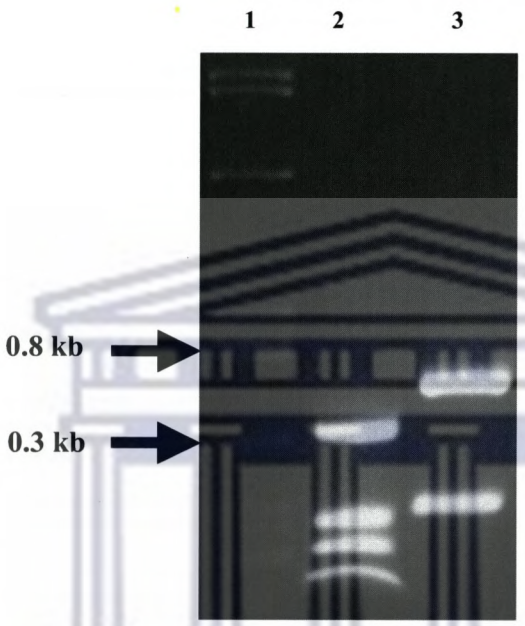


Figure 3.17: Agarose (2.5%) gel electrophoregram showing RFLP pattern of *recA* gene. Lane1: *Pst*I-digested λDNA marker, Lane 2: *Alu*I, and Lane 3: *Hae*III.

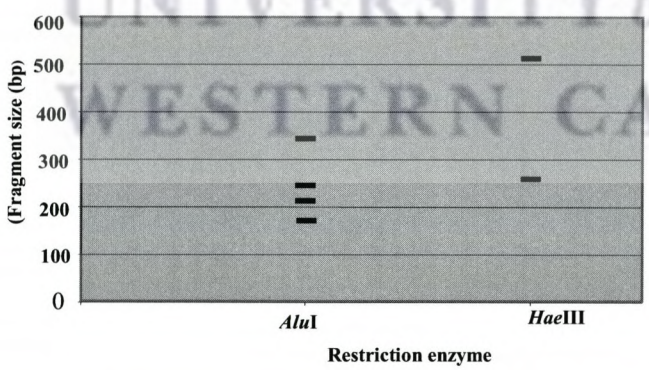


Figure 3.18: *In silico* generated *B. multivorans* (accession number AF143775) *rec*A gene RFLP pattern produced with *Alu*I and *Hae*III restriction enzymes. (Patterns were generated using Web-cutter and drawn in Microsoft Excel).

On the basis of these results (i.e., 16S rRNA gene sequence and ARDRA, 16S rRNA based phylogram analysis, species specific primer evaluation, *recA* and r*ecA* RFLP analysis) the UWC10 isolate was confidently designated as *B. multivorans*. The 16S RNA gene nucleotide sequence of this isolate was deposited in GenBank under the Accession number DQ103700.

3 .8 Discussion

To date, the number of reports describing the presence of EFH activity in bacteria is low. This has been mainly attributed to the failure of many phenolic acid esterase-screening protocols to support the growth of many microorganisms, particularly bacteria (Donaghy and McKay, 1994). The main goal of the present study was to isolate microbial cultures that could convert the model substrate ethyl ferulate to ferulic acid. An agar-screening assay developed by Donaghy *et al.* (1998) was used to screen bacterial isolates with EFH activity. A bacterial isolate, designated UWC10, demonstrated high EFH activity during the screening program and was selected for further studies.

Enzyme localization studies indicated that EFH activity was intracellularly located (Fig. 3.3). EFH activity from fungal species, including *Aspergillus niger* (Faulds *et al.*, 1997), *Humicola insolens* (Hatzakis *et al.*, 2003), *Fusarium oxysporum* (Topakas *et al.*, 2003a) and *Sporotrichum thermophile* (Topakas *et al.*, 2003b) is typically extracellular. However, there are conflicting observations regarding the localization of EFH activity from bacterial sources. The ferulic acid esterase activities from the number of actinomycetes such as *Streptomyces* strains C-248 and C-254 (Johnson *et al.*, 1988), *S.*

averitilis strain 3339 (Garcia *et al.*, 1998), and *S. olivochromogenes* (Faulds and Williamson, 1991) have been reported to be extracellular, while FAE activities from a number of *Bacillus* and *Lactobacillus* species were reported to be intracellular (Donaghy *et al.*, 1998).

The effect of carbon source on the specific EFH activity of isolate UWC10 was investigated in liquid media (Table 3.3). The production of EFH activity under the experimental conditions studied was insensitive to catabolite repression from simple carbon sources such as glucose, and appeared to be constitutively expressed. Conversely, FAE expression in *S. overmitilis* was reported to be dependent on the presence lignocellulose substrates (Garcia *et al.*, 1998), and the FAE activity of *A. niger* has been shown to be induced by the presence of ferulic acid (Faulds *et al.*, 1997). There is only one report describing the regulation of FAE expression at the gene level (De Vries *et al.*, 1997). A carbon catabolite repressor protein, CreA, was identified in *A. niger* and shown to prevent transcription of the *faeA* gene in the presence of glucose and fructose.

Growth and activity profiles (Fig. 3.4) suggested growth-related expression of EFH activity. The specific EFH activity was the highest at the transition from exponential to stationary phase, followed by a decrease in specific EFH activity during the stationary growth phase. These observations suggested the cessation of expression of the EFH encoding gene(s) during stationary phase, most probably with concurrent induction of other degradation pathways (Morihara, 1974).

Biotransformation studies of EF, using both the growing and resting cells, were also performed. TLC and HPLC analysis suggested that ferulic acid produced during biotransformation of EF was further metabolized to vinyl guaiacol. Bioconversion of FA to vinyl guaiacol by isolate UWC10 as observed in this study was not surprising. There are a number of reports describing the bioconversion of ferulic acid to vinyl guaiacol by bacteria, including *Ps. cepacia* (now *B. cepacia*) (Andreoni *et al.*, 1984), *Fusarium solani* (Nazareth and Mavinkurve, 1986), mutant strains of *Ps. fluorescens* (Andreoni *et al.*, 1995), *B. pumilus* (Degrassi *et al.*, 1995) and *Amycolatopsis* sp. (Achterholt *et al.*, 2000).

Detailed molecular identification based on two conserved genes, 16S rRNA and *recA,* was undertaken in order to establish the identity of isolate UWC10. A BLAST search revealed that the 16S rRNA gene had high sequence identity to a number of *Burkholderia* species but did not support identification to species level. Further PCR analysis using species-specific primers showed that genomic DNA could only be amplified with G1/SPR3 primer pair set. Based on the PCR algorithm developed to identify *Burkholderia* species, isolate UWC10 was tentatively classified as *B. multivorans* (Whitby *et al.*, 2000).

To confirm the assignment, a *B. multivorans*-specific primer pair (Mahenthiralingam *et al.*, 2000), targeting the *recA* gene was used giving the expected 1043 bp PCR amplicon. In addition, ARDRA and RFLP analyses with *Dde*I and *Hae*III of the 16S RNA and *recA* genes gave the predicted patterns (Seo and Tsuchiya, 2005; Payne *et al.*, 2005). Based on

these results, isolate UWC10 is confidently designated as *Burkholderia multivorans* strain UWC10.

CHAPTER FOUR

Cloning and sequencing analysis of the lipase and esterase encoding genes from *Burkholderia multivorans* UWC10

4.1 Introduction

The novelty of an enzyme can be assessed most explicitly if its primary structure is known. Functional analysis alone frequently does not permit two enzymes of similar functional class to be distinguished. This problem is of particular importance for the ester hydrolase class of enzymes, which catalyze a wide range of substrates. For example, the standard ρ-nitrophenyl esters, which are used to assay esterase and lipase activities, can also be hydrolyzed by some proteases and peptidases.

In addition, some organisms produce multiple isozymes. The enzymatic properties of a crude enzyme preparation from such an organism will represent an average of the traits of all isozymes present. For the detailed and accurate analysis of enzymatic properties, it is critical to clone individual enzyme genes to allow the separation of genes encoding different isozymes. In addition, the over-expression of cloned genes greatly facilitates subsequent protein purification, and is an important step in determination of protein structure (i.e., by crystallization and X-ray diffraction).

Considerable numbers of lipase and esterase genes from different sources have been cloned and in some cases successfully expressed. This has been facilitated by the development of fast, reliable and robust *in situ* plate screening assays (see review by

Beisson *et al.* (2000) and Gupta *et al.* (2003). Amongst the various *in situ* plate screening assay systems, the tributyrin and olive oil agar plate assays for esterases and lipases respectively are widely used.

This chapter reports the construction of a *Burkholderia multivorans* UWC10 gene library in *E. coli*, and the subsequent screening of the library for identification of genes encoding lipase and esterase activities.

4.2 *Burkholderia multivorans* UWC10 library construction and screening

In earlier experiments, *B. multivorans* UWC10 was found to show good esterase and lipase activities in both qualitative and quantitative assays (Chapter 3: Sections 3.2 and 3.3). To identify and isolate esterase and lipase genes, a shotgun genomic DNA library of this strain was constructed using the high copy number cloning vector pUC18. *E. coli* (DH5α), which exhibits very low endogenous esterase activity, was used as the host.

Genomic DNA of *B. multivorans* UWC10 was partial digested with *Sau*3A1 (Fig. 4.1). The resulting 2-8 kb fragments were recovered and ligated into pUC18 which was digested with *Bam*HI and treated with shrimp alkaline phosphatase. The ligation product was used to transform competent *E. coli* (DH5α) cells. The library contained 10^4 colony-forming unit (CFU) with 82% recombinants based on blue and white screening.

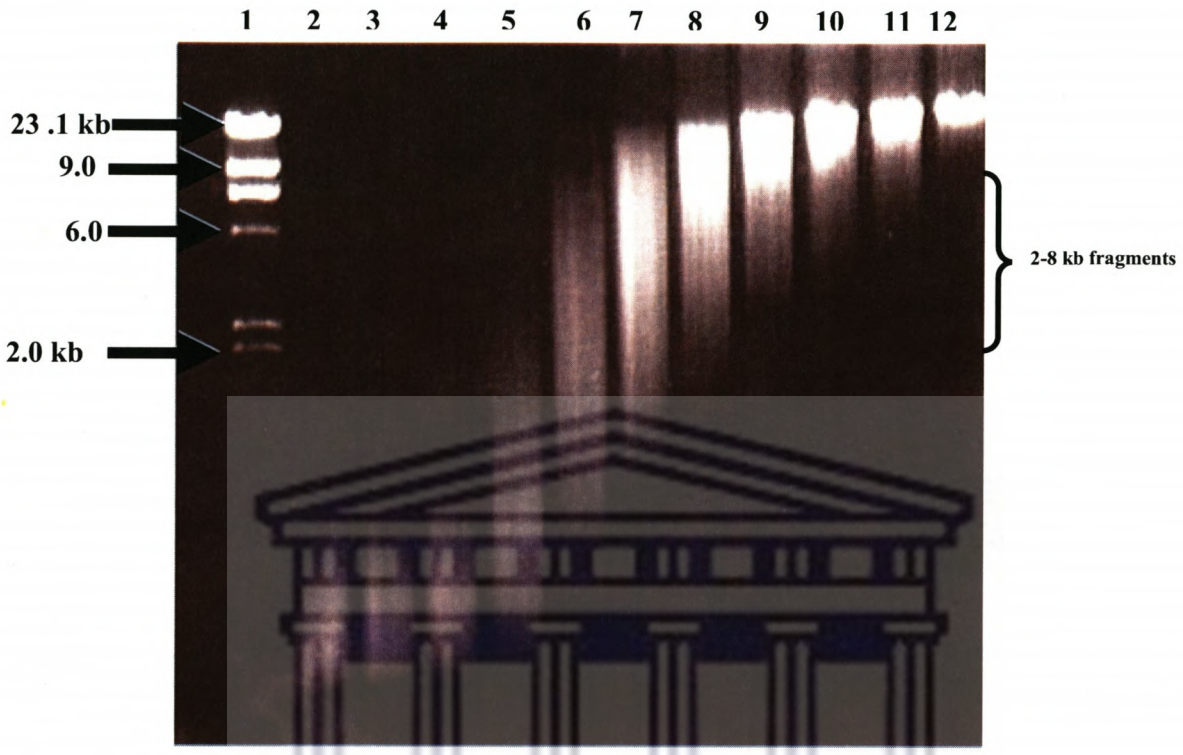


Figure 4.1: Agarose (0.8%) electrophoretogram showing partial digestion of *B. multivorans* UWC10 genomic DNA by *Sau*3A1: Lane 1: *Hind*III-digested λDNA markers; Lane 2: 0.002 U; Lane 3: 0.004 U; Lane 4: 0.008 U; Lane 5: 0.02 U; Lane 6: 0.031 U; Lane 7: 0.0625 U; Lane 8: 0.125 U; Lane 9: 0.25 U; Lane 10: 0.5 U; Lane 11: 1.0 U; Lane 12: 0 U.

Library screening for esterase and lipase gene(s) was carried out in three steps. Firstly, colonies with halo-forming activity on tributyrin agar plates were isolated. Since both esterases and lipases can hydrolyze tributyrin, the colonies with halo-forming activity were then tested for lipase activity using an agar plate assay containing olive oil and Rhodamine B. Thirdly, colonies were further screened for specific EFH activity using FAE plate assay. Approximately, 10 000 ampicillin-resistant recombinants were screened and several clear halo-forming clones were identified. Three highly active clones (which

produced large clear halos on tributyrin agar plates after two days of incubation at 30 °C) were selected. These clones, designated pRASH14, pHOLA6 and pTEND5 were sub-cultured and their plasmids isolated for further studies. The results of the qualitative activity assays of the three clones are shown in Table 4.1. All three recombinant clones formed halos in the absence of the inducer IPTG, indicating that the genes encoding esterolytic or lipolytic activities were being expressed in *E. coli* under the control of their intrinsic promoters.

Table 4.1: Esterase and lipase activities in *E. coli* (DH5α) transformants identified during the screening of the *B. multivorans* UWC10 genomic library

Clones	Esterase activity	Lipase activity	EFH activity
pRASH14	+	+	-
pHOLA6	+	-	-
pTEND5	+	-	+

+ = Activity; - = No activity

4.3 Sub cloning and sequencing of pRASH14

4.3.1 Subcloning of pRASH14

Restriction endonuclease digestions analysis of the pRASH14 lipase-positive clone revealed the presence of a plasmid DNA insert of about 7.5 kb (Fig. 4 2). Since this clone contained a DNA insert too large for immediate sequencing, active subclones containing smaller inserts were constructed. Restriction endonuclease digestion of the pRASH14 with *Sma*I resulted into two distinct DNA bands (Fig. 4.2, lane 4). A top band (labelled

A) was religated directly (because it was shown to contain a vector backbone) to yield a plasmid designated pRASH14A, while a bottom band (labelled B) was blunt-end cloned into *Sma*I restricted pUC18 and was designated pRASH14B.

Figure 4.2: Agarose (1%) gel electrophoregram showing restriction endonuclease digestions of the pRASH14 plasmid DNA: Lane 1: *Pst*I-digested λDNA markers; Lane 2: pRASH14 digested with *Hind*III; Lane 3: pRASH14 digested with *Hinc*II; Lane 4: pRASH14 digested with *Sma*I.

Following transformation and screening on tributryin agar plates, active clones could only be identified with *E. coli* transformants harbouring the pRASH14B plasmid. Restriction endonuclease mapping revealed DNA inserts of about 3.0 kb from these

active clones. A plasmid DNA was isolated from one of the representative clones and was used to sequence the insert in both directions using a primer walking technique (Myrick and Gelbart, 2002). Primers used for sequencing the 3.0 kb DNA insert of the pRASH14B are shown in Table 4.2.

Table 4.2: Primers used to sequence the pRASH14B DNA insert

Primer name	5′→3′	Positions	Reference
M13F/pUC18	GTAAAACGACGGCCAGT	380-397[a]	Yanisch Perron *et al.*,1985
M13R/pUC18	CAGTATCGACAAAGGAC	464-481[a]	Yanisch Perron *et al.*,1985
Lip281	ATCGCTTCGATTCCGTTCGAAC	386-307[b]	This study
Lip649	GAACAATTGCTGGCATACGTC	654-674[b]	This study
Lip1302	GAACTTCAAATGGAATCATCTCG	1307-1329[b]	This study
Lip1731	CTATTGCCTGACTGCACAGAGC	1736-1756[b]	This study
Lip1981	GACTGGAGCGAACCGTTTTTCG	1986-207[b]	This study
Lip2401	TGTTTACGAAACCGGGCGAAGC	2407-2429[b]	This study
Lip398	TAATCGGCATAGCGGCTTTGC	403-423[b]	This study
Lip964	TCTGTGCAGTCAGGCAATAGTCGAAG	968-993[b]	This study
Lip1381	TCGAGATGATTCCATTTGAAGTTCG	1494-1418	This study
Lip1783	GTCCTGGTTCGTGTTGTGACTGC	1788-1810[b]	This study
Lip2501	GTTTGCGTGCCGACTAGATACG	2506-2527[b]	This study
Lip2041	TTGACGTATGCCAGCAATTGTTC	2055-2077[b]	This study

[a]Positions in relation to pUC18 (Accession number L09136)

[b]Positions in relation to sequenced pRASH14B DNA insert

4.3.2 Sequence analyses of the pRASH14B

To establish the existence of a lipase-encoding region within the cloned *Sma*I fragment of the pRASH14B, a 2710 bp sequence was analyzed. Sequence analysis of the pRASH14B insert revealed two complete open reading frames (ORFs) (Fig. 4.3). ORF1 consisting of 1095 pb (nt. positions 331-1426) commencing with an ATG start codon and encoding 356 amino acid residues terminated with a TGA stop codon. ORF2, comprising of 1050 pb, encoded 350 amino acids and was located downstream from ORF1, starting with an initiation ATG codon at nt. position 1429 and extending to a TGA termination codon at nt. position 2481. The identities of the two ORFs were deduced with the aid of BLAST searches (Appendix 4.I) and were shown to be highly homologous to prokaryotic lipases and lipase chaperones, respectively. The genes encoding ORF1 and ORF2 were designated *lipA* and *lipB* respectively.

The lipase chaperone gene, *lipB* was located 3 bp downstream of the lipase structural gene *lipA* (Fig. 4.3). The *lipA* and *lipB* gene fragments showed high GC contents (66 and 73%, respectively). A putative ribosome binding site (RBS) for the *lipA* gene was found located 7 bp upstream of the *lipA* start codon (^{318}AGGAG). Further upstream of the *lipA* gene was a putative -10 transcriptional initiation site (^{312}ATAATCA), separated by 20 bp from a -35 region (^{280}TTCGAT). No promoter-like regions were identified in the sequences preceding the *lipB* coding region. The first 250 bp of the 5′ region of the *lipB* gene had a high GC content (77%), while the 26-nt. region (nt. positions 1523-1549) of the *lipB* contained direct (–CGCGTCGCC-) and indirect (-GCCCGT-) repetitive sequences.

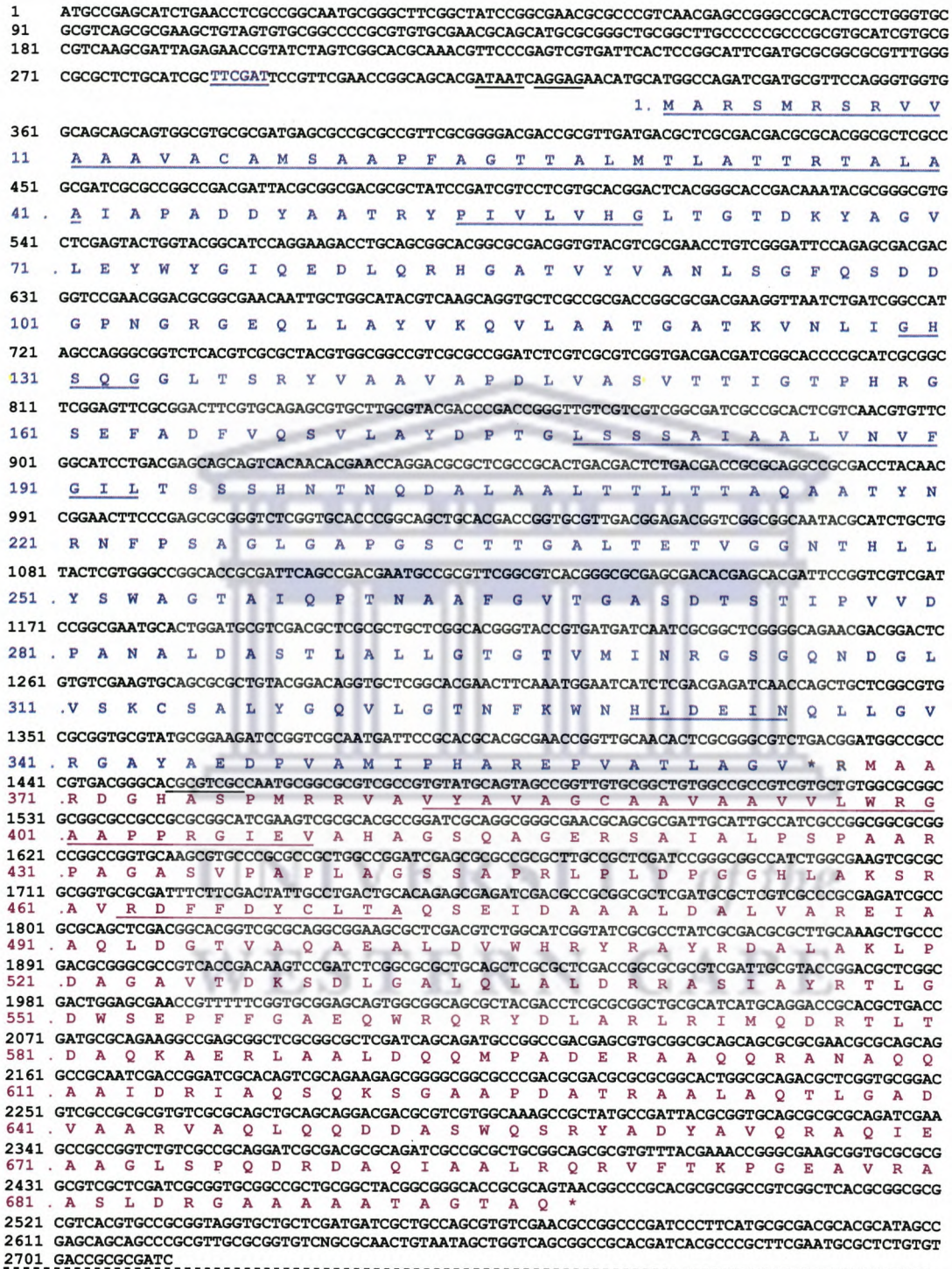


Figure 4.3: Nucleotide sequence (black) and deduced amino acid sequences of the lipase ORF 1 (blue) and ORF 2 (pink) from *B. multivorans*. Motifs discussed in the text are underlined.

4.3.3 Analysis of the deduced amino acid sequences of LipA and LipB

The *lipA* gene encoded a 356 amino acid protein with an estimated molecular weight of 37.3 kDa (Fig. 4.3), designated LipA. Based on analysis by SignalP 3.0 (Bendten *et al.*, 2004), the LipA pre-protein was predicted to contain a putative 40 amino acid N-terminal signal or leader peptide (with a maximum cleavage site probability of 0.9 between two alanine residues (a.a. positions 40-41) which would be cleaved to form a mature protein of 324 a.a. with predicted molecular weight and pI values of 33.2 kDa and 4.93, respectively (http://www.cbs.dtu.dk/services/SignalP/). The deduced amino acid sequence of LipA exhibited typical α/β-hydrolase fold features at amino acid positions 50-340, based on the InterProScan tertiary structure prediction program (http://www.expasy.ch).

The GHSQG motif (a.a. positions 129-133) corresponded well with the pentapeptide GxSxG motif which is conserved in lipases, esterases and many other hydrolytic enzymes and normally harbours a catalytic serine. The putative active site serine (Ser131) is encoded by an AGC codon in the same manner as many other lipases. Based on the high homology (78% identity, with 96% similarity) of the primary structure of LipA with that of lipase P22088 from *B. cepacia* (Jorgensen *et al.*, 1991), for which the three dimensional structure has been solved (Kim *et al.*, 1997a), it can be deduced that Ser131, Asp308, and His330, probably comprise the catalytic triad of LipA while the cysteine residues Cys234 and Cys314 are involved in disulphide bond formation. The aspartic residues Asp286 and Asp332 may form a calcium binding site in the mature LipA. The region from amino acid position 178 to position 193 corresponds well to the 'lid' domain

as previously shown in the 3D structures of *B. cepacia* and *B. glumae* lipases (Noble *et al.*, 1993; Kim *et al.*, 1997a). A putative oxyanion hole region (YPIVLVHG) comprised of the His-Gly (HG) dipeptide and a short hexapeptide of hydrophobic residues upstream of HG was observed (a.a. positions 54-60) within the LipA primary sequence. The HXDXZN motif (where X represents any amino acids and Z a hydrophobic residue) where the active site histidine is normally located was also observed in the LipA primary structure (a.a. positions 330-334).

The *lipB* gene encoded a 350 amino acid protein designated LipB with an estimated molecular weight of 37.3 kDa and predicted pI value of 8.64. Database searches revealed that LipB belongs to a class of proteins known as lipase chaperones (Appendix 4.I), that do not show any significant homology to any known class of proteins (Rosenau *et al.*, 2004). This class of proteins has been recently defined as lipase specific foldase and is involved in the folding and secretion of lipases via a two-step Type II secretion pathway (Rosenau *et al.*, 2004). Analysis of the LipB primary structure by the method of Eisenberg *et al.* (1984) identified a hydrophobic stretch of amino acids (extending from Val383 to Arg409) comprising a putative membrane anchor domain. Adjacent to the putative membrane anchor domain is a proline-alanine-rich stretch (a.a. positions 413-471) of amino acid residues covering almost 20% of the entire LipB protein. The RDFFDYCLTA motif (a.a. positions 453-462) corresponds to -RX_1X_2FDY-(F/C)L(S/T)A-, the only conserved motif in lipase specific foldases.

Prediction algorithms (PHD: available on www.expasy.org) revealed that the lipB was composed of the following secondary structure elements: α-helical (67.1%) and random coil (33%). The amino acid sequences of LipA and LipB were deposited in the GenBank under the accession number AAZ39650 and AAZ39651, respectively. The physical map of the *B. multivorans* lipase operon is shown in Fig. 4.4.

Multiple sequence alignments of LipA with other Group I proteobacterial lipase family members showed complete conservation of the putative catalytic triad residues (Ser183, Asp350 and His396) throughout the whole family (Fig. 4.5). The pentapeptide GxSxG motif (positions 181- 185) also appeared to be conserved with an exception that the residue after the catalytic serine was either histidine or glutamine. The oxyanion dipeptide HG (position 110-111) region was highly conserved. The hydrophobic stretches that flank the HG region were also typically present but with only proline (Pro105) being strictly conserved throughout the Group 1 proteobacterial lipase family. The cysteine residues (Cys290 and Cys377) that are involved in disulphide bond formation are conserved throughout the Group I proteobacterial lipases with the exception of *Ps. fluorescens* lipase. Aspartic residues Asp217 and Asp350, which could form the calcium binding site, were highly conserved. The HXDXZN motif (positions 396-401) that harbours the catalytic histidine (H396) is also strictly conserved. The lid domain (-LSSSAIAALVNVFGIL-) stretching from position 187-2002 was only strictly conserved in lipases from the *Burkholderia* group.

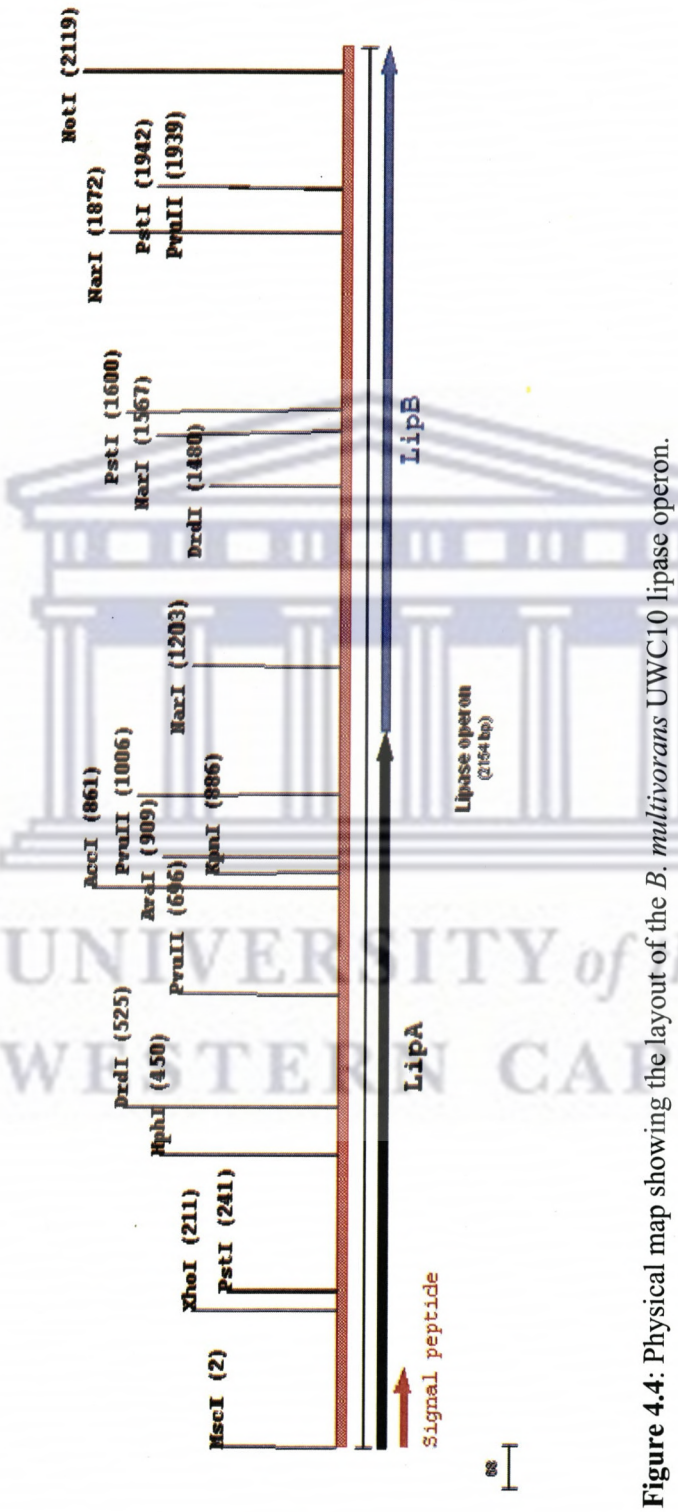


Figure 4.4: Physical map showing the layout of the *B. multivorans* UWC10 lipase operon.

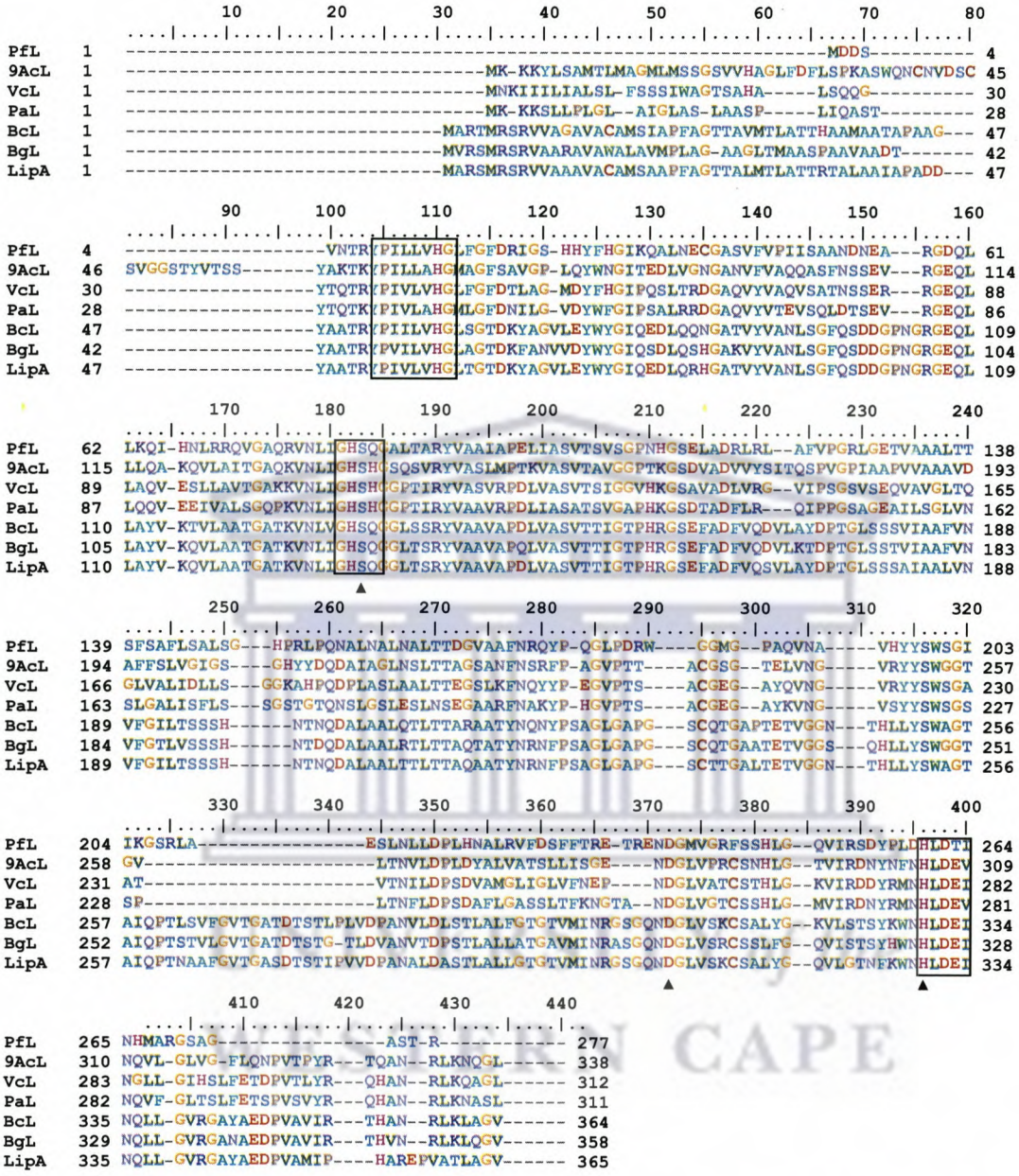


Figure 4.5: Multiple sequence alignments of *B. multivorans* lipase, LipA (AAZ39650) and other Group I proteobacterial lipases represented by lipases from *Ps. fluorescens* IFO, PfL (X14033); *Ac. calcoaceticus* RAG1, AcL (AF047691), *V. cholera* EIT, VcL (Y00557), *Ps. aeruginosa* TE3, PaL (AB008452), *B. glumae* PG1, BgL (X703540), and *B. cepacia* DMS, BcL (M58494). Catalytic triad residues are indicated by solid triangle. The motifs discussed in the text are boxed.

4.3.4 Comparison of the deduced amino acids sequences of LipA and LipB with other lipases

The deduced amino acid sequence of LipA was compared with other Group 1 proteobacterial lipases (Table 4.3). The LipA amino acids sequence revealed high sequence similarity with lipases classified in subfamily I.2 (the so-called *Burkholderia* lipase group). The LipA was 86.8% identical and 96.1% similar to the lipase from *B. cepacia*, 77.0% identical and 94.1% similar to *B. glumae* and 86% identical and 96.3% similar to the lipase from *Pseudomonas* sp. KWI-51. Homology searches between LipA and lipases from subfamilies I.1 and I.3 was also performed. The lipases from subfamily I.1 represented by *Ps. aeruginosa* and *Acinetobacter* showed 38-42% sequence identity to LipA while subfamily 1.3 represented by *Ps. fluorescens* and *P. vulgaris* showed 38-40% sequence identity. The *Vibrio* group showed 44-46% sequence identity to LipA.

A search of the available protein sequence databases revealed that LipB has high homology to lipase chaperone (i.e., lipase specific foldase). Amino acid sequence comparisons indicate that LipB is 77% (with 99% similarity), 62 % (with 92% similarity) and 79% (96 % similarity) identical to lipase chaperones from *B. cepacia*, *B. glumae* and *Pseudomonas* sp. KWI-51, respectively (Table 4.4). The lipase chaperones from *Pseudomonas*, *Acinetobacter*, *Vibrio* groups showed low sequence identities to LipB, 27-29 %, 25-27% and 24-26%, respectively.

Table 4.3: Comparison of the *B. multivorans* lipase (LipA) sequence with corresponding sequences from the Group 1 Proteobacterial lipase family

Lipase producing strain	Accession No.	Identity (%)	Similarity (%)
Burkholderia cepacia DSM	M58494	86.78	96.13
Pseudomonas species KWI-51	D10069	86.23	96.23
Burkholderia glumae PG1	X70354	77.03	94.12
Pseudomonas aeruginosa TE3585	AB008452	42.3	82.93
Pseudomonas aeruginosa	P26876	41.18	82.68
Pseudomonas pseudoalcaligenes M-1	A08195	40.72	82.74
Acinetobacter calcoaceticus RAG-1	AF047691	41.30	81.05
Acinetobacter sp. SY-01	AF518410	42.77	77.84
Acinetobacter calcaceticus BD413	X80800	38.39	74.69
Pseudomonas fluorescens C9	AF031226	42.30	82.63
Pseudomonas fluorescens IFO	X14033	40.44	79.04
Proteus vulgaris K80	U33845	38.39	79.86
Vibrio cholera TE3	Y00557	44.01	81.87
Vibrio parahaemolyticus	NC004603	43.89	82.68
Vibrio vulnificus	AE016804	46.18	81.3

Table 4.4: Comparison of the *B. multivorans* chaperone (LipB) sequence with corresponding sequences from lipase chaperone families I-IV.

Chaperone producing strain	Accession Number	Identity (%)	Similarity (%)
Pseudomonas species KWI	P25276	79.39	96.66
Burkholderia cepacia DSM	B39133	77.98	96.66
Burkholderia glumae	Q05490	62.28	92.69
Pseudomonas pseudoalcaligenes	CAA02276	29.22	73.20
Pseudomonas aeruginosa PAO1	CAA44998	28.33	71.33
Pseudomonas mendocina	AAM14702	27.16	72.84
Acinetobacter calcoaceticus BD413	Q43961	25.46	73.20
Acinetobacter calcoaceticus RAG-1	Q9X2S4	25.69	77.37
Acinetobacter sp.	AF518410	20.91	62.36
Vibrio cholera	Y00557	26.55	69.91
Vibrio parahaemolyticus	NP_ 797559	24.00	72.73
Vibrio vulnificus	AE016804	26.91	68.36

For further comparisons of the LipA amino acid sequence with corresponding sequences from other bacterial lipases, a phylogenetic analysis was performed. Comparisons of the LipA amino acid sequences with corresponding sequences from Group 1 Proteobacterial lipases using Neighboring-Joining Distance analyses yielded the phylogenetic trees shown in Fig. 4.6. The lipase phylogenetic tree shows that, as expected, LipA clusters with lipases from *Burkholderia* species, confirming that LipA belongs to lipase subfamily I.2 (Arpigny and Jaeger, 1999). The good bootstrap support (100%) and high sequence identity (80%) further supported this conclusion.

Phylogenetic analysis was also carried out to explore the relationship between LipB and other lipase chaperones. Figure 4.7 suggests that the phylogeny of these enzymes is largely concordant with the Group I proteobacterial lipases. Lipase chaperones are classified into four families (I-IV), with family II comprising of *Burkholderia* lipase chaperone family. Based upon a sequence identity of greater than 70% with family II lipase chaperones and placement in the phylogenetic tree (Fig. 4.7), it was concluded that LipB belong to family II (Rosenau *et al.*, 2004).

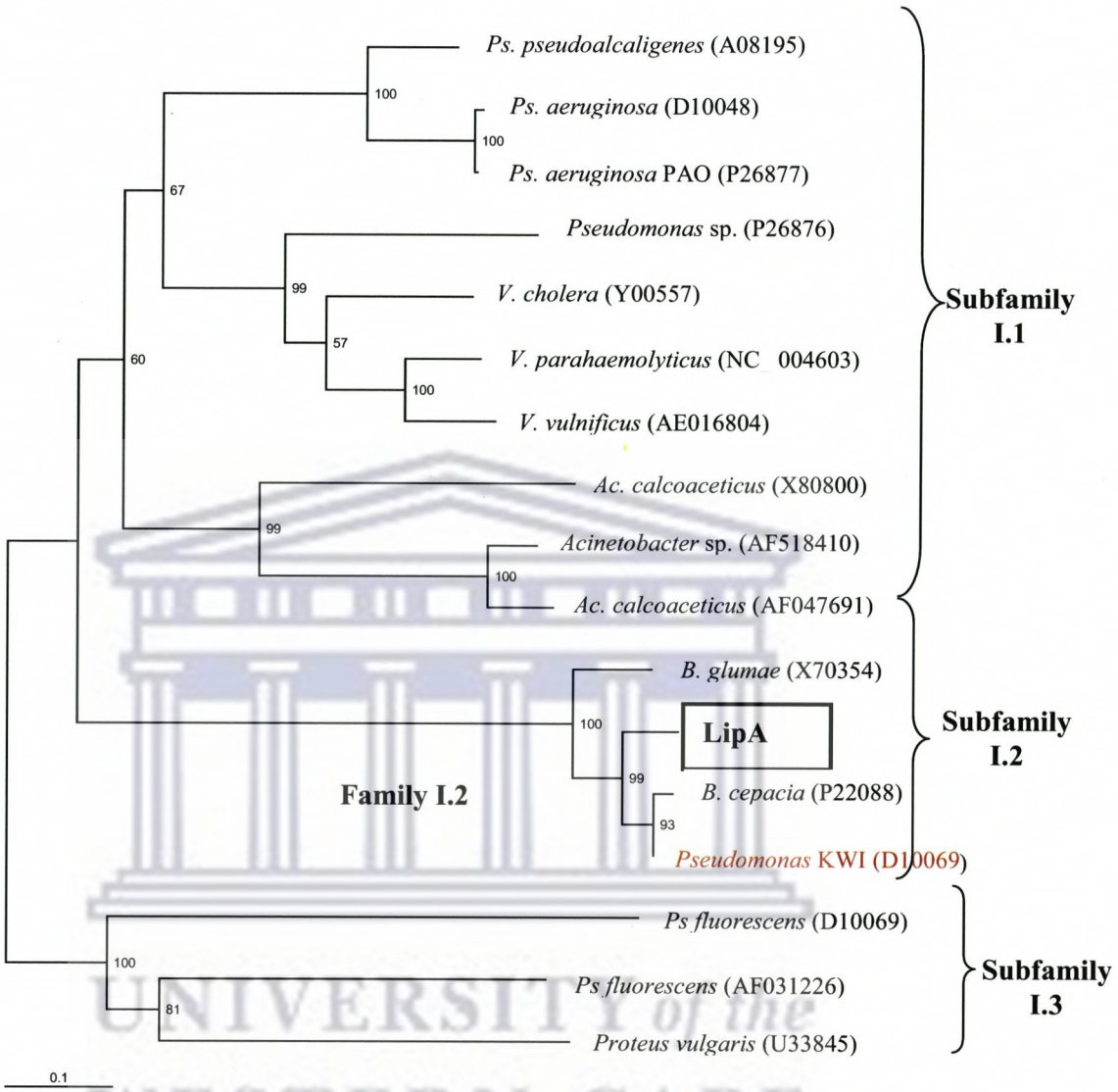


Figure 4.6: Evolutionary distance phylogram showing the position of lipase (LipA) from *B. multivorans* in relation to Group1 proteobacterial lipases.

Sequences were aligned using CLUSTALX. The phylogenetic tree was constructed using PHYLOWIN package based on Neighbor-Joining method. The scale represents 0.1amino acid substitution per position. The numbers on the tree show the levels of bootstrap support based on 500 resamplings. Sequences were obtained from GenBank.

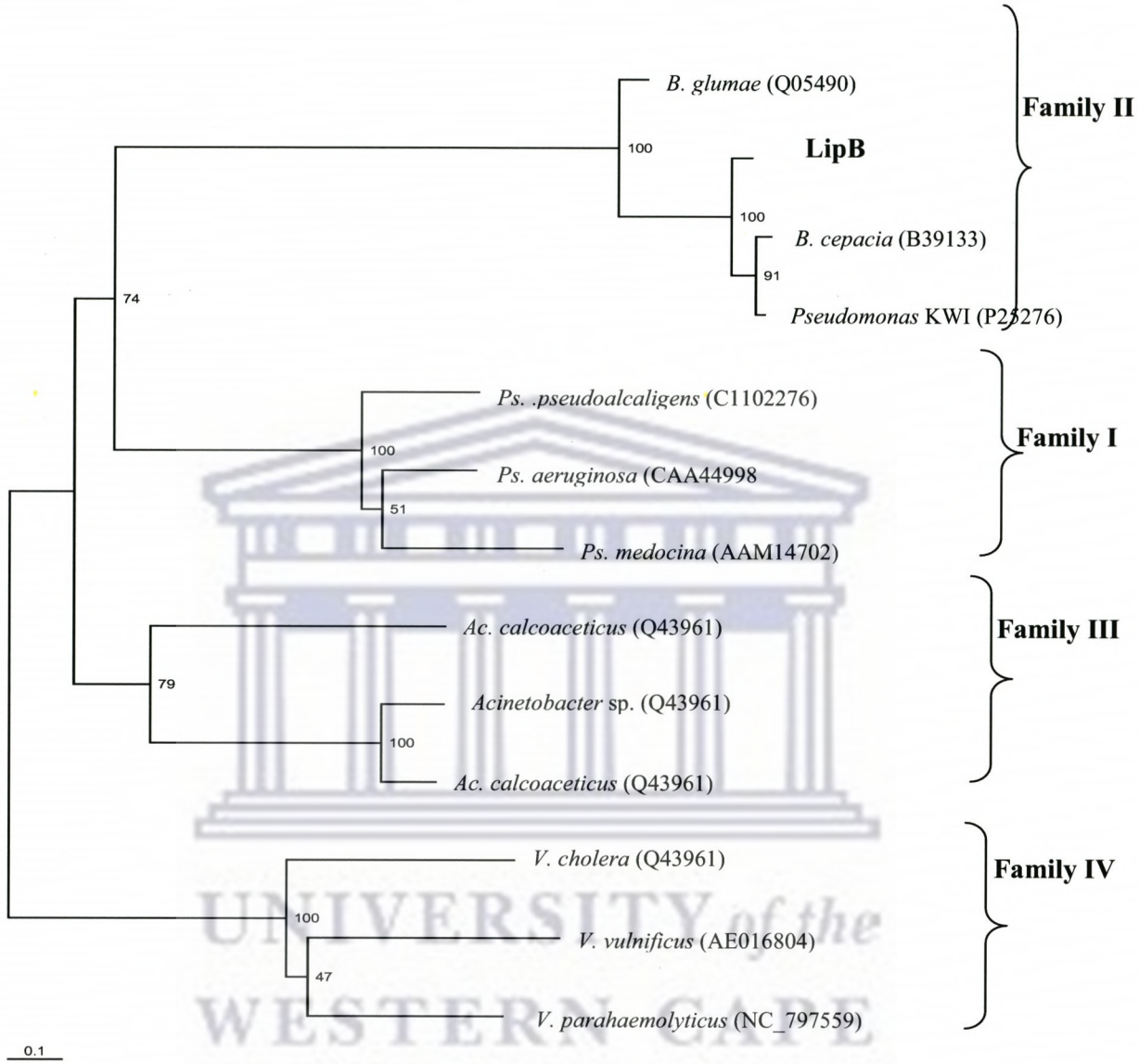


Figure 4.7: Evolutionary distance phylogram showing the position of lipase-chaperone (LipB) from *B. multivorans* in relation to Group1 proteobacterial lipase-chaperone.

Sequences were aligned using CLUSTALX. The phylogenetic tree is based on Neighboring-Joining analysis using PHYLOWIN package. The scale represents 0.1 amino acid substitution per position. The numbers on the tree show the levels of bootstrap support based on 500 resamplings. Sequences were obtained from the GenBank.

4.4 Sub cloning and sequencing of pTEND5

4.4.1 Subcloning of pTEND5

E. coli transformant (pTEND5) showed a zone of clearance on both tributryin and FAE plate during library screening. Restriction analysis indicated that this clone contained a DNA insert of about 2.1 kb (Fig 4.8). The insert was then sequenced in both directions using the primer-walking method (Myrick and Gelbart, 2002). The primers used for the sequencing of the complete 2 .1 kb DNA insert fragment of this clone are shown in Table.4.5.

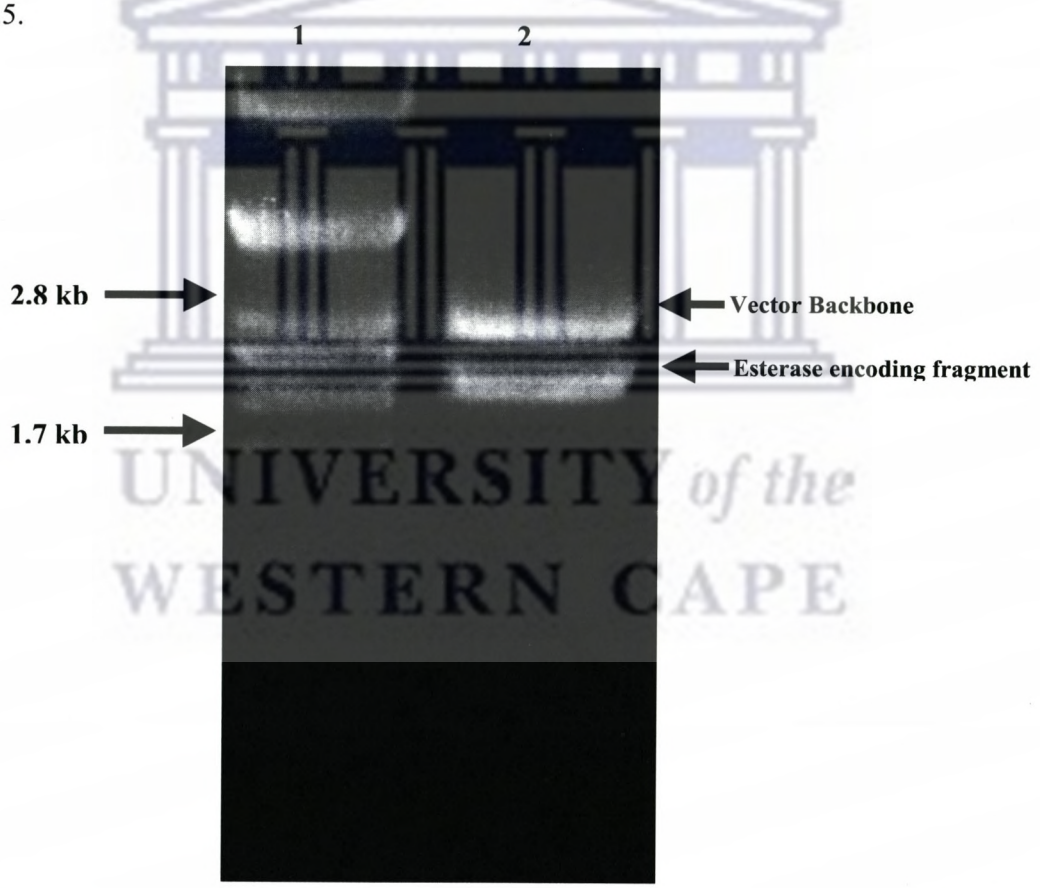


Figure 4 8: Agarose (1%) gel electrophoregram showing restriction digestion of pTEND5 plasmid DNA. Lane 1: *Pst*I-digested λDNA markers; Lane 2: pTEND5 digested with *Sma*I and *Xba*I.

Table 4.5: Primers used to sequenced the pTEND5 DNA insert

Primer name	5'→3'	Position	Reference
M13F/pUC18	GTAAAACGACGGCCAGT	380-397 [a]	Yanisch-Perron *et al.*,1985
M13R/pUC18	CAGTATCGACAAAGGAC	464-481 [a]	Yanisch-Perron *et al.*,1985
EF448	AGTTTCATCAGAACTACGTGCTCG ATCTCG	448-478 [b]	This study
ER747	CATCCACTATGTCGAATACGGCAGC	747-771 [b]	This study
ER1168	CT TCGTGTCGTGGACCTTCG	1148-1168 [b]	This study
EF1627	TGGTAGGCCTTGTTGGCCTGTGTCG	1627-1651 [b]	This study
ER1677	CGACGCAGTTTGACGCAAGCCAATG	1677-1701 [b]	This study
ER1287	CACAGCAGGTCGGTCGCGGTCGCATAG	1287-1313 [b]	This study
ER1910	GAACACGTTCGCCTGCGATC	1910-1930 [b]	This study
ER614	CGACGATGCCGAGCAGGATGTTCAGG	614-640 [b]	This study
ER181	CTGTACATCATCCCTTTATAGGACACC	181-207 [b]	This study

[a]Position in relationship to pUC18 (Accession number L09136)

[b]Position in relation to sequenced pTEND5 DNA insert

4.4.2 Problems with sequencing the pTEND5 insert DNA

Initial sequencing of the pTEND5 insert DNA using M13F and M13R primers yielded high quality sequencing data. However, an attempt to "primer walk" to recover the complete nucleotide sequence using a series of designed primers (Table 4.5) largely failed and often resulted in a very short sequence reads. Literature searches of the

previously sequenced DNA fragments from *Burkholderia* species revealed a relatively high molar GC content (>60%) (Holden *et al.*, 2004; Nierman *et al.*, 2004). Furthermore, most of the templates revealed the presence of direct and indirect repeat sequences. Based on this information, it was concluded that unsuccessful sequencing attempts might be attributed to the formation of secondary structure on the nascent strands and/or slippage of the DNA polymerase accompanied by premature chain termination in repeat regions (Motz *et al.*, 2000).

In order to overcome these problems, sequencing reaction conditions were determined empirically. Long sequence reads were obtained after addition of 0.2 M betaine and 2% dimethyl sulfoxide (DMSO) as denaturants in the sequencing reaction mixtures (Motz *et al.*, 2000; Seto *et al.*, 1995). In addition, the denaturation and annealing times were increased to 98°C for 6 min and 50 °C for 20 min, respectively in the presence of the denaturants to allow successful primer extension for linear growth of sequencing reaction products (Seto *et al.*, 2000).

4.4.3 Sequence analysis of the cloned pTEND5 DNA insert

The nucleotide sequence of the 2.1 kb *Sau*3A1 insert of pTEND5 was determined (Fig. 4.9). Sequence analysis revealed two distinct ORFs. One of the putative open reading frames (ORF1, 597 bp; nt. positions 2-598) was found truncated by the *Sau*3A1 at the N-terminus and terminated with a TGA stop codon (Fig. 4.9). The deduced amino acid sequence derived from this putative incomplete ORF1 revealed a polypeptide of 181 a.a. with significant homology to the number of bacterial flavin-containing monooxygenases,

TrkA (Appendix 4.II). The amino acid sequence of this ORF was 77% (138/177) identical to monooxygenase from *B. pseudomallei* (Accession no. CAH39447), 62% (110/117) to *Caulobacter crescentus* (Accession no. AAK2529), 54% (96/176) to *R. metallidurans* (Accession no. AEN53123), and 51% (89/172) to *Bradyrhizobium japonicum* (Accession no. BAC51207). The incomplete monooxygenase gene was designated *trkA* and the deduced corresponding polypeptide was designated TrKA.

The second putative open reading frame (ORF2) was identified in the second translation frame immediately downstream of the ORF1, overlapping with ORF1 by four nucleotides. ORF2 consisted of 978 bp commencing with an ATG start codon at nt. position 595 and extending to a TGA stop codon at nt. position 1570. The last two nucleotides of the ATG start codon of ORF2 formed part of the TGA stop codon of the truncated *trkA* gene. Possible nucleotide sequences corresponding to typical elements of *Burkholderia* or *E. coli* promoter sequences were not found in the region upstream the predicted ORF2. However, an -AGGC- sequence located 4 bp upstream the ATG start codon was presumed to be the putative RBS for the ORF2, based on its close similarity to the consensus Shine-Dalgano (-AGGA-) sequence. Translational analysis of the ORF2 revealed a polypeptide of 326 a.a. encoding a protein of 34.83 kDa with a predicted pI value of 10.02. A very high GC content (74%) was found within the predicted ORF2. Furthermore, the ORF2 nucleotide sequence revealed a number of direct and indirect sequences repeats (Table 4.6). These repetitive sequences are likely to form secondary structures. The possibility of secondary structures formation within the pTEND5

nucleotide sequence is presumed to have contributed to initial sequencing problems, subsequently resolved by addition of denaturants.

Table 4.6: Summary of the positions and sizes of the direct and indirect nucleotide sequence repeats deduced within the ORF2 of the pTEND5

Direct Repeats				Indirect repeats				
Sequences	**Size**	**Pos 1**	**Pos2**	**Sequences**	**Size**	**Pos 1**	**Pos2**	**dG**
CGCGCTCGCG	10	652	1012	CGGCATCGTC	10	634	1489	2.3
CGCGCTCGCG	10	652	1423	CGCGCGCCGC	10	679	1553	-1.6
GCTCGCGCTG	10	655	1054	CGTCGACCGGC	11	865	1363	-1.3
CGGCTCGCGC	10	836	1052	CGTCGACCGGC	11	1002	1063	-3.1
CCGGCCGCTCG	11	882	1108					
CGTGTTCGCGC	11	919	1207					
CGGCGGCGCG	10	1000	1551					
CGCGCTCGCG	10	1012	1423					
GCCGCTCGTGC	11	1111	1129					
CGGCGGCGGCG	11	1542	1548					

Analysis of the deduced amino acid sequence of ORF2 revealed no putative signal peptide (Fig. 4.9). The N-terminus of ORF2 was homologous to pancreatic lipases (Appendix 4.II). A classical signature motif, the pentapeptide GHSLG motif (a.a. positions 133-137); corresponding to the GxSxG motif which harbours the catalytic serine for many esterases, lipases and other hydrolases, was observed. The serine (Ser135) within the GHSLG motif is believed to be catalytic serine. InterProScan tertiary structure predictions showed classical α/β-hydrolase fold features stretching from position 8-311 (http://www.expasy.ch). Primary structure analysis of ORF2 using the

method of Eisenberg *et al.* (1984) and the topology prediction program (PredictProtein available on http://cubic.bioc.columbia.edu) identified four hydrophobic stretches at the following amino acid positions: Leu5-Val30, Gly133-Gly137, Phe185-Thr198, and Trp208-Arg223. The secondary structure prediction (PROFsec on http://cubic.bioc.columbia.edu) and prediction algorithms (PHD available on www.expasy.org) revealed that ORF2 contained the following secondary structure elements: α-helices (45%), β-sheets (42%) and random coil (13%). The long hydrophobic stretch (Leu5-Val30) was found located within an α-helix, and could potentially form a membrane anchor domain (Appendix 4.III). Only one cysteine residue (Cys240) was identified eliminating the possibility of disulphide bond formation.

Further amino acid sequence analysis of the ORF2 revealed a PIVFVHG (a.a. positions 61-67) motif corresponding well with a putative oxyanion region comprised of the HG dipeptide and a short hydrophobic stretch upstream the HG dipeptide. This oxyanion region is believed to stabilize the tetrahedral intermediates of the hydrolytic reactions. A putative oxyanion binding pocket HGDAL motif (a.a. positions 175-179), which corresponds well with the HXDXZ motif (where X represents any amino acids and Z a hydrophobic residue) was also observed. In subfamilies I.1 and I.2 lipases this segment is conserved as HLDEI or HILDEV respectively, and normally harbours the catalytic histidine. The physical map of the deduced putative proteins and the endonuclease restriction analysis of pTEND5 DNA insert is shown in Fig. 4.10. The presence of high GC rich region, direct and indirect sequence repeats and putative membrane anchors within the ORF2 might adversely affect protein expression in *E. coli.*

```
1     CCGCGCGTCGATCGTCACCGACGAGATCGAACGCTTCACGCCGACCGGGCTGAAACTGAAAAGCGGCGCACAGCTCGACGCCGACGTGAT 90
1       G  R  A  S  I  V  T  E  I  E  R  F  T  P  T  G  L  K  L  K  S  G  A  Q  L  D  A  D  V  V 30
91    CGTCACCGCGACCGGCCTGAAGGTGAAGATGCTCGGCGGCGCACGCGTGACGGTCGACGGCCGCGCGGTCGACCTGCCGCAGACGGTGTC 180
30      I  T  A  T  G  L  K  V  K  M  L  G  G  A  R  V  T  V  D  G  R  A  V  D  L  P  Q  T  V  S 60
181   CTATAAAGGGATGATGTACAGCGACGTGCCGAACCTCGCGTCGTCGTTCGGCTATACGAACGCGTCGTGGACGCTGAAAGCCGAACTGAT 270
61      Y  K  G  M  M  Y  S  D  V  P  N  L  A  S  S  F  G  Y  T  N  A  S  W  T  L  K  A  E  L  I 90
271   CGCGCGCTACGTGTGCCGGCTGCTGAACCACATGCGCGCGCACGGGTACGATACCTGCGTGCCGCGCCTCGCGCCCGGCGATCTCGGCGA 360
91    A  R  Y  V  C  R  L  L  N  H  M  R  A  H  G  Y  D  T  C  V  P  R  L  A  P  G  D  L  G  D 120
361   CGTGCCGGCCGTGAACCTGAGCTCCGGCTATATCCAGCGCGCGGCCGGCATCCTGCCGAAGCAGGGCCACCGCAAGCCGTGGAAGTTTCA 450
121     V  P  A  V  N  L  S  S  G  Y  I  Q  R  A  A  G  I  L  P  K  Q  G  H  R  K  P  W  K  F  H 150
451   TCAGAACTACGTGCTCGATCTCGCATCGCTGAAGTTCAGCGGCCTCGCCGACTCGGCAATGCATTTCGAACGCCGCGCGACAGCCGGCCC 540
151     Q  N  Y  Y  V  L  D  L  A  S  L  K  F  S  G  L  A  D  S  A  M  H  F  E  R  R  A  T  A  G  P  180
541   GGCGGCGACGCCGGCGACCGAACCCGTACTCGAAAACCCGCTGAGGCCGCCATGAGCCCGACGCTCCATCTGATCCTGAACATCCTGCT 630
181     A  A  T  P  A  T  E  P  V  L  E  N  P  L  R  P  P  *
                                                                1 M  S  P  T  L  H  L  I  L  N  I  L  L  13
631   CGGCATCGTCGCGGTGCTCGCCGCGCTCGCGCTGTTTTCCGGCTACGTCGCGCGCCGCGTGACGCGCGCGTTTCCGCCCGAAGGGCGCGT 720
14      G  I  V  A  V  L  A  A  L  A  L  F  S  G  Y  V  A  R  R  V  T  R  A  F  P  P  E  G  R  V  43
721   CGTCGAGGTCGGCGGCGACCGCATCCACTATGTCGAATACGGCAGCGGCACGCCGATCGTGTTCGTGCACGGGCTCGCCGGCCAGTTGCG 810
44      V  E  V  G  G  D  R  I  H  Y  V  E  Y  G  S  G  T  P  I  V  F  V  H  G  L  A  G  Q  L  R  73
811   CAACTTCGCGTATCTGCCGCTCGCGCGGCTCGCGCAGCATCACCGCGTGATTCTCGTCGACCGGCCGGGCGCCGGCCGCTCGACCCGCGG 900
74      N  F  A  Y  L  P  L  A  R  L  A  Q  H  H  R  V  I  L  V  D  R  P  G  A  G  R  S  T  R  G  103
901   CGCCGGATCGCAGGCGAACGTGTTCGCGCAGGCGCGCACGATCGCCGCGTTCATCGACGCGCTGCGGCTCGACCGCCCCGTGCTCGTCGG 990
104     A  G  S  Q  A  N  V  F  A  Q  A  R  T  I  A  A  F  I  D  A  L  R  L  D  R  P  V  L  V  G  133
991   CCATTCGCTCGGCGGCGCGATCGCGCTCGCGGTCGGGCTGAACCATCCGGAGCGCGTGAGCCGGCTCGCGCTGATCGCGCCGCTGTCGCA 1080
134     H  S  L  G  G  A  I  A  L  A  V  G  L  N  H  P  E  R  V  S  R  L  A  L  I  A  P  L  S  H  163
1081  TCCGCAGTCCGAGCCGCCCGCGCCGTTCCGGCCGCTCGTGCTGCCGTCGCCGCTCGTGCGCCGCTTCGTGTCGTGGACCTTCGCGATCCC 1170
164     P  Q  S  E  P  P  A  P  F  R  P  L  V  L  P  S  P  L  V  R  R  F  V  S  W  T  F  A  I  P  193
1171  GATGACGATCCTGACCGGCCGCCAGGCCGTGCGCCTCGTGTTCGCGCCCGAGGACGTGCCGCACGACTTCGCGGTCAAGGGCGGCGGCCT 1260
194     M  T  I  L  T  G  R  Q  A  V  R  L  V  F  A  P  E  D  V  P  H  D  F  A  V  K  G  G  G  L  123
1261  GCTCGGGCTGCGGCCTGCCAGCTTCTATGCGACCGCGACCGACCTGCTGTGCGCGCCCGACGATCTGCCCGCGATGGAAAGCCGCTACGC 1350
224     L  G  L  R  P  A  S  F  Y  A  T  A  T  D  L  L  C  A  P  D  D  L  P  A  M  E  S  R  Y  A  153
1351  GGAACTCGCGGTGCCGGTCGACGTGCTGTATGGACGCGAGGACCGGATCCTCGACTGGCGCGCCCACGGCGACGCGCTCGCGAAGAAGTC 1440
254     E  L  A  V  P  V  D  V  L  Y  G  R  E  D  R  I  L  D  W  R  A  H  G  D  A  L  A  K  K  S  183
1441  GGCGCGCGTGCGGCTGACGGTCGTCGAAGGCGGGCACATGCTGCCCGTGACGATGCCGGAAGCGACCGCCGACTGGCTGCTCGAGGTGGC 1530
284     A  R  V  R  L  T  V  V  E  G  G  H  M  L  P  V  T  M  P  E  A  T  A  D  W  L  L  E  V  A  213
1531  CGCCGCGCCGGTGACGGCGGCGGCGGCGCGCGGCGCACTGAACGGTGCGCTGCCCGGCGCACTGAGCGGCGCAGACAGCGGGCGCGGC 1620
314     A  A  P  V  T  A  A  A  A  R  A  A  H  *                                                243
1621  CTGGTAGGCCTTGTTGGCCTGTGTCGGCTTGCGTCGGCTTGCGTCGGCTTGCATTGGCTTGCGTCAAACTGCGTCGGACCGCGTCCGACT 1710
1711  CCCTCGCTCGTCGGTTGGCGATCGTGCCCAGCAGCGACGCGCGCCGTGGCGTTACGCGGCGGGCGGCCGTGGAAGCCGTGGTCGAATCAG 1800
1801  CTGCCGTCGTCCGGCAACTCGAGGNCGAATTCGGCAATCACGTCGCGCGCGCTGCGGAACGCCTCGACCGCGGCGGGCGCGCCCGCATCC 1890
1891  AGCGCGCTGTGCAGCAGCCCCCTCGCGGAATTTCGACGAGCGTCGCGCCGTTGTTCAGCGCGCCGATCGTGCCCCGGCCGTGGAAGCGCG 1980
1981  GTGGAAGACAGCGACGCGCGCCGTGGCGTCGAGGACGAATTCGGCAATCACGTCGCGCGCGCTGCGGAACGCCTCGACCGCGGCGGGCGC 2070
2071  CCGCATCC
```

Figure 4.9: The nucleotide (Black) and deduced amino acid sequences of TrkA (Red) and EstEFH5 (blue) from *B. multivorans* UWC10. Amino acids are aligned with the second nucleotide of each codon. Sequence motifs discussed in the text are underlined.

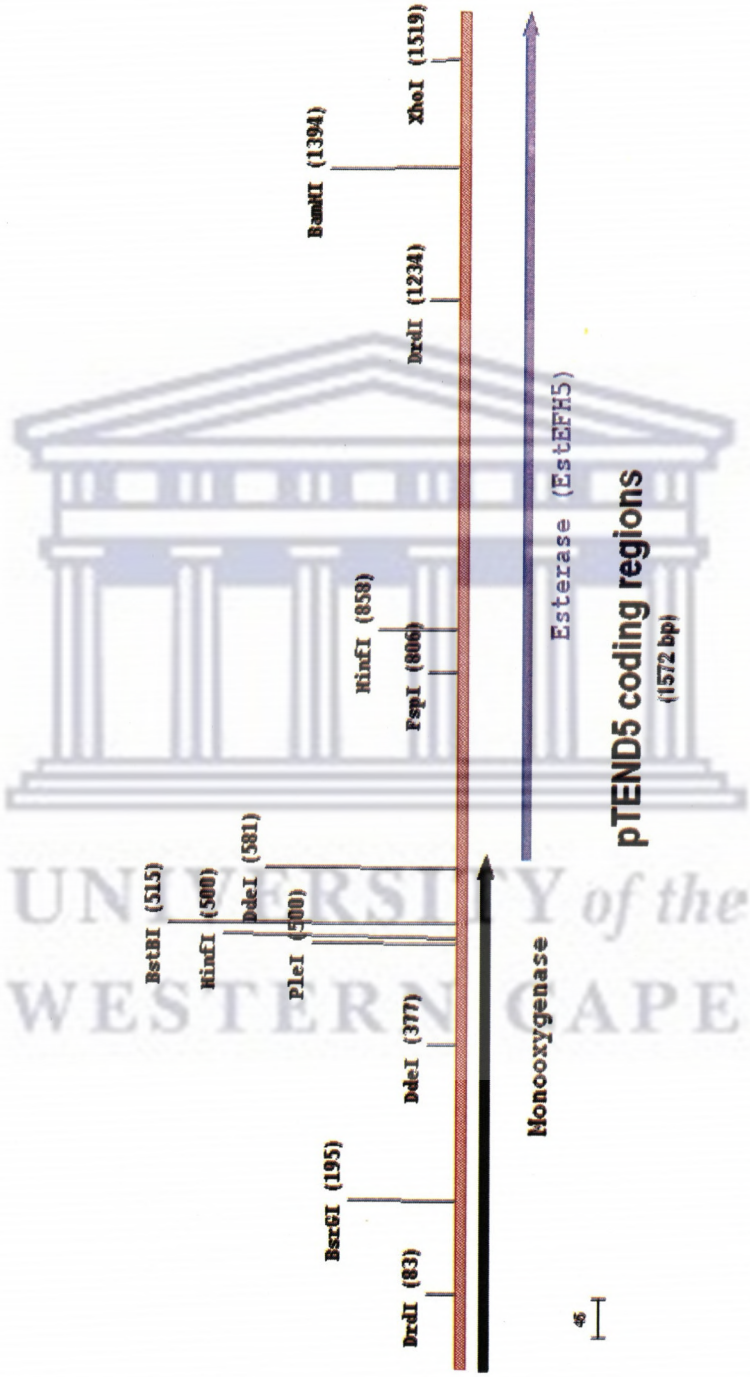


Figure 4.10: Physical map and genetic organization of the pTEND5 insert DNA. Arrows indicate the proposed direction and extension of the putative proteins.

4.4.4 Comparison of the deduced amino acid sequence of ORF2 with other esterases

The deduced amino acid sequence (326 a.a.) of the ORF2 was employed to search for homologous proteins from NCBI and EMBL protein databases. The search report revealed that the ORF2 had a strong sequence homology to esterases of prokaryotic origin. The best scores (36-61% identity) are shown in Table 4.7. ORF2 was designated EstEFH5 and the corresponding gene was designated *estEFH5*. The EstEFH5 amino acid sequence was deposited in GenBank under the accession number AAV97951.

The amino acid sequence deduced from EstEFH5 exhibited sequence identity to number of lactone hydrolyzing esterases, arylesterases and chloroperoxidases (Table 4.7). These enzymes are classified as aromatic compounds hydrolyzing enzymes based on gene oncology (Go) Annotation (http://www.ebi.ac.uk/ego/DisplayGo). Primary structure analysis revealed that, EstEFH5 had a high sequence homology (64% identical and 92.2% similar) to the esterase (SS3) from *B. pseudomallei*. Furthermore, deduced amino acid sequence comparisons showed that EstEFH5 is 61.3% identical (with 89.1% similarity), 46.0% (with 82.3% similarity), 44.3% (with 85% similarity), and 36% (with 78.8% similarity) to lactone hydrolyzing esterases from *R. mannitoliltytica* (Est1), *Ps. fluorescens* (EstF1), *Pseudomonas* sp. (EstPF5) and *Mesorhizobium loti* (EstMA), respectively (Table 4.7). EstEFH5 showed low sequence identity, 31-32 to arylesterases and 26-31 to chloroperoxidases (Table 4.6). In order to gain further information on the primary structural features of EstEFH5, multiple sequence alignments were performed with the five closely related esterase sequences (Fig. 4.11).

Table 4.7: Comparison of the amino acid sequence similarities between EstEFH5 and other related proteins

Esterase producing organisms	Accession number	Identity (%)	Similarity (%)
Lactone hydrolyzing esterases group			
Burkholderia pseudomallei (SS3)	Q63IU6	64.0	92.2
Ralstonia mannitolilytica (Est1)	Q6EIRO	61.3	89.1
Pseudomonas sp. (EstPF5)	Q9AE76	44.3	85.1
Mesorhizobium loti (EstMA)	Q98HN0	36.6	78.8
Pseudomonas fluorescens (EstF1)	O87637	46.0	82.3
Arylesterases group			
Rhodopseudomonas palustris (RPA)	Q6N9A2	32.3	69.3
Bradyrhizobium japonicum (EstBJ)	Q89D95	31.6	79.0
Peroxidases group			
Burkholderia. cepacia (PerC)	Q8VT61	27.1	65.7
Xanthomonas axonopodis (PerNH)	Q8PKX2	31.0	67.8
Agrobacterium tumefaciens (PRXC)	Q8UKN5	24.8	70.2
Methanosarcina acetivorans (PerCH)	Q8TS14	26.8	63.8
Pseudomonas syringae (ArylEst)	Q9A376	27.8	65.3

Alignment of the EstEFH5 amino acid sequence with Est1, EstF1, EstMA EstPF5 and SS3 sequences clearly demonstrates the conservation of the classical GxSxG pentapeptide signature motif (a.a. 137-141) which is shared by many esterases, lipases and other hydrolases (Fig. 4.11). From the alignment it could be deduced that Ser139, Asp219, and His279 form the catalytic triad, which corresponds to S135, His 275, and Asp215 as the putative catalytic triad in the EstEFH5 primary structure (Fig. 4.11).

The HG-dipeptide (a.a. positions 70-71) typical of the oxyanion region in esterases and lipases was also conserved. This region stabilizes the negatively charged oxygen anion of the tetrahedral intermediate during acylation or deacylation reactions. Among the eight residues of this region, three were identical and four were similar within the consensus. High sequence homologies between EstEFH5 and other esterases were noted at a.a. positions 223-229 (GGGLLG) and 299-301 (GHM). The functions of these motifs have yet to be elucidated.

The amino acid sequence of the EstEFH5 was aligned with other esterases by Neighbor-Joining distance analyses, yielding the phylogenetic tree depicted in Fig. 4.12. The phylogram indicates that EstEFH5 is clustering together with lactone hydrolyzing esterases. Application of the Arpigny and Jaeger (1999) classification scheme suggested that EstEFH belongs to Family V. Esterases in this family share significant homology to non lipolytic enzymes such as epoxide hydrolases, peptidases, dehalogenases and haloperoxidases all of which possess the typical α/β hydrolase fold.

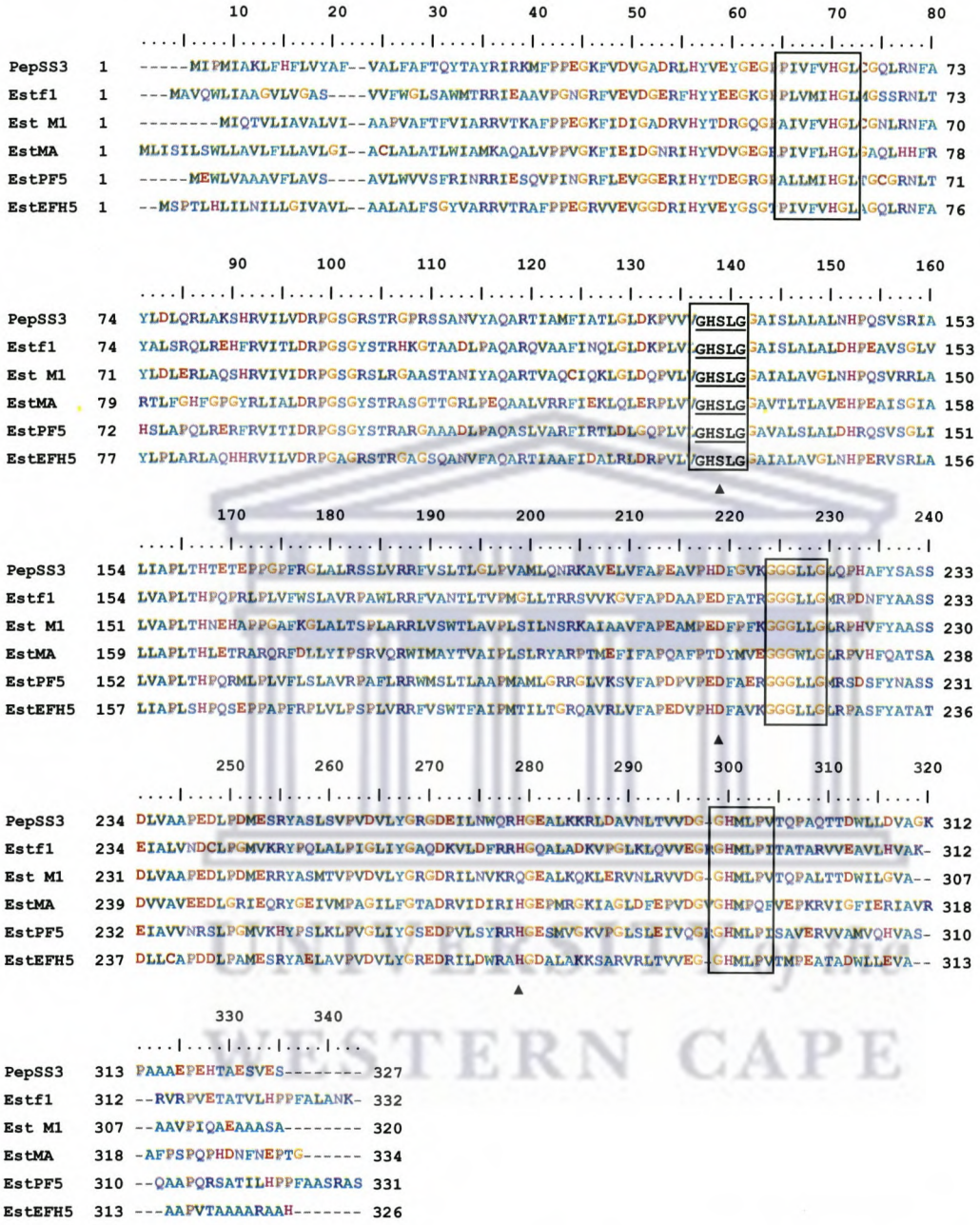


Figure 4.11: Multiple sequence alignment of *B. multivorans* esterase EstEFH5 (AAV97951) and other related proteins represented by SS3 from *B. pseudomallei* (YP11976), Estf1 *Ps. fluorescens* DMS01016 (AAC36352), Est1 from *R. mannitoliltyica* M1 (AAQ83881), EstMA from *Mesorhizobium loti* (BAB47791) and EstPF5 from *Pseudomonas* sp. (AAY92206). Catalytic triad residues are indicated by solid triangle. Motifs referred to in the text are boxed.

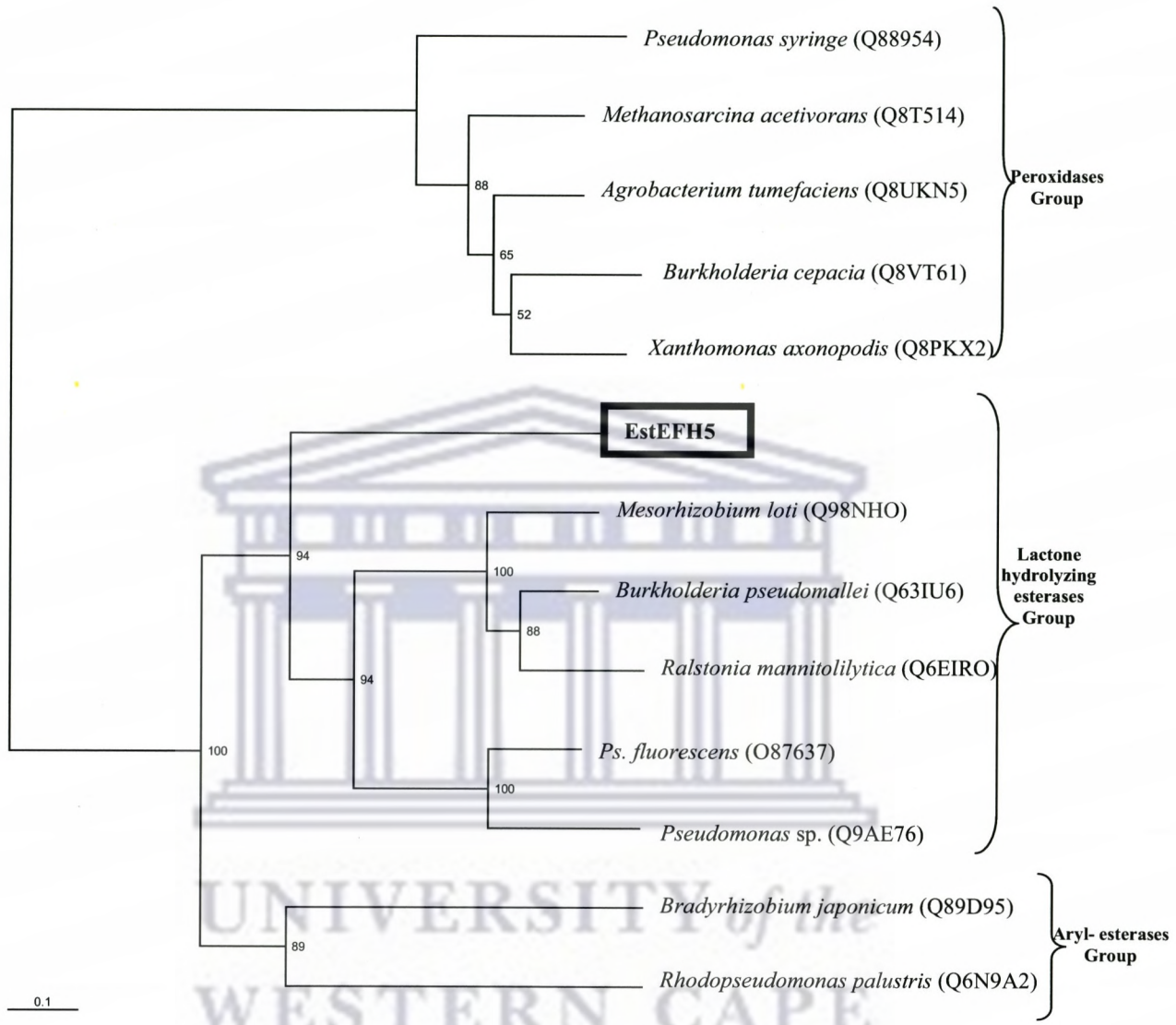


Figure 4.12: Evolutionary distance phylogram showing the position of esterase (EstEFH5) from *B. multivorans* in relation to other representative esterase members.

The phylogenetic tree is based on Neighboring-Joining analysis using PHYLOWIN. The scale represents 0.1 amino acid substitutions per position. The numbers cited show the levels of bootstrap support based on 500 resamplings. Sequences were obtained from the EMBL database.

4.5 Subcloning and Sequencing of pHOLA6

4.5.1 Subcloning of pHOLA6

A clone designated pHOLA6 that showed esterase activity during library screening was obtained. Restriction endonuclease digestion of the pHOLA6 plasmid DNA with *Hinc*II, *Hind*III and *Sma*I revealed a DNA insert of about 4.5 kb. Since this clone contained a DNA insert too large for immediate sequencing, active subclones containing smaller inserts were constructed. Restriction endonuclease digestion of the pHOLA6 plasmid DNA with *Hinc*II resulted into two distinct DNA bands of about 2.8 and 4.5 kb respectively (Fig. 4.13, lane 4). Based on plasmid restriction mapping, the top band (band A) was shown to contain a vector backbone and was directly religated, yielding a plasmid designated pHOLA6A with an insert DNA of about 1.7 kb. The bottom band (band B) was ligated into the *Hinc*II restricted site of desphosphorylated pUC18; yielding a plasmid designated pHOLA6B with a DNA insert of about 2.8 kb.

Following *E. coli* (DH5α) transformation and screening on the tributyrin agar medium, active recombinant clones were only identified with clones harbouring plasmid pHOLA6B. These clones exhibited activity in the absence of IPTG, indicating that the cloned esterase gene(s) was being transcribed from the intrinsic promoter(s). The complete 2.8 kb DNA insert was sequenced in both directions using primer-walking method together with a series of designed oligonucleotides (Table 4.8). Sequencing reactions were modified as described in Section 4.3.2.

4.5.2 Sequence analyses of cloned pHOLA6B insert DNA

To establish the existence of an esterase-encoding region within the cloned *Hinc*II DNA fragment of pHOLA6B, the complete nucleotide sequence (2830 bp) was determined (Fig. 4.14). Translational analysis of the nucleotide sequence of the *Hinc*II DNA fragment revealed two complete ORFs, in the same orientation. ORF1, consisting of 1194 bp (nt. positions 151-1344), commenced with ATG start and terminating with TGA stop codon. This ORF1 encoded a protein of 398 a.a. with an estimated molecular weight and pI values of 41,895 Da and 5.54, respectively. The possible RBS site -GAGG- was located 8 bp upstream of the ATG start codon.

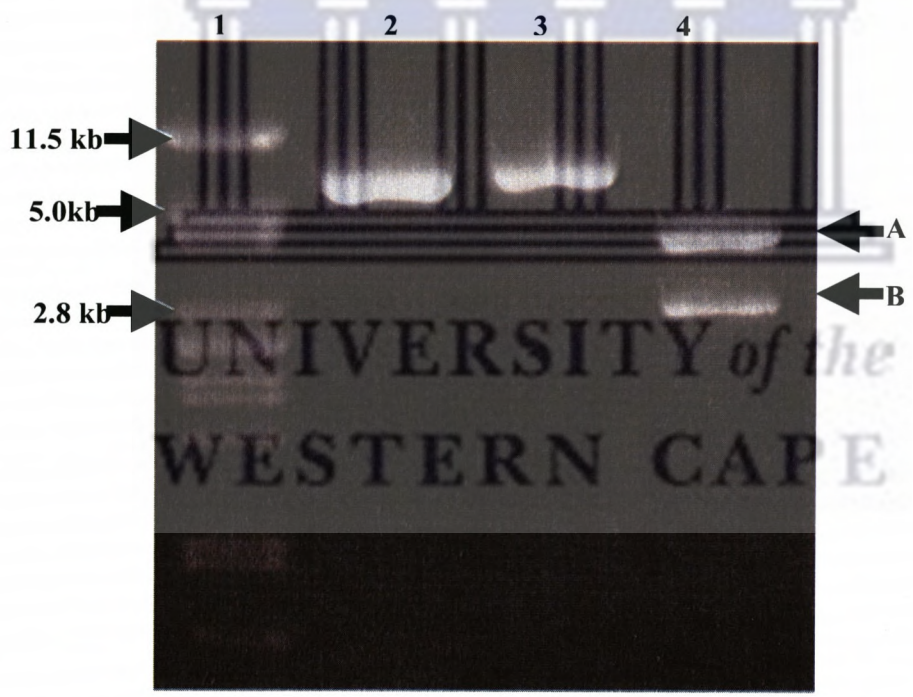


Figure 4.13: Agarose gel (1%) gel electrophoregram showing restriction endonuclease digestions of the pHOLA6 plasmid DNA. *Pst*I-digested λDNA markers; Lane 2: pHOLA6B digested with *Hind*III; Lane 3: pHOLAB digested with *Sma*I; Lane 4: pHOLAB digested with *Hinc*II.

Table 4.8: Primers used to sequence the complete pHOLA6B DNA insert

Primer name	5' →3'	Position	References
M13F/pUC18	GTAAAACGACGGCCAGT	380-397 [a]	Yanisch-Perron *et al.*,1985
M13R/pUC18	CAGTATCGACAAAGGAC	464-481 [a]	Yanisch-Perron *et al.*,1985
EstF1	CATCACGTTGCGGCATCTGCTGTCG	501-525 [b]	This study
EstF2	ATTGCACGTCGATCGTACGAACTG	1350-1374	This study
EstF3	TGCTGACCGTGATGTTTGCGTACG	2017-2041 [b]	This study
EstF4	ATGCTTGTGACGGTGTGGAATCTCG	2377-2401 [b]	This study
EstR1	CGAACACGAGCAGCACCGCATCGACCTGC	2106-2134 [b]	This study
EstR2	CCTCGTACAGCGTGTTGGACAG	1255-1276 [b]	This study
EstR3	CGACCACGTCGATCGCGAGCGAATAG	690-716 [b]	This study

[a]Position in relationship to pUC18 (Accession number L09136)

[b]Position in relation to sequenced pHOLA6B DNA insert

Downstream of ORF1 was a second complete open reading frame (ORF2) consisting of 1059 bp (nt. positions 1492-2830). ORF2, encoded a protein of 353 a.a. with predicted molecular weight and pI values of 35,840 Da and 11.38, respectively. A putative RBS site -GAAG- was located 26 bp upstream from presumptive ATG start codon of the ORF2. The DNA sequences of the two open reading frames, ORF1 and ORF2, had high GC content, 73% and 72%, respectively.

4.5.3 Analysis of the deduced amino acid sequences of ORF1 and ORF2

Analysis of the deduced amino acid sequence of ORF1 revealed two motifs; GVSDG (a.a. positions 147-151) and GTSGG (a.a. positions 176-180), which correspond well to the classical pentapeptide GxSxG motif. The motif SVTK (a.a. positions 74-77), corresponding to the SxxK motif, which is conserved in β-lactamases, peptidases, and esterases of family VIII, was also observed. The SVTK motif was shown to harbour catalytic serine as demonstrated by mutational studies with esterases from *B. gladioli* (Petersen *et al.*, 2001) and *Brevi. linens* (Sakai *et al.*, 1999). The serine within GxSxG motifs was shown not to be essential for catalysis in these esterases.

The deduced primary structure of the ORF1 revealed significant high homology to Family VIII esterases, class C β-lactamases and peptidases (Table 4.9). Based on high sequence scores (26-68% identity) to a number of esterases, the deduced protein was designated EstBL and the encoding gene was designated *estBL*. The physical map of the pHOLA6B coding region is shown in Fig. 4.15. Due to incorrect annotation by genome sequencing programs and the high homology of these esterases with the N-terminus of β-lactamases, the CD-search protein domain annotations (Marchler-Bauer and Bryant, 2004), have incorrectly annotated most of the Family VIII esterases as β-lactamases (Appendix IV).

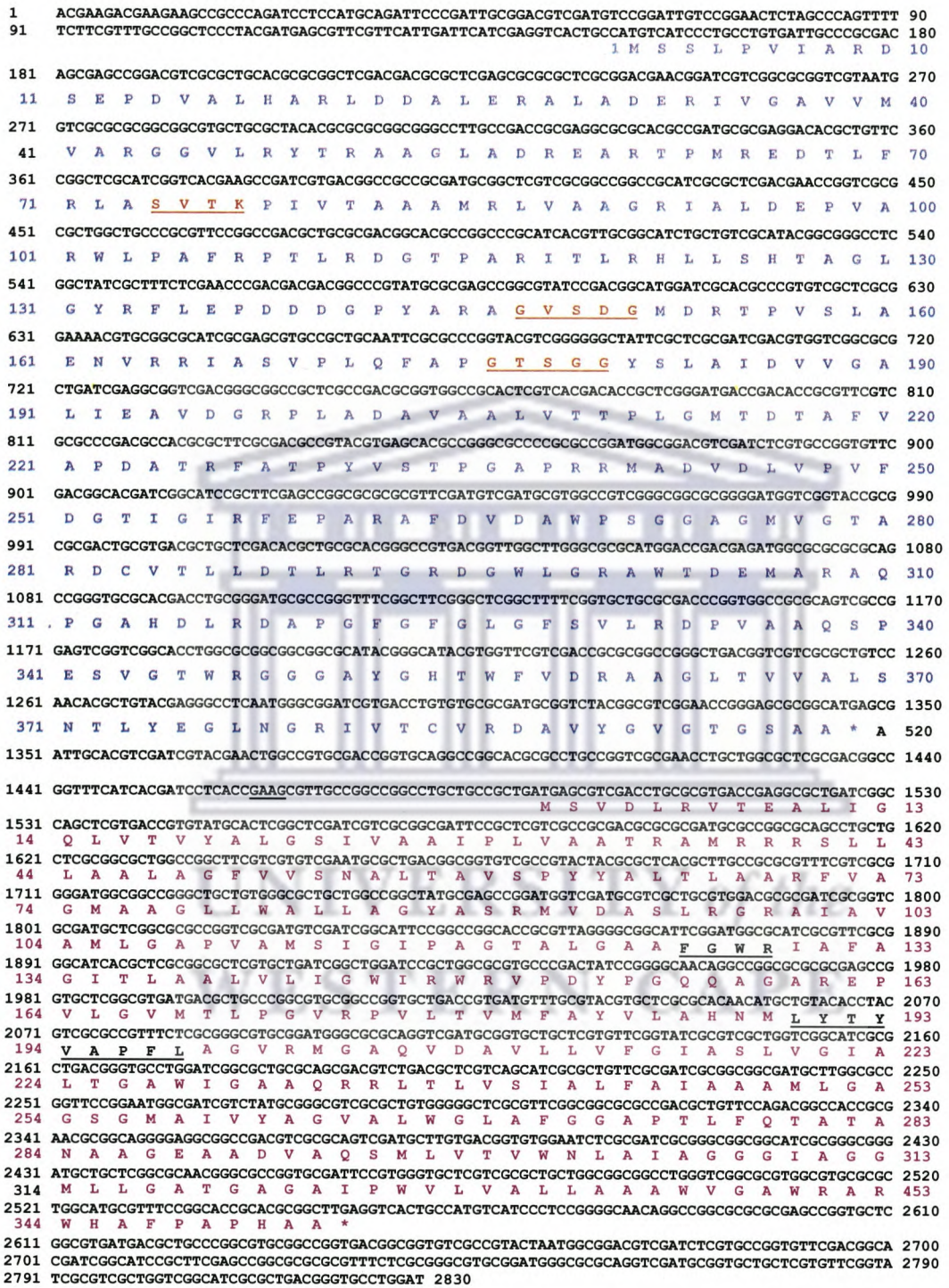


Figure 4.14: The nucleotide (Black) and deduced amino acid sequences of EstBL (Blue) and Aep (pink) from *B. multivorans* UWC10. Amino acids are aligned with the second nucleotide of each codon. Sequence motifs discussed in the text are underlined.

The deduced 353 amino acid sequence of ORF2 was used to scan for homologous protein in the protein databases. The CD-search protein domain annotations from the NCBI server indicated that the primary structure of the ORF2 was related to the arabinose efflux permeases (AraJ) (Appendix IV). The FGWR (a.a. position 126-129) motif which is conserved in these enzymes was observed. A second motif (LYTYVRPFL; a.a. position 190-198) which corresponds to the highly conserved LYTYXZPFL motif (where x is the hydrophobic amino acid residue and Z represents any amino acids) within the permease family was also observed. Amino acid sequence comparisons showed three arabinose efflux permease (AraJ) enzymes which exhibited considerable sequence homology to ORF2; AraJ from *Ps. fluorescens* (59% identity, 75% similarity; Accession number: ZP_00263612), AraJ from *Ps. aeruginosa* (36% identity, 52% similarity; Accession number: ZP_00136659) and AraJ from *Rhodospillum rubrum* (30% identity; 50% similarity; Accession number ZP_00270152).

Analysis of the multiple sequence alignments of EstBL and other related esterase revealed a number of conserved motifs (Block A-G) as depicted in Fig. 4.16. The SxxK motif which is believed to harbour the catalytic serine (Ser75) was found to be highly conserved. A consensus motif (KTG) which is normally located at the C-terminus of all known β-lactamases was not observed (Knox *et al.*, 1996). The amino acid sequence motif corresponding to the classical signature pentapeptide (GxSxG) motif at a.a. positions 149-153 was much conserved throughout the aligned esterases. A second GTSGG motif (a.a. position 179-183) within the EstBL primary structure which corresponds to the pentapeptide GxSxG motif was not conserved in any of the aligned

esterases sequences. The functions of the other conserved motifs identified (Blocks B, D, E, F, G) are as yet unknown.

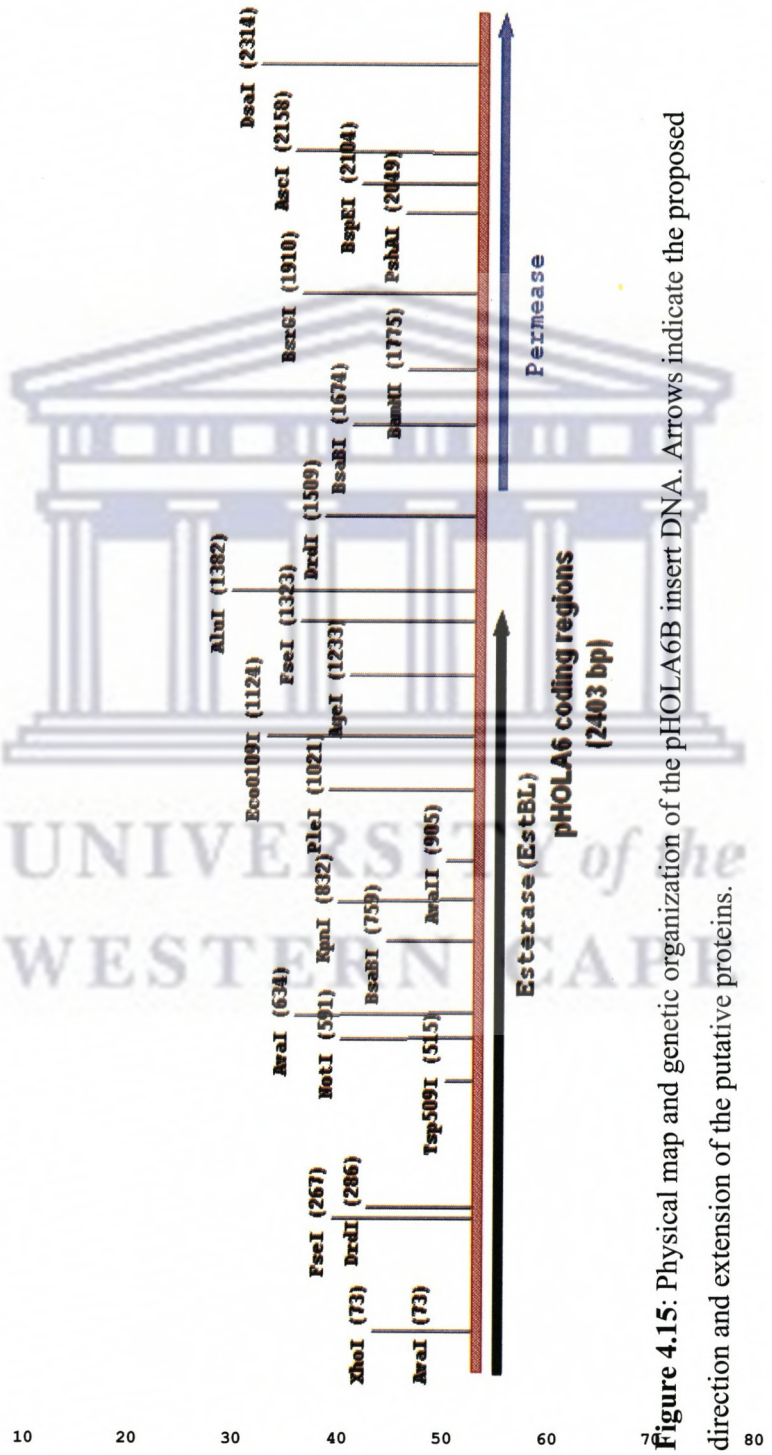


Figure 4.15: Physical map and genetic organization of the pHOLA6B insert DNA. Arrows indicate the proposed direction and extension of the putative proteins.

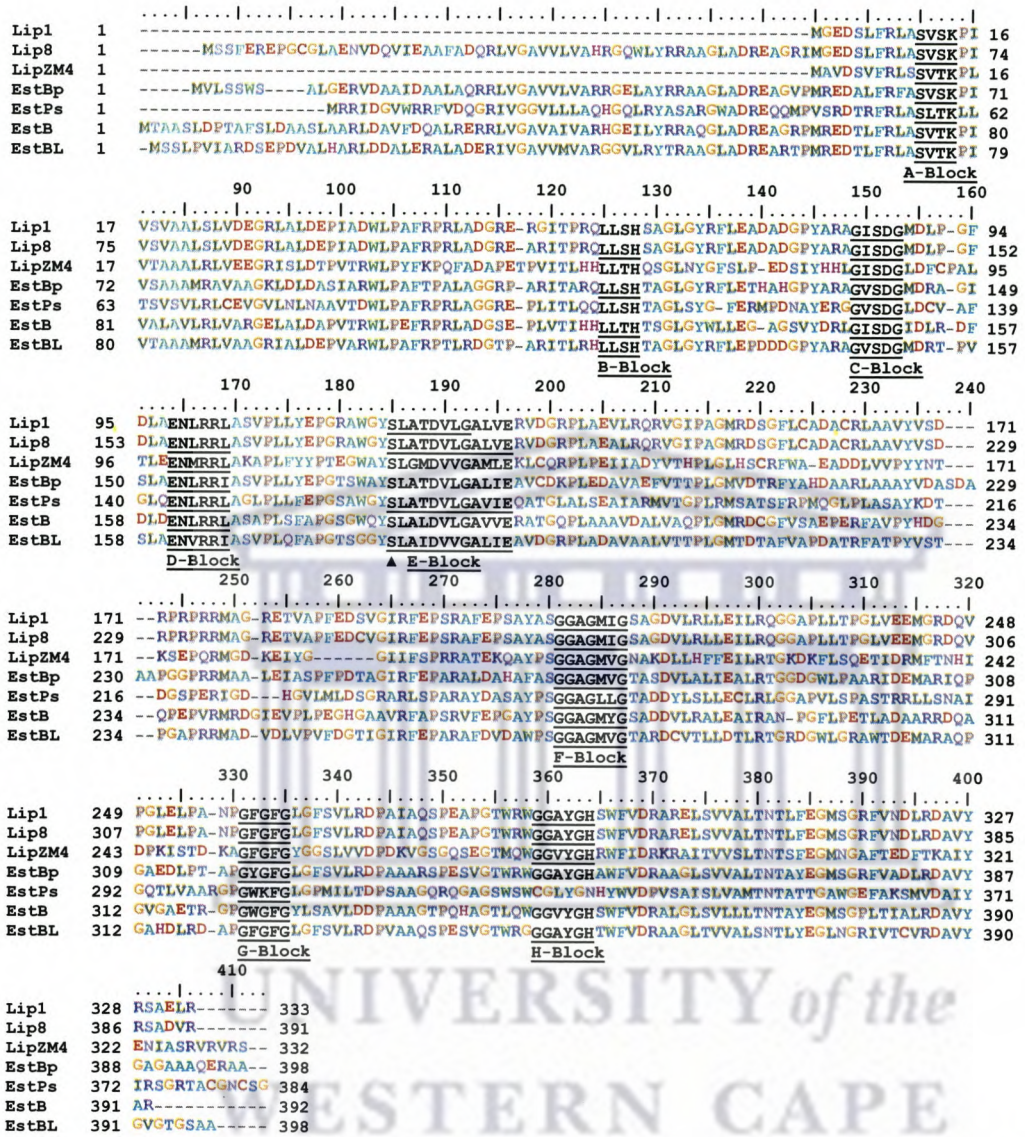


Figure 4.16: Multiple sequence alignment of *B. multivorans* esterase EstBL (AAV97951) and other related proteins represented by Lip1 from *Pseudomonas* sp. nov. 109 (CAA43847), Lip8 from *Ps. aeruginosa* LST-03 (BAD69792), LipZM4 from *Zygomonas mobilis* ZM4 (AAG42401), EstBp from *B. mallie* (AAU49131), EstPs from *Ps. Syringae* (AA056448) and EstB from *B. gladioli* (AAF59826). Motifs discussed in the text are indicated as underlined blocks.

4.5.4 Comparison of the deduced amino acids sequences of EstBL with other related enzymes

In order to gain an understanding on the possible evolutionary relationship of the EstBL, a phylogenetic analysis based on the primary structures of family VIII esterases and other related enzymes peptidases and β-lactamases was performed (Table 4.9). As expected, EstBL from *B. multivorans* UWC10 displayed high identities (51.9-68.6%) to other members of *Burkholderia* esterases. The homology between the esterase EstBL from *B. multivorans* and the esterases from genus *Pseudomonas* esterases varied from 26.9% to 42.4%. EstBL exhibited low sequence identity (24%) to β-lactamases from *Enterobacter* (D4479) and *E. coli* (X74512). Sequence identity of EstBL with peptidases from *B. cereus* (CAA09676) and *Streptomyces* R61 (P15555) was between 26-29%. Comparison of the EstBL amino acid sequence with corresponding sequences from other related primary structures was performed using Neighbor-Joining analysis yielding the phylogenetic tree depicted in Fig. 4.17. The phylogram of EstBL with the related esterases shows that EstBL clusters independently from peptidases and β-lactamases. Furthermore, EstBL was also shown to cluster independently from the *Burkholderia* and *Pseudomonas* esterase groups, regardless of high sequence identity, 51-68% and 60-69%, respectively.

Table: 4.9: Comparisons of *B. multivorans* esterase (EstBL) sequence with related proteins.

Esterase producing organisms	Accession number	Identity (%)	Similarity (%)
Esterases group			
Pseudomonas sp. (EstA)	A44832	27.92	71.47
Burkholderia gladioli (EstB)	AAF59826	51.93	86.38
Burkholderia mallei (EstBp)	AAU49131	68.64	93.06
Pseudomonas fluorescens (EstC)	AAC60471	26.33	70.31
Pseudomonas syringae (EstPs)	AA056448	42.44	82.49
Lipolytic group			
Pseudomonas sp. 109(Lip1)	CAA43847	60.36	90.36
Pseudomonas aeruginosa LST-03 (Lip8)	BAD69792	69.10	90.02
Zygomonas mobilis ZM4 (LipZM4)	AAG42401	46.77	80.77
Peptidases group			
Bacillus cereus (Adp)	CAA09676	29.25	71.94
Streptomyces R61 (Ddcp)	P15555	27.45	71.15
B-Lactamases group			
Enterobacter cloacae (EntBL)	D4479	24.93	63.61
E. coli (EcoliL)	X74512	24.36	65.33

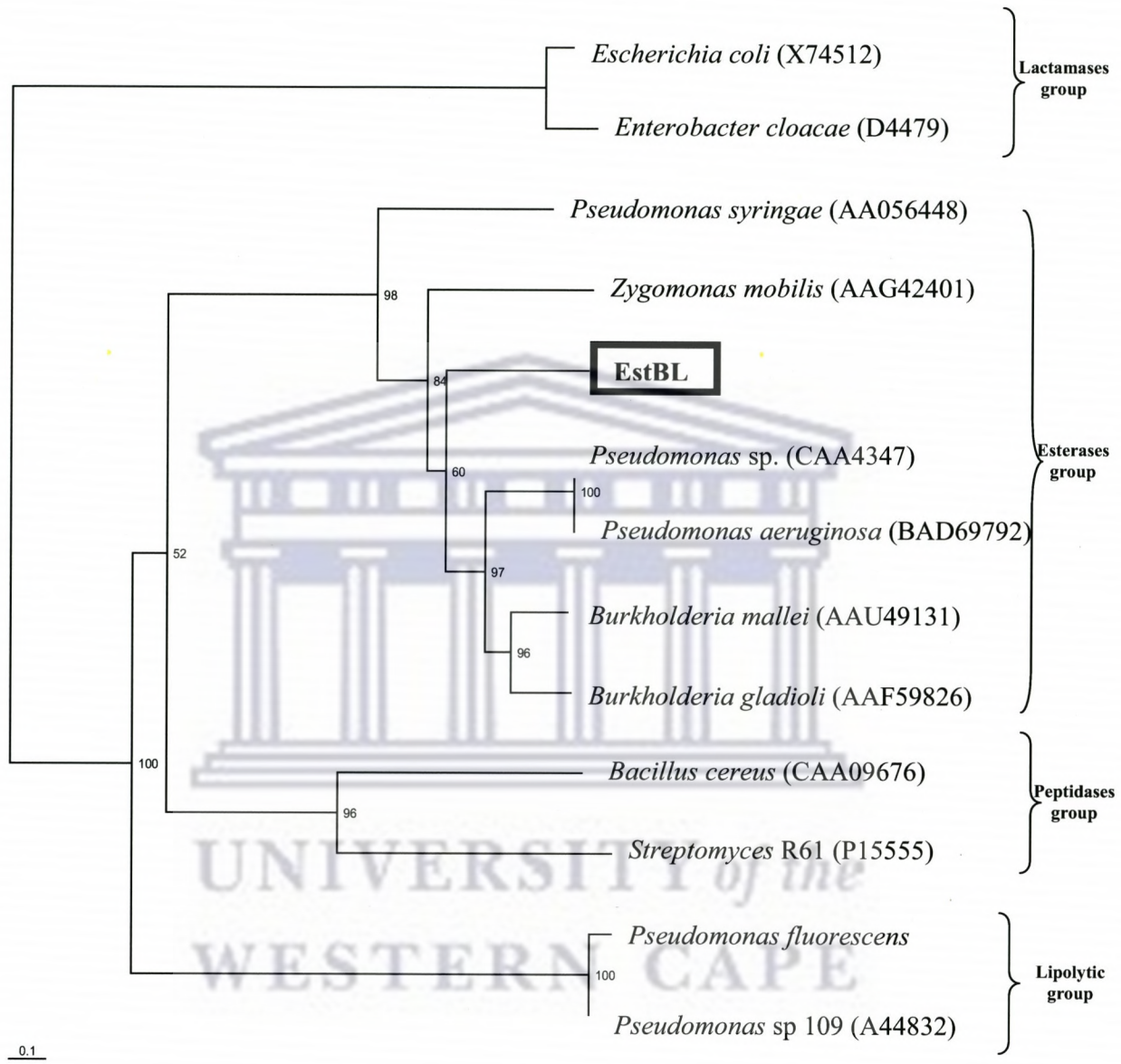


Figure 4.17: Evolutionary distance phylogram showing the position of esterase (EstBL) from *B. multivorans* in relation to other esterase, β-lactamases and peptidase representative members.

The phylogenetic tree is based on Neighbor-Joining analysis using PHYLOWIN package. The scale represents 0.1amino acid substitution per position. The numbers on the tree show the levels of bootstrap support based on 500 resamplings. Sequences were obtained from the GenBank.

4.6 Discussion

This chapter describes the cloning and nucleotide sequencing of various *B. multivorans* UWC10 genes encoding lipase, esterase and other putative activities as summarized in Table 4.10. Functional screening indicated that a clone harbouring pRASH14 encoded a lipase-like protein while clones pTEND5 and pHOLA6 both encoded esterolytic activities. The deduced primary structures of all three proteins (LipA, EstEFH5 and EstBL) showed a pentapetide GxSxG motif (GHSQH, GHSLG and GVSDH in LipA, EstEFH5 and EstBL, respectively). The catalytic serine is generally believed to be located within such a pentapeptide motifs (Arpingy and Jaeger, 1999; Bornscheuer, 2002). However, the EstBL primary structure also contained the SVTK motif, the serine of which has been implicated in catalytic activity (Sakai *et al.*, 1999; Petersen *et al.*, 2001).

Table 4.10: Properties of the three recombinant clones

Clone	Enzyme	Activity[a]	Family[b]	Motifs	Homology	Size[c] (kDa)
pRASH14	LipA	Lipolytic	I	GHSQG	Group 1 Proteobacterial lipases	33
pTEND5	EstEFH5	Esterolytic	V	GHSLG	Lactone hydrolyzing esterases; chloroperoxidases	35
pHOLA6	EstBL	Esterolytic	VIII	GVSDG SVTK	Class C β-lactamases; Peptidases, Family VIII esterases	42

([a]= Based on qualitative agar plate assays; [b]=classification based on Arpigny and Jaeger, (1999); [c]=deduced from the amino acid sequence)

All lipases and esterases can be divided into two evolutionarily distinct classes on the basis of the codon for the active site serine, which can either be AGY or TCN, (where N is any nucleotide and Y is C or T, (Petersen and Drablos, 1994)). The TCN codon is found in many esterases while the AGY codon if found in many lipases. The LipA putative catalytic serine (Ser131) was encoded for by AGC codon, consistent with the early classification scheme of bacterial lipases based on the serine codon (Jaeger *et al.*, 1994). The putative catalytic serine (Ser135) within the EstEFH5 primary structure was encoded by a TCG codon consistent with the suggestion that EstEFH5 is an esterase. Both possible catalytic serine residues (Ser74) and (Ser149) within the EstBL primary structure were encoded for by the TCG codon.

4.6.1 Significance of the leader sequence

Primary structure analysis of LipA indicated the presence of a putative 40 a.a. N-terminal leader peptide characterized by a positively charged N-terminus, a hydrophobic core region and a polar C-terminal region. This observation suggests that LipA is an extracellular enzyme, which is directly, exported through the membrane with the aid of the leader sequence. Literature searches revealed that lipases from Group I proteobacterial lipases such as lipases from *Pseudomonas* (Chihara-Siomi *et al.*, 1992), *Burkholderia* (Jorgensen *et al.*, 1991; Frenken *et al.*, 1992), *Acinetobacter* (Kok *et al.*, 1995; Sullivan *et al.*, 1999) and *Vibrio* (Ogierman *et al.*, 1996) possess the N-terminal signal sequence which serves as a potential substrate for Sec-homologous protein export (Rosenau and Jaeger, 2000).

4.6.2 A bicistronic lipase operon

An analysis of the deduced amino acid sequence of pRASH14B identified a second ORF immediately downstream of the lipase structural gene. A detailed analysis of the 50 bp intergenic region between the translational stop codon of the *lipA* gene and the downstream translation initiation codon of *lipB* revealed the presence of both direct and indirect repetitive sequences. These repetitive sequences could potentially form mRNA with characteristic secondary structural features (e.g., hairpin loops) which may function as transcriptional terminator or processing sites for mRNA degradative enzymes during the expression of the lipase operon. The organization of the LipA and LipB genes, the absence of putative Shine-Dalgano sequence upstream of the LipB gene and the presence of potential hairpin structures just downstream of the LipA gene suggest that the two genes are possibly transcribed as a single bicistronic operon. Although the intergenic regions between lipase and lipase-chaperone genes vary from 3 to 83 bp in length, similar transcriptional organization is used in the lipase operons from *Ps. fluorescens* (Chihara-Siomi *et al.*, 1992), *B. cepacia* (Jorgensen *et al.*, 1991), *B. glumae* (Frenken *et al.*, 1992), *P. vulgaris* (Kim *et al.*, 1996) *V. harveyi* (Teo *et al.*, 2003) and *R. solanacearum* (Ouyen *et al.*, 2004). Although similar transcriptional organization is found in the lipase operons of *Acinetobacter* (Sullivan *et al.*, 1995), these operons differ in that the structural gene is located downstream of the lipase chaperone gene.

4.6.3 Significance of lipase chaperone (LipB)

The second open reading frame of the cloned lipase operon codes for a protein designated LipB, with significant sequence similarity with a class of proteins known as lipase chaperones or lipase specific foldases (Rosenau *et al.*, 2004). The secretion of lipases from *Burkholderia* and other Group I proteobacterial lipases is typically via type II secretion pathway (Rosenau and Jaeger, 2000). Through this mechanism a lipase is firstly transported across the cytoplasmic membrane via the Sec system (which uses a leader peptide as a substrate), then folded in the periplasm and into the native conformation and is transported as a mature protein across the outer membrane via the Xcp system. The folding of lipase depends on general periplasmic folding catalysts such as disulphide isomerase DsbA and lipase specific foldase (Rosenau and Jaeger, 2000).

The role of lipase chaperones is well studied in *B. glumae* (El Khattabi *et al.*, 2000). It is believed that the lipase chaperone is anchored to the inner membrane by an N-terminal hydrophobic segment, with the large C-terminal domain exposed to the periplasm. Molecular mechanism studies of the lipase chaperone indicate that these proteins are steric chaperone that act by providing essential steric information to their cognate lipases rather than by preventing an off-pathway reaction such as aggregation (Shibata *et al* 1998). It is therefore postulated that LipB accelerates the rate-limiting step by lowering energy barrier along the LipA folding pathway.

4.6.4 EstEFH5 is related to lactones hydrolyzing Esterase

An incomplete ORF1 from the cloned DNA insert of plasmid construct pTEND5 encoded a polypeptide with significant homology (51-60% amino acid sequence identity) to monoxygenases from number of organisms including *B. pseudomallei Caulobacter crescentus*, *R. metallidurans*, *Bradyrhizobium japonicum*. A second complete open reading frame (ORF2) found in this plasmid encoded a protein designated EstEFH5, with significant homology (36-61% amino acid sequence identity) to lactone hydrolyzing esterases. Analysis of the EstEFH5 primary structure showed the presence of a GxSxG motif and residues consistent with a catalytic triad. Unique feature of EstEFH (and related lactone hydrolyzing esterases) include the presence of a 30 a.a. N-terminal extension and a very high pI value (10-11). Hydropathy plots (Fig. 4.18) indicate that this N-terminal region was highly hydrophobic, suggesting that this sequence may anchor the enzyme in the cell membrane. It has been noted that lactone hydrolyzing enzymes possesses the sequence typically requiring detergents to maintain the protein in a soluble state during expression and/or purification (Khalameyzer *et al.*, 1999).

The importance of ORF1 encoding monooxygenase upstream the EstEFH5 was also noted. The location of monooxygenase genes adjacent to the lactone hydrolyzing esterase genes have reported from previous studies. These include the Est1 gene from *Ps. fluorescens* located upstream the monooxygenase gene (Khalameyzer *et al.*, 1999) while in *Acinetobacter* sp. the esterase gene is located downstream the monooxygenase gene (Taschner *et al.,* 1988). A mechanistic model has been proposed, which suggests that

esterases located adjacent to monooxygenases are involved in the hydrolysis of lactones formed by Baeyer-Villiger monoxygenases (Onakunle *et al.*, 1997).

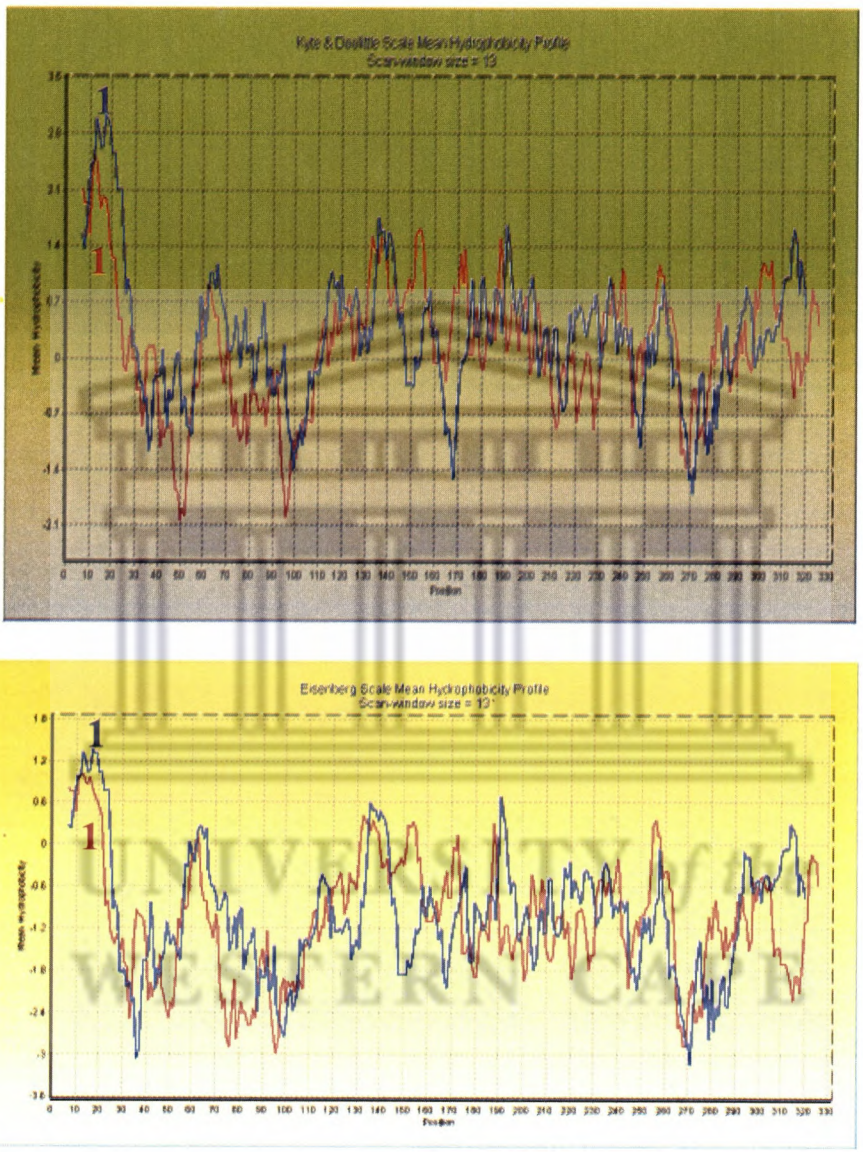


Figure 4.18: Comparison of (A) Kyte and Doolittle and (B) Eisenberg hydrophobicity plots of EstEFH5 (blue) and EstF1 (red) from *Ps. fluorescens* Accession number AAC36352 (Khalameyzer *et al.*, 1999). Putative membrane anchor domains are numbered for both proteins.

The primary structure of EstEFH5 contains a GGGL sequence, which corresponds to GGGX (X-denoting hydrophobic residue) motif. Henke and co-workers have shown that the presence of the GGGX motif is linked with the specificity for tertiary alcohols (Henke *et al.*, 2002). All enzymes bearing this motif including pig liver esterase, several acetyl choline esterases and an esterase from *B. subtilis* were found active towards acetates of tertiary alcohols, while enzymes bearing the more common GX motif did not catalyze the model substrates (Henke *et al.*, 2002). It is noted however that the GGGX motif in these esterases was located adjacent to the oxyanion region unlike in EstEFH5, where this motif was located towards the C-terminus of the protein. While the significance of the locality of the GGGX motif is unclear, the ability of these enzymes to accept bulks substrates is consistent with the activity of EstEFH towards ethyl ferulate.

4.6.5 EstBL contains β-lactamase fold

Primary structure analysis of EstBL revealed two potential serine hydrolase motifs (G-V-**S_{149}**-D-G) and a motif for class C β-lactatames (F-R-L-A-**S_{74}**-R-L-A-S-V-T-K-P-I-V-T-A-A-A-M-R-L) that are presumed to represent the active site residues. However, in the case of the (G-V-**S_{149}**-D-G) motif, sequences upstream this motif do not match with the lipase or esterase signatures (such as the oxyanion hole region) present in the PROSITE motif databases (http://www.expasy.ch). Comparison of the amino acids of EstBL by BLAST search showed significant homologies to penicillin-binding proteins, β-lactamases and DD-peptidases and esterases. Multiple sequence alignments showed a number of other conserved regions including the G-(V/I)-S-D-G) and SxxK motif. The region around the second putative GxSxG motif, represented by G-T-**S_{215}**-G-G, shows no

conservation and is therefore unlikely to harbour the catalytic serine. The KTG motif, which is highly conserved at the C-terminus of the β-lactamases, was not conserved within the primary structure of aligned esterases (Knox *et al.*, 1996). Functional characterization of the well studied EstB from *B. gladioli* showed no β-lactamase activity towards a variety of the β-lactam substrates (Petersen *et al* 2001). Site directed mutagenesis of EstB catalytic serine residue Ser75 to Ala75 indicated that the serine residue within the β-lactamase motif is essential for esterase activity (Petersen *et al.*, 2001).

The identity of residues which might act as a general base in the catalytic mechanism of esterases/ β-lactamases is still a matter of debate. Crystallographic studies have revealed that a lysine adjacent to the catalytic serine in β-lactamases plays a role analogous to histidine in esterases and other many hydrolases (Damblon *et al.*, 1996). To function as a base, the pK_a of the lysine ε-amino group would have to be reduced drastically since the environment must be able to stabilize the energetically unfavourable deprotonated form. This is believed to occur as a result of substrate binding (Swaren *et al.*, 1995). Although, the experimental evidence in support of this speculation is unconvincing (Swaren *et al.*, 1995), the mutational studies have indicated that the replacement of Glu66 results in long-lived acyl-intermediates (Banerjee *et al.*, 1998). It is generally believed that various amino acids could act as general bases for the acylation and deacylation steps, possibly with the involvement of a conserved water molecule (Lamotte *et al.*, 1999). EstB from *B. gladioli* is the only crystallographic structure from the Family VIII esterases that has been

solved to date (Wagner *et al.*, 2001), but this study did not confirm unequivocally which of the residues serves as a base.

CHAPTER FIVE

Heterologous expression, Purification and Characterization

5.1 Introduction

A reliable method for confirming that a gene encodes a particular protein is by heterologous expression in a suitable host. Amongst the available expression host systems, *E. coli* remains the most commonly used. This is largely due to its rapid and reliable growth, well-characterized genetics and wide range of vectors available (Baneys *et al*, 1999). However, expression in *E. coli* does not guarantee that a protein will correctly fold or will accumulate in significant amount, or will be functional even if present in high levels. Expression vector systems that allow the expression of the protein of interest as a fusion partner to improve both solubility and purification have been developed. Some of the more common fusion partners include maltose-binding protein (MBP), thioredoxin, glutathione-S-transferase (GST) and pectate lyase (pelB) (LaVallie and McCoy, 1995).

Attempts to express proteins from the genus *Burkholderia* in a soluble biologically active state have been marked with both success and failure. For instance, genes encoding esterases from *B. gladioli* (Reiter *et al*., 2000, Petersen *et al*., 2001) and *B. cepacia* (Kim *et al*., 2004a) have been successfully expressed in *E. coli*, while the expression of lipase genes from this genus has so far failed (Jorgensen *et al*., 1991; Quyen *et al*., 1999). This chapter reports the design and use of expression constructs for over-expression of cloned

B. multivorans lipase and esterase genes in *E. coli*, and the purification and characterization of the recombinant enzymes.

5.2 Expression of EstBL

5.2.1 Expression strategy

The recent expression of esterase genes from *B. gladioli* (Petersen *et al.*, 2001) and *Ps. aeruginosa* (Ogino *et al.*, 2004) which share significant homologies to the EstBL gene revealed that heterologous expression of these genes did not require any flanking sequences or genes. Furthermore, the absence of the N-terminal leader peptide within the primary structure of EstBL suggested that EstBL could be expressed as a full-length protein. Based on this information, it was thus assumed that the expression of the EstBL gene in *E. coli* would not require any flanking genes.

The expression strategy employed involved the PCR amplification of the EstBL structural gene using restriction sites engineered in the primers, and subsequent cloning into pMOS*Blue* vector. The EstBL structural gene could therefore be excised by means of these restriction sites before insertion into the corresponding sites of the pMS470Δ8 expression vector. The pMS470Δ8 is a *tac*-promoter based *E. coli* expression vector system with the Shine-Dalgano sequence of T7 gene 10 (Balzer *et al.*, 1992). Expression plasmid pMSEstBL was constructed by replacing the *Nde*I/*Hind*III stuffer–fragment with the EstBL-encoding fragment. In this construct, the EstBL-coding fragment was fused at the start ATG codon to the T7 gene 10 translation initiation signal of the pMS470Δ8 vector.

5.2.2 Construction of EstBL expression vector

In order to design an EstBL expression vector, the EstBL structural gene was amplified using a set of engineered restriction site-containing Beta(F)/Beta(R) PCR primers (Fig. 5.1). Based on the engineered restriction sites, the gene was expected to be inserted in the correct orientation and reading frame just downstream of the ribosome binding site (RBS) of the pMS470Δ8 expression vector. The ATG start codon of the EstBL gene was used to create an *Nde*I restriction site for the 5′ (sense strand) insertion point while *Hind*III restriction site was added to the end of the 3′ primer (anti-sense strand). Restriction sites and stop codon were included in such a manner as to avoid the addition of C or N-terminal tags to the expressed protein.

A

B

Beta (F): 5′-GCCCATATGTCATCCCTGCCTGTGATTG-3′ (*Nde*I)

Beta(R): 5′-ATTAAGCTTTCATGCCGCGCTCCCGGTTCC-3′ (*Hind*III)

Figure 5.1: (A) Layout of the pHOLA6B coding regions of the *B. multivorans* UWC10 and the organization of the primer pair represented by red and pink arrows (b) The nucleotide sequences of the Beta (F) and Beta (R) primers and the engineered restriction sites are underlined.

A DNA fragment containing the EstBL gene was PCR-amplified with *Pfu* DNA polymerase using Beta (F) and Beta (R) as primers and pHOLA6B as a template. An annealing temperature of 55 °C was used: all other conditions were as previously described (Chapter 2: Section 2.10). The expected PCR product of approximately 1180 bp was obtained (Fig. 5.2). Following gel purification and blunt-end cloning into the pMOS*Blue* vector, plasmid construct pMOSEstBL was obtained. The insert was excised by means of the engineered restriction sites and cloned into the corresponding sites of the expression vector pMS470Δ8, yielding an expression construct pMSESTBL. The expression construct pMSESTBL was used to transform electrocompetent *E. coli* (BL21) cells to yield recombinant clones of *E. coli* (BL21)/pMSESTBL expression host. An overview of the EstBL gene subcloning procedures is shown in Fig. 5.3.

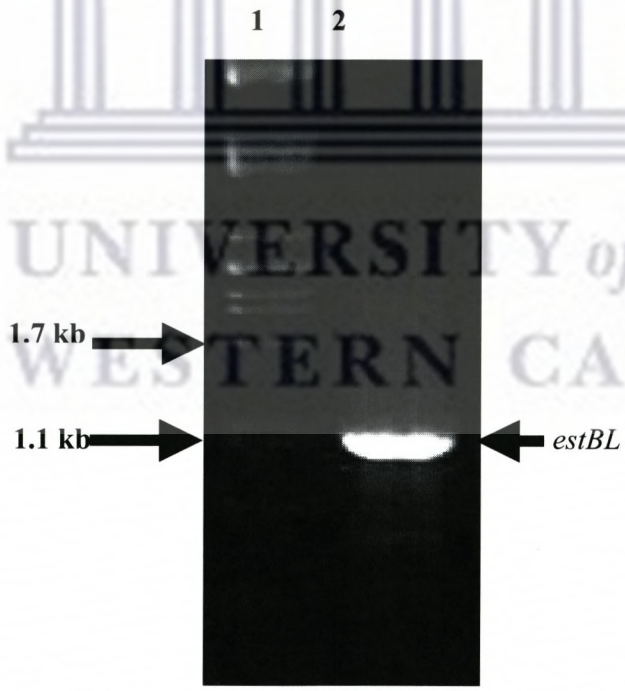


Figure 5.2: Agarose (1%) gel electrophoregram showing PCR amplification of the *estBL* encoding DNA fragment: Lane 1: *Pst*I-digested λDNA markers; Lane 2: *estBL* gene.

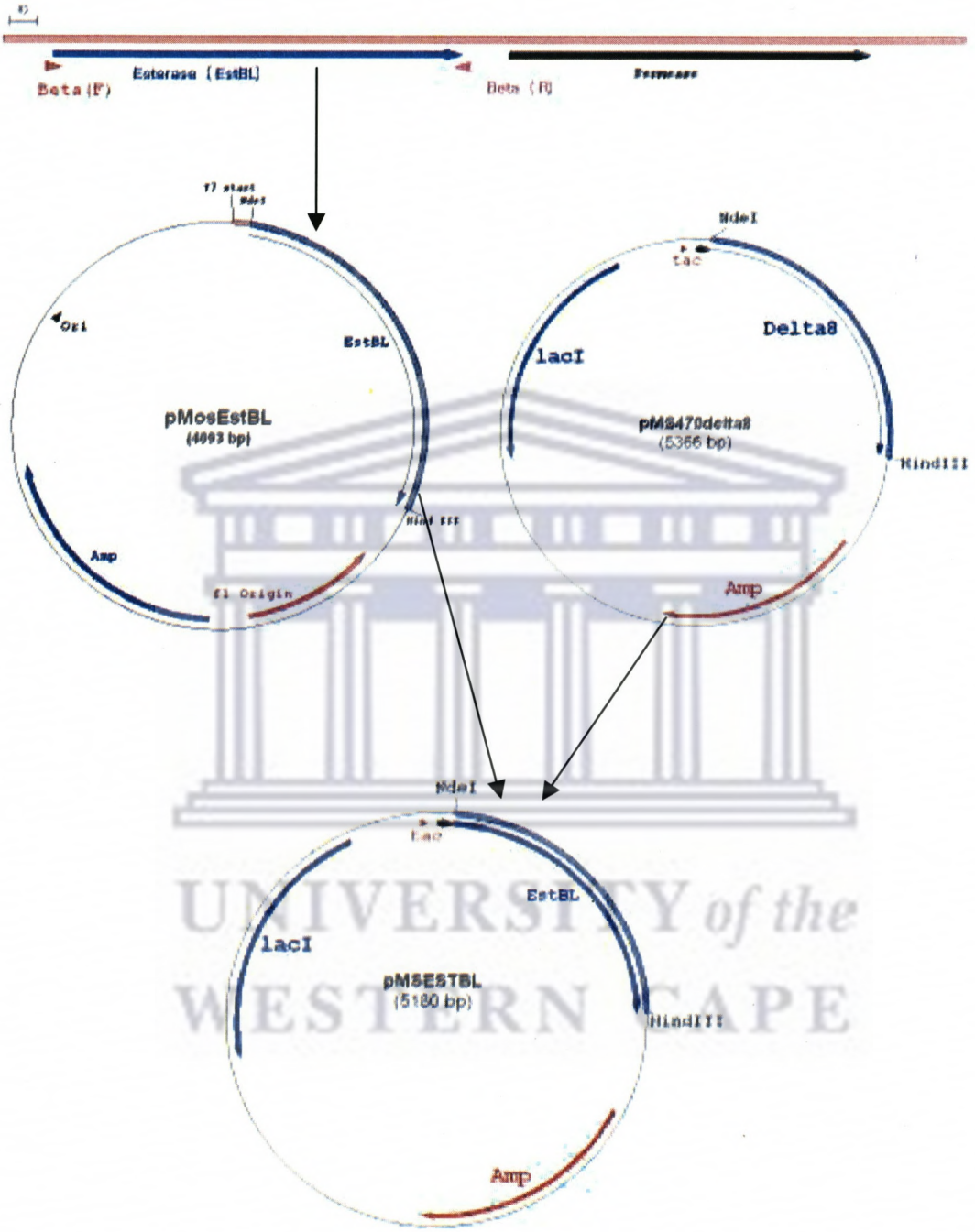


Figure 5.3: Schematic representation of the PCR cloning and subcloning of the esterase (estBL) gene of *B. multivorans* into the pMS470Δ8 expression vector.

5.2.3 Heterologous expression of EstBL in *E. coli*

Expression optimization studies under standard conditions (Chapter 2 Section 2.13, procedure 1) yielded a protein extract with a specific activity of 0.7 U. mg^{-1} after 2.5 h of incubation, following induction with 1 mM IPTG at 37 °C. Given that the gene was being transcribed from a strong *tac*-promoter, this value was low on the basis of comparison with the 0.4 Umg^{-1} specific activity observed with the uninduced fraction. Alternative culturing conditions were tested to optimise *estbBL* expression including changing the growth medium, growth temperature, inducer concentration and incubation period after induction. However, these experiments did not increase the specific activity (Data not shown).

A second expression optimization protocol, involving induction at the point of inoculation, followed by growth at 18 °C was also investigated (Chapter 2: Section 2.13). Initial EstBL expression studies at 37 °C with 1 mM IPTG showed a specific activity of 1.1 U mg^{-1}, which was still very low based in comparison with the 0.5 U mg^{-1} "leaky" activity observed in the uninduced soluble fraction. After attempting different expression conditions including varying temperature and IPTG concentrations, optimal expression conditions were found to be the following: 18 °C, incubation temperature, 0.1 mM IPTG concentration (Fig. 5.4). These findings suggested that both growth temperature and the IPTG concentration were important parameters in achieving high level expression of soluble EstBL. A specific activity of 9.1 U mg^{-1} was estimated in the crude soluble fraction of the induced culture, which was 10-fold higher than that observed when

induction was carried out at higher growth temperatures and IPTG concentrations (Fig. 5.4).

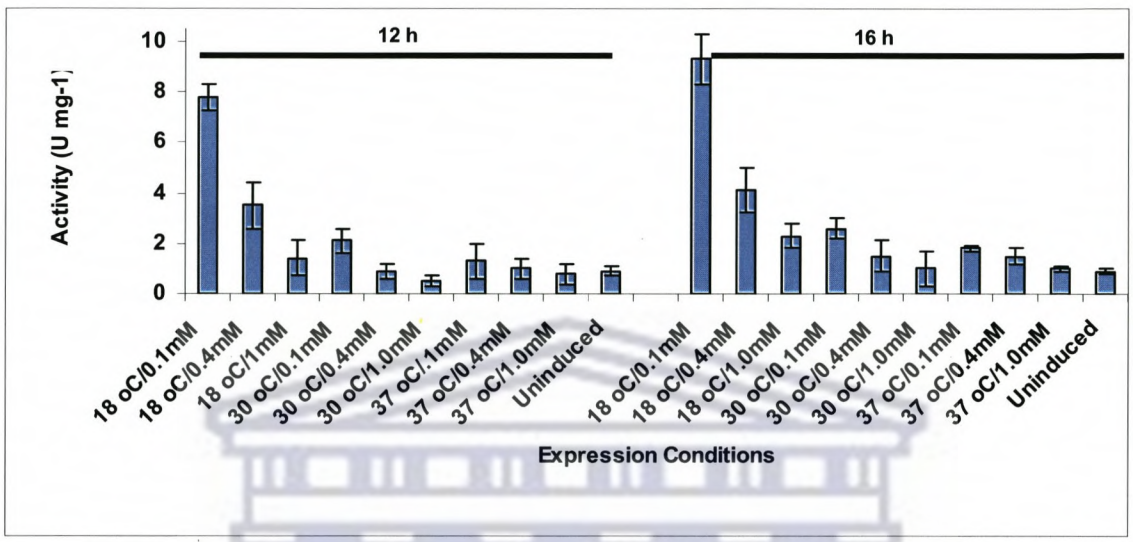


Figure 5.4: Influence of inducer concentration and induction temperature on EstBL activity in the crude extracts of *E. coli* BL21 (DE3) harbouring plasmid pMSESTBL.

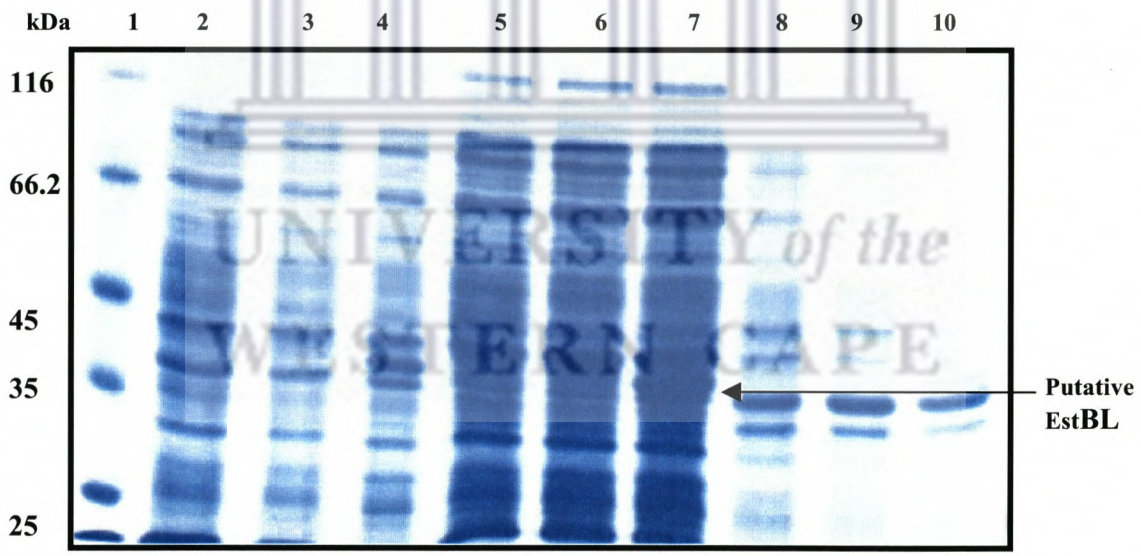


Figure 5.5: SDS-PAGE (12%) electrophoregram showing the expression of EstBL in pMS470Δ8 vector. Lane 1: Marker (kDa), Lane 2: Total protein of *E. coli* (BL21), Lane 3: Total protein of *E. coli*/ pMSEstBL (Uninduced), Lane 4: Total protein pMSEstBL (Induced), Lane 5: soluble fraction of *E. coli* (BL21), Lane 6: soluble fraction of *E. coli*/pMSEstBL (Uninduced), Lane 7: soluble fraction of *E. coli*/ pMSEstBL (Induced), Lane 8: Insoluble fraction. *E. coli* (BL21), Lane 9: Insoluble fraction of *E. coli*/ pMSEstBL (Uninduced), and Lane 10 Insoluble fraction (*E. coli*/ pMSEstBL) induced.

The total cell protein, the soluble and insoluble fractions of the *E. coli* (BL21) cells harbouring pMSESTBL under these conditions were analysed by SDS-PAGE (Fig. 5.5). An extra protein band at 42 kDa corresponding to the predicted molecular weight of EstBL (41.895 Da) was observed in the induced soluble cytoplasmic fraction. The induced insoluble protein fraction showed no indication of equivalent band suggesting that expression at low temperature and inducer concentration favoured only the formation of soluble EstBL.

5.2.4 Purification of the EstBL

Conventional purification procedures were followed since the expression construct was designed without fusion tag sequences at either the C- or N- termini. This was done in order to ensure that the recombinant enzyme was structurally homologous to the native enzyme.

5.2.4.1 Ammonium sulphate precipitation

In order to reduce protein contamination and to facilitate the binding of the enzyme in the subsequent hydrophobic interaction column, crude cell homogenate with a specific esterase activity of 9.9 mg ml^{-1} was subjected to ammonium sulphate precipitation at 20% saturation. This step resulted in a slight increase in esterase specific activity to 10.4 U mg^{-1}. About 94.3% of initial esterase activity was recovered after this step.

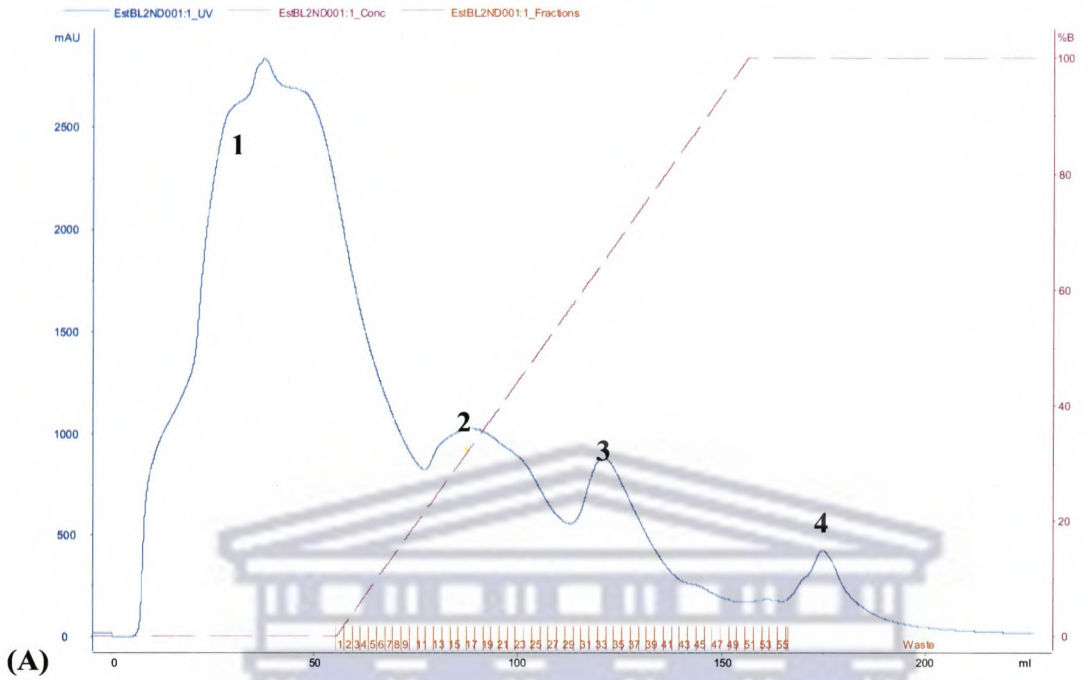
5.2.4.2 Hydrophobic interaction

For further purification, the ammonium sulphate fraction containing EstBL was loaded onto a Phenyl-Sepharose hydrophobic interaction column (HIC). The esterase elution profile from the HIC column is shown in Fig. 5.6A. Assays of the flow-through fraction (Peak 1) did not indicate any esterase activity, suggesting that the enzyme was bound to the HIC column. Elution of bound proteins with a decreasing linear gradient (0.6-0 M) of ammonium sulphate over 5 column volumes resulted in three distinct peaks, labelled 2, 3, and 4 on the chromatogram (Fig. 5.6A).

Esterase activity could only be detected in the fractions corresponding to peak 3, which was eluted at around 290 mM ammonium sulphate. SDS-PAGE analysis of the esterase-positive fractions showed a strong putative EstBL protein band (Fig. 5.6B). The specific activity of the pooled fractions increased to 39.9 U mg^{-1} with 52% of the esterase activity recovered, indicating a loss of above 20% of total esterase activity. The HIC step gave a 4-fold increase in specific activity.

5.2.4.3 Ion exchange chromatography

The pooled fraction containing EstBL activity from the HIC column was concentrated and dialyzed before loading onto a Q-Sepharose ion exchange column. The elution profile of EstBL activity from this chromatographic step is shown in Fig. 5.7A. No esterase activity was detected in the flow-through fraction suggesting that EstBL was bound to the column. Elution of the bound proteins with an increasing linear gradient of NaCl (0-0.5 M) resulted in three peaks (labelled 1-3).



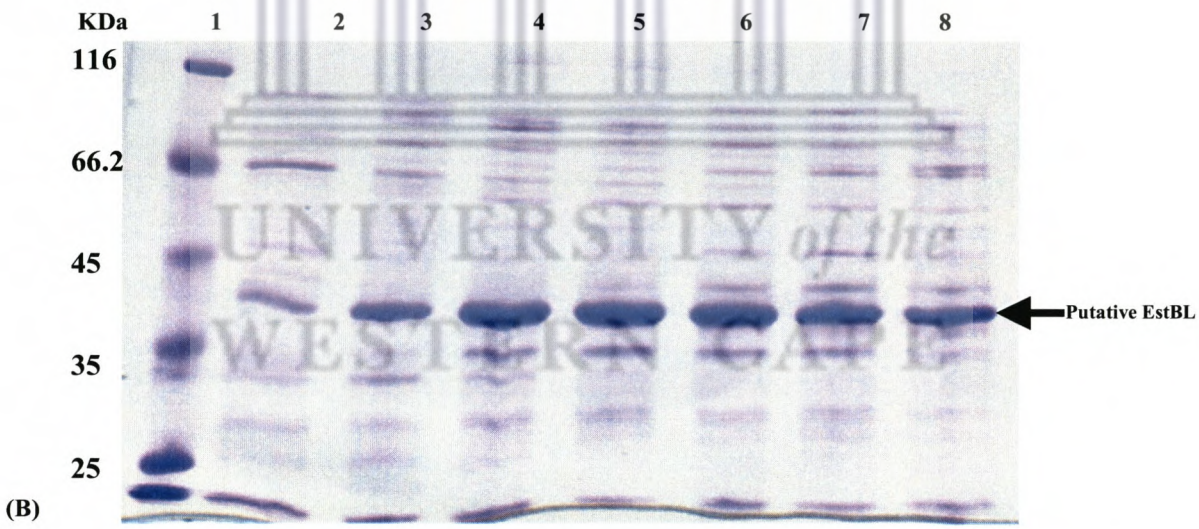


Figure 5.6: (A) Phenyl Sepharose hydrophobic interaction chromatogram showing the elution profile of putative *B. multivorans* esterase (EstBL). (B) SDS-PAGE (12%) electrophoregram showing the eluted esterase-containing fractions (Lanes 2-8 correspond to the fraction numbers 31-37 on the chromatogram).

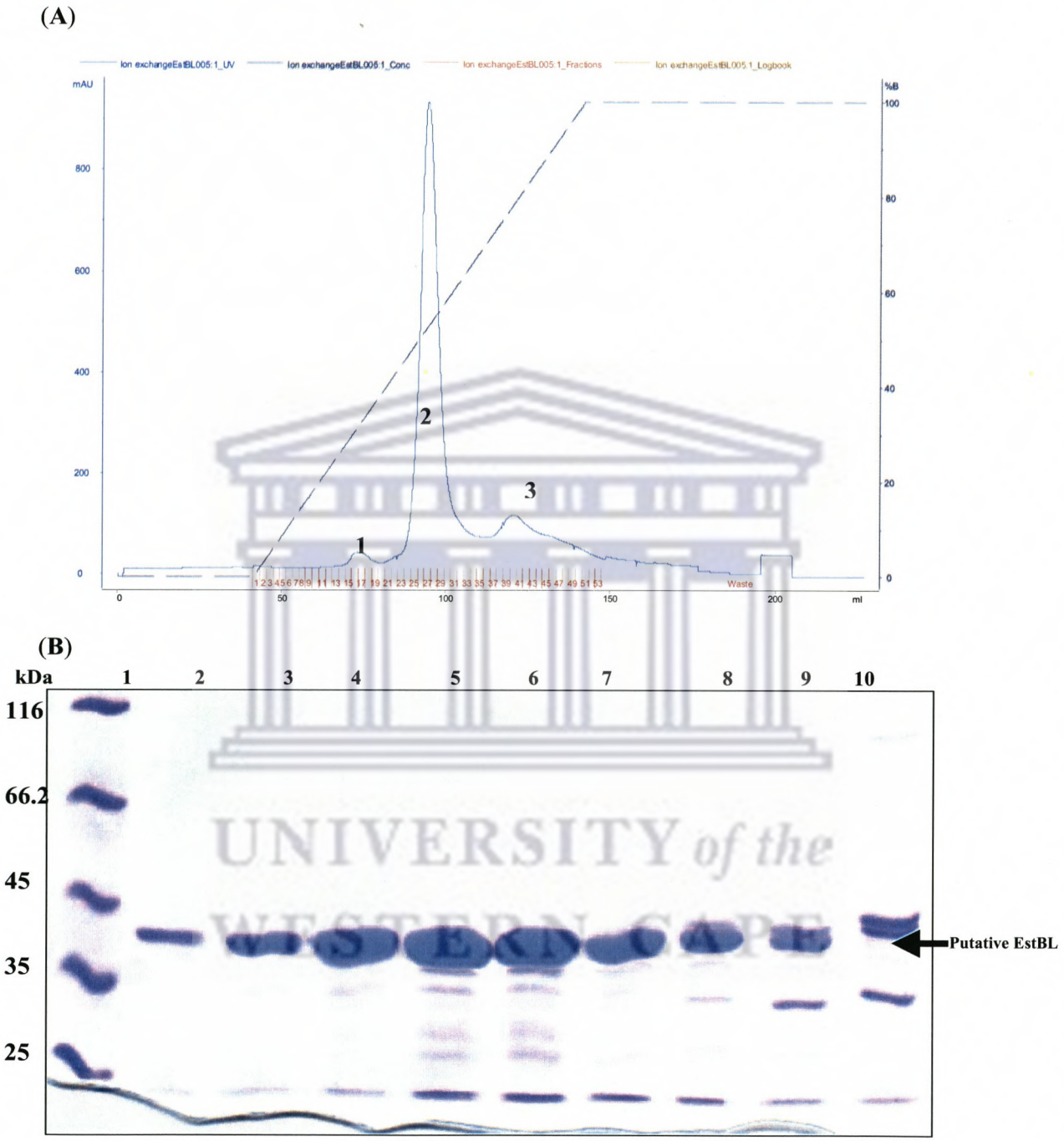


Figure 5.7: (A) Q-Sepharose ion exchange chromatogram showing the elution profile of putative *B. multivorans* esterase (EstBL). (B) SDS-PAGE (12%) electrophoregram showing the eluted esterase containing fractions (Lane 2-10 correspond to the fraction number 24-32 on the chromatogram).

The activity assay indicated the presence of esterase activity in fractions corresponding to peak 2, which was eluted at a NaCl concentration of around 270 mM. SDS-PAGE analysis of the peak 2 fractions revealed a dominant putative EstBL band with only minor contaminating proteins (Fig. 5.7B). The specific activity significantly increased from 39.9 U mg^{-1} to 58.5 U mg^{-1} with a purification of 5.9-fold. Approximately 47% of the esterase was recovered after this step.

5.2.4.4 Size exclusion chromatography

Pooled esterase-containing fractions from Q-Sepharose ion exchange column were loaded onto a Superdex 75 size exclusion column. The elution profile from the Superdex 75 column (Fig. 5.8A) showed a single peak. Activity assays showed that fractions under this peak contained esterase activity. The specific esterase activity of the pooled fractions (19-22) was 113.2 U mg^{-1} with 37.1% of the EstBL recovered. The SDS-PAGE analysis of the esterase-positive fractions showed the presence of a single protein band of the expected molecular weight (Fig. 5.8B). The purification factor from this chromatographic step was 11.5-fold.

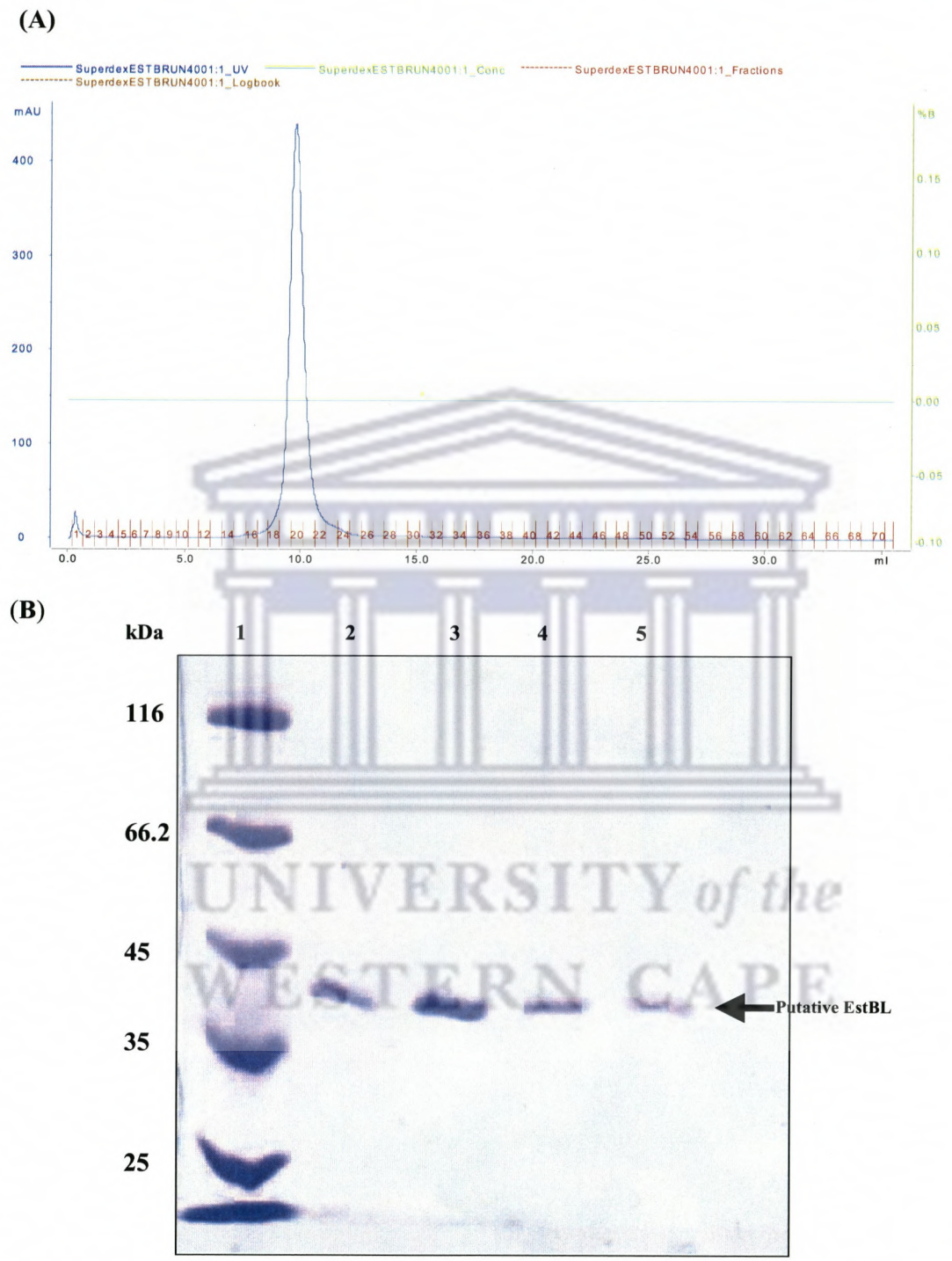


Figure 5.8: (A) Superdex 75 size exclusion chromatogram showing the elution profile of the putative *B. multivorans* esterase (EstBL). (B) SDS-PAGE (12%) electrophoregram showing the eluted esterase-containing fractions (Lane: 2-5 correspond to the fraction number 19-22 on the chromatogram).

5.2.5 Subunit molecular weight determination

A purification table and the SDS-PAGE analysis of EstBL-containing fractions at different purification steps are shown in Table 5.1 and Fig. 5.9, respectively. The final purification step yielded a single protein band corresponding to a molecular weight of 42 kDa. The repeated correlation between protein and esterase activity-containing fractions, the symmetry of the Superdex 75 elution peak, and the apparent molecular weight of the putative enzyme band all suggest convincingly that recombinant EstBL was purified to homogeneity by this series of chromatographic steps.

Table 5.1: A summary of EstBL purification steps

Steps	Volume (ml)	Activity[a] (U)	Protein (mg/ml)	Specific activity (U/mg)	Recovery (%)	Purification (-fold)
Crude extract	20	70	7.1	9.9	100	1.0
$(NH_4)_2SO_4$ precipitation (20%)	22	60	5.7	10.4	94.3	1.1
Phenyl-Sepharose	14	52	1.3	39.9	52.0	4.0
Q-Sepharose	18	37	0.6	58.5	47.8	5.9
Superdex 75	4	26	0.23	113	37.1	11.5

[a] =Activity was measured with 1 mM ρ-nitrophenyl acetate as a substrate

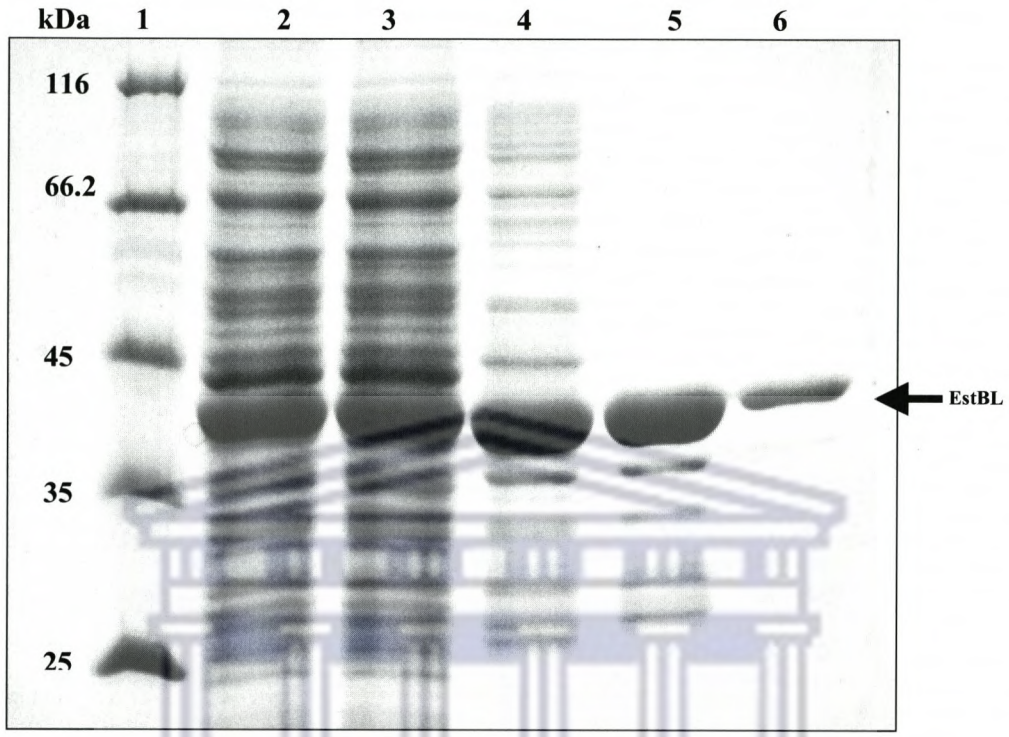


Figure 5.9: SDS-PAGE (12%) electrophoregram of EstBL from different purification steps: Lane 1: Molecular weight markers, Lane 2: Crude cell free fraction, Lane 3: Ammonium sulphate fraction, Lane 4: Phenyl-Sepharose hydrophobic interaction fraction, Lane 5: Q-Sepharose ion exchange fraction, Lane 6: Superdex 75 size exclusion fraction.

5.2.6 EstBL activity under non-denaturing PAGE conditions

The molecular weight of the purified EstBL was also determined under non-denaturing PAGE conditions. Under native PAGE, EstBL migrated as a single protein band with a molecular weight of approximately 42 kDa, strongly suggesting that EstBL is a monomer (Fig 5.10A). It is acknowledged that molecular weight estimation on native PAGE is generally very unreliable, due to the effects of both molecular size and charge density on protein mobility.

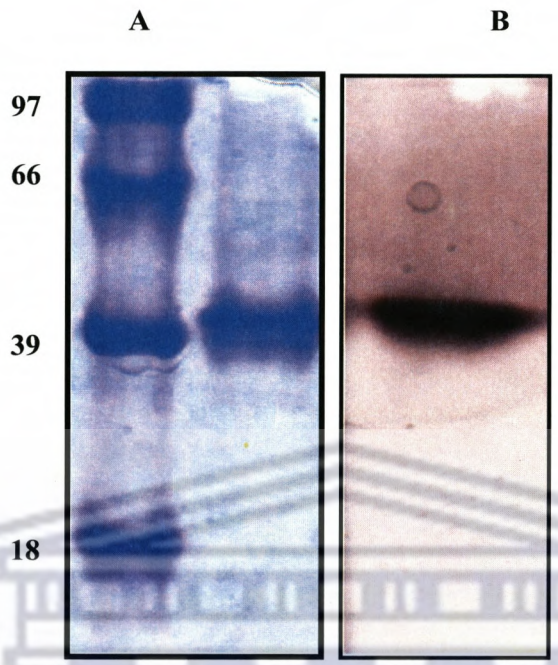


Figure: 5.10: Non-denaturing PAGE (12%) electrophoregram showing (A) Coomassie-stained EstBL protein (B) activity stained EstBL proteins.

Activity of the purified EstBL was detected after non-denaturing PAGE by staining the gel in an α-naphthol acetate substrate solution in the presence of fast blue B dye (Fig. 5.10b). The principle of the assay involves the binding of the α-naphthol product with the dye forming a purple (or dark red) colour (Gudelj *et al.*, 1998). Following gel staining in the substrate-dye solution, a dark red band was observed within one minute at room temperature. Comparisons of the migration pattern of the activity stained band and a parallel loaded EstBL solution stained with Coomassie brilliant blue revealed that both bands migrating to an apparent molecular weight of 42 kDa (Fig. 5.10A).

5.2.7 Biochemical characterization

5.2.7.1 Temperature and pH stability profiles

The thermostability profile of the recombinant EstBL was presented as the percentage of the initial activity remaining after a 30 min incubation at a range of temperatures 15-70 °C (Fig. 5.11). EstBL retained full activity at temperatures between 25 and 30 °C. At elevated temperatures the enzyme was inactivated. For example, 15% of activity was lost at 40 °C while activity loss at 50 °C was 37%. No activity could be detected after a 30 min incubation at temperatures above 60 °C.

The pH stability of EstBL was also investigated and the profile was presented as the percentage of the initial activity remaining after incubation at pH range of 6.0-10.0 for 30 min at 25 °C (Fig. 5.12). The pH stability below pH 6.0 and above pH 10 could not be investigated due to low extinction coefficient of the ρ-nitrophenol product and high background hydrolysis of the substrate, respectively. EstBL was found to be stable over a pH range of 7.0-8.5 with approximately 95% of activity retained. EstBL lost half of initial activity at pH 6.0 while 75% of activity was lost at pH 10.

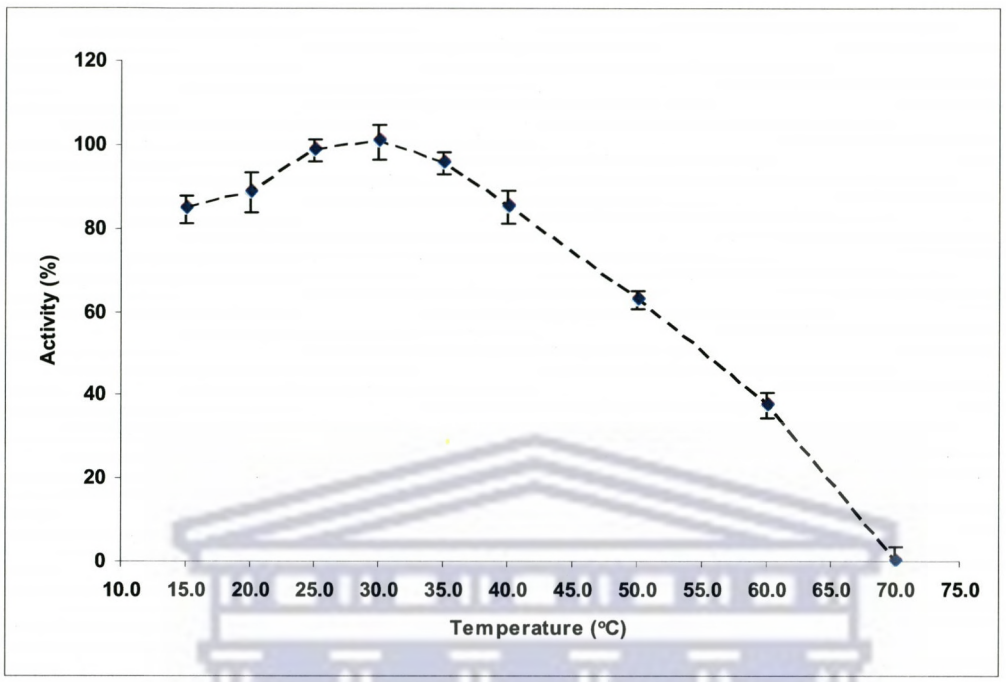


Figure 5.11: Thermostability profile of the recombinant EstBL as defined by residual activity after 30 min incubation at specific temperature.

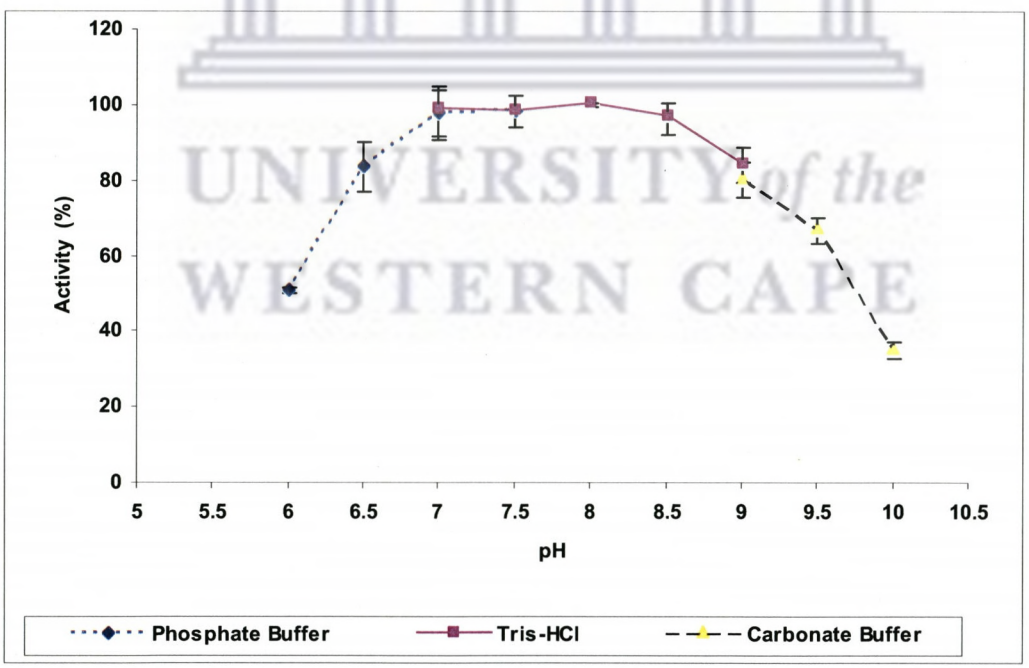


Figure 5.12: pH stability profile of the recombinant EstBL as defined by residual activity after a 30 minute incubation at specific pH values.

5.2.7.2 Substrate specificity

5.2.7.2a EstBL activity against ρ-nitrophenyl- and β-naphthyl-esters

The hydrolytic activity of EstBL against different fatty acid esters was investigated using a range of ρ-nitrophenyl esters; (acetate, C2; propionate, C3; butyrate, C4; caproate, C6; caprylate, C8; caprate, C10; myristate, C14; palmitate, C16) and β-naphthyl (C2-C16) esters. The results are summarized in Fig. 5.13. Analysis of the EstBL hydrolytic activity against ρNP esters showed a strong preference towards short length acyl chains of ρNP-C2, ρNP-C3 and ρ-NP-C4 esters, with ρNP-C2 being the most easily hydrolysed substrate. The hydrolytic activity of EstBL against medium chain esters of ρ-NP-C6, ρ-NP-C8 and ρ-NP-C10, decreased drastically. EstBL activity was 6.8, 7.4 and 17.9-fold lower for ρ-NP-C6, ρ-NP-C8, and ρ-NP-C10 respectively, relative to that of ρNP-C2 (100%). EstBL did not show any detectable activity against long chain substrates (ρ-NP-C14 and ρ-NP-C16).

EstBL hydrolytic activity against β-naphthyl esters showed a similar trend, with highest activity exhibited against β-naphthyl-C2 and β-naphthyl-C3 esters (Fig. 5.13). EstBL activity against medium chain ester, β-naphthyl-C5 was 5.6 fold lower relative to β-naphthyl-C2. EstBL was not active against β-naphthyl esters with *n*-acyl chain lengths greater than C8.

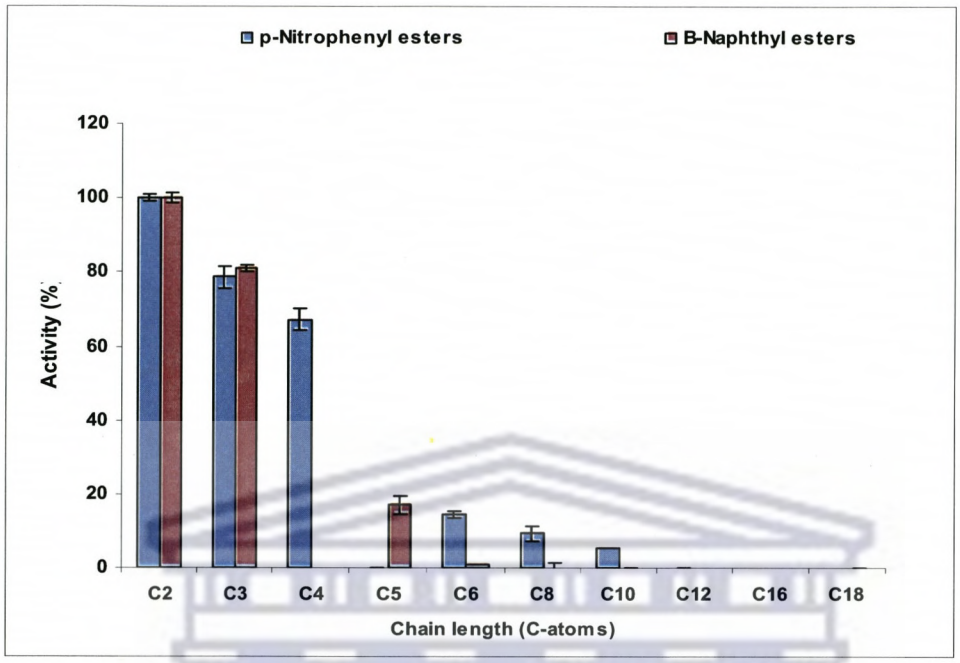


Figure 5.13: Influence of acyl chain length on EstBL activity (Activity against ρ-NP-C2was taken as 100%).

5.2.7.2b Reaction kinetics

The kinetic parameters of EstBL were determined for short chain length substrates (ρ-NP-C2, ρ-NP-C3, ρ-NP-C4). To determine EstBL reaction kinetics accurately, the experimental data of initial velocity versus substrate concentration with coefficients of variation of ≤ 5% were fitted to the Michaelis-Menten equation by weighted non-linear regression (Hyperbola) analysis using GraphPad Prism version 4.0 (Appendix 5.I). The data for the hydrolysis of ρ-NP-C2, ρ-NP-C3 and ρ-NP-C4 by EstBL are also presented as a Lineweaver-Burk plot for display purposes, in order to compare non-linear regression data to linear regression plot as shown in Appendix 5.II. Although 1/v vs 1/s plots resulted in a reasonable good fits ($r^2 \geq 0.90$) to the experimental data for EstBL, it

should be stressed that this data was not considered for discussion, for reasons reviewed elsewhere (Einsenthal *et al.*, 1974; Cornish-Bowden and Endrenyi, 1981).

A summary of the kinetic constants of EstBL hydrolysis of the three most readily hydrolyzed substrates, ρ-NP-C2, ρ-NP-C3, and ρ-NP-C4, is shown in Table 5.2. The K_M value was highest for ρ-NP-C2, followed by ρ-NP-C3 and ρ-NP-C4, which were 1.6 and 3.6-fold lower relative to ρ-NP-C2. These data suggest that the increase in acyl chain length of the substrate results in an increase in substrate affinity to the enzyme. This in turn directly affected the maximum velocity of the reactions as shown by a decrease in V_{max} values with an increase in chain length of the substrates. The catalytic turnover (k_{cat}) also showed a decreasing trend with increase in substrate chain length for the three substrates, with ρ-NP-C3 and ρ-NP-C4, being 2 and 6-fold lower relative to ρ-NP-C2. The catalytic efficiencies (as measured by k_{cat}/K_M) suggest that the three substrates were being converted at different efficiency under the conditions investigated.

Table 5.2: Kinetic parameters for hydrolysis of various ρ-nitrophenyl esters

Substrate	K_M (mM)	V_{max} (U mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
Acetate (ρ-NP-C2)	8.1 ± 1.1	1675.3 ± 116.5	636.7 ± 46.4	78.6
Propionate (ρ-NP-C3)	5.0 ± 0.8	804.8 ± 62.6	319.4 ± 24.6	63.8
Butyrate (ρ-NP-C4)	2.2 ± 1.7	276.1 ± 8.9	109.5 ± 3.5	49.8

Enzyme concentration in the assay= 9.6 μg ml^{-1}

k_{cat} was calculated assuming a molecular weight of 42 kDa estimated by SDS-PAGE, and a single active site per monomeric protein.

5.2.7.3 EstBL activity against β-lactam substrates

The high homology of the EstBL sequence to a number of class C β-lactamases and esterases belonging to family VIII led to the suggestion that EstBL could be a bi-functional enzyme with both esterolytic and β-lactamase activity. This observation prompted an investigation of the activity of EstBL enzyme on β-lactam substrates. While a commercial *B. cereus* β-lactamase showed significant activity against β-lactam substrates, EstBL showed no activity against these substrates (Table 5.3). These observations led to a conclusion that EstBL is not functionally related to class C β–lactamases, regardless of the presence of the β-lactamase motif within its primary structure (Chapter 4, Section 4.5).

Table 5.3 EstBL activity against β–lactam substrates

Substrate	BC-lactamase ($U\ mg^{-1}$)	EstBL ($U\ mg^{-1}$)
Ampicillin	6.1 ± 0.510	BQL
Cephalosporin C	0.69 ± 0.003	0.002±0.006
Cephalotin	0.20 ± 0.002	BQL

BC-Lactamase=*Bacillus cereus* β-lactamase

BQL=below quantification limit

5.2.7.4 Effect of metal ions

The effects of metal ions on enzyme activity may vary considerably. Metal ions may induce conformation changes in the enzyme that may in turn induce or inhibit the activity of the enzyme. In some cases, however metal ions have been shown to play an important role in stabilizing enzymes by binding at specific sites and hence reducing the flexibility of the enzyme and its susceptibility to partial unfolding (Gregory *et al.*, 1993). The effect of different metal ions on EstBL activity was investigated at 1mM concentration after 30-min incubation at 25 °C and results are shown Table 5.4. Taking experimental errors into account, Mg^{2+} Li^{2+} and Ca^{2+} showed no effect on EstBL activity. Several divalent cations (e.g., K^{2+}, Ni^{2+}, Co^{2+}, Mn^{2+} and Zn^{2+}) partially inhibited EstBL activity. Cu^{2+}, Fe^{3+} and Ag^{2+} markedly inhibited EstBL activity with more than 60%, 75% and 95% of activity lost, respectively.

Table 5.4: Effect of metal ions on esterase activity

Metal	Specific activity (U mg^{-1})	Residual activity (%)[b]
Control	153.2 ± 6.7	100
Mg^{2+}	148.4 ± 5.0	97
Li^{2+}	146.3 ± 9.5	96
Ca^{2+}	139.8 ± 4.0	91
K^{2+}	119.7 ± 6.4	83
Ni^{2+}	127.7 ± 9.1	78
Co^{2+}	117.7 ± 3.8	77
Mn^{2+}	113.5 ± 2.4	74
Zn^{2+}	84.2 ± 4.5	55
Cu^{2+}	53.8 ± 2.0	35
Fe^{3+}	21.1 ±2 .0	14
Ag^{2+}	6.7 ± 4.3	4.0

[b]=100% activity represents esterase activity towards ρ-nitrophenyl acetate with no cations added.

5.2.7.4 Effect of detergents

The effect of different detergents on EstBL activity was investigated, and the results are shown in Fig. 5.14. Data are the presented as the percentage of residual activity after 30-minute incubation with the detergent at 25 °C. Non-ionic detergents (Triton X-100 and Tween 20) at 0.5% (v/v) did not result in a significant loss of EstBL activity. EstBL retained more than 80% of the activity in the presence of a zwitterionic detergent (CHAPS) at 0.5% (w/v). EstBL activity was markedly decreased in the presence of mono-ionic detergents. More than 90% of EstBL activity was lost in the presence of SDS (0.5% v/v), while lauroylsarcosine and sodium cholate resulted in 31 and 46% loss of esterase activity, respectively.

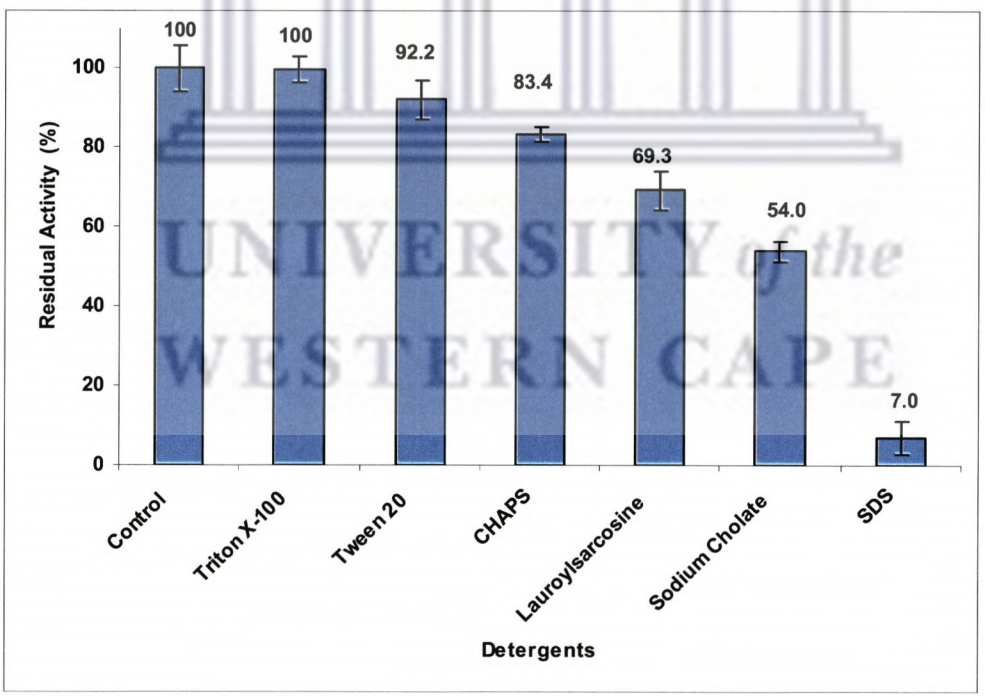


Figure 5.14: Effect of detergents on EstBL activity

5. 2.7.5 Effect of inhibitors

The activity of EstBL in the presence of various inhibitors (phenylmethylsulfonyl fluoride, (PMSF), ethylenediaminetetracetic acid (EDTA) and dithiothreitol (DTT)) at various concentrations was investigated (Table 5.5). Data are presented as percentage of residual activity after 30-minute incubation with the inhibitor at 25 °C. EstBL was significantly inhibited by PMSF with only 23% of activity remained at 1 mM. Taking experimental errors into account, the reducing agent (DTT) at 0.001, 0.1 and 1 mM did not result in any significant lost of activity. EstBL retained most of the activity (≥ 90%) in the presence of the metal chelator, EDTA, at 1mM.

Table 5.5: Effect of inhibitors on EstBL activity

Inhibitor	Concentration (mM)	Activity remaining[b] (%) ± SD
Control	-	100 ± 5.0
DTT	0.001	99 ± 3.1
	0.1	92 ± 6.0
	1	88 ± 11.0
EDTA	0.001	100 ± 5.4
	0.1	97.3 ± 8.0
	1	95.1 ± 3.0
PMSF	0.001	66.5 ± 7.1
	0.1	44.9 ± 5.0
	1	23.3 ± 2.0

[b]=100% activity represents esterase activity towards ρ-nitrophenyl acetate with no inhibitor added.

5.3 Expression of EstEFH5

The work presented in Chapter four led to the identification and isolation of pTEND5 clone which, based on qualitative data, was shown to encode for both non-specific esterolytic and ethyl ferulate hydrolyzing activities. A full-length gene sequence believed to encode for these esterolytic activities was identified and the primary structure of the corresponding protein deduced. The putative esterase, named EstEFH5, was shown to be encoded by a GC rich template which could affect the expression in *E. coli* due incompatibility in codon usage. This section describes attempts to heterologously express the EstEFH5 gene in *E. coli*.

5.3.1 Expression strategy

Hydrophathy plot analysis of the EstEFH5 primary structure revealed the presence of a hydrophobic stretch of amino acids (30 a.a.), possibly comprising a membrane anchor domain. The presence of a putative membrane anchor domain led to an assumption that this domain might affect the expression in *E. coli* by anchoring the expressed protein to the membrane. Based on this observation, two expression constructs were investigated; one with the full-length esterase gene and the other one with a truncated esterase gene lacking the membrane anchor coding sequence. Based on the expression data from the EstF1 homologue (Khalameyzer *et al.*, 1999), it was assumed that heterologous expression of the EstEFH5 gene would not require additional flanking sequences. The expression strategy employed in this study involved the PCR amplification of EstEFH5 encoding structural gene using restriction-site engineered primers, with subsequent blunt-end cloning into the pMOS*Blue* vector. As a result, the correct inserts carrying the

EstEFH5 structural gene could be screened by restriction analysis before insertion into the corresponding sites of the pMS470Δ8 and pET22b (+) expression vectors.

The pET22b (+) expression vector allows stringent control of the expressed gene and can only be used with an appropriate lysogenic host strain, such as an *E. coli* BL21 (DE3) that contains the gene for T7 RNA polymerase (T7pol). The target genes are inserted into the expression vector downstream T7*lac* promoter that combines the T7 transcriptional promoter and *lac* operator. The vector also carries the natural promoter and the coding sequence for the *lac* repressor (*lac*I), enabling the control of expression through a constitutive repression of transcription of both the T7pol gene from the host chromosome and of the target gene by any T7pol produced, where repression is removed upon the addition of IPTG.

5.3.2 Construction of expression vectors

Two primer sets (EstN/EstH and EstNT/EstH) were designed to target the full length and truncated (without the membrane anchor coding region) esterase gene, respectively (Fig. 5.15). By engineering restriction sites into the PCR primers, the genes could be inserted in the correct orientation and reading frame just downstream of the RBS of pET22b (+). The ATG start codon of the EstEFH5 encoding gene was used to create an *Nde*I site for the 5′ (sense strand) insertion point while *Hind*III sites were added to the end of the 3' primers (anti-sense strand).

(A)

(B)

Est (N): ATACATATGGCGCGTGACGCGCGCGTTTCCG (*Nde*I)

Est (NT): TGAACATATGATCCTGCTCGGCATCGTCGC (*Nde*I)

Est (H): ATTCGCAAGCTTGCAGCGCACCGTTCAGTG (*Hind*III)

Figure 5.15: (A) Layout of the pTEND5 coding regions of the *B. multivorans* UWC10 and the organization of the primer pair represented by red and pink arrows. (B) The nucleotide sequences of the primers used and the engineered restriction sites are underlined.

The 3′ end of the EstEFH5 encoding gene contained a GC rich region that posed a serious problem with regard to designing a reverse primer. The PerlPrimer v1.14 software indicated the presence of reverse primer repeats around this region, with >98% probability of forming primer dimers. In order to overcome this problem, a reverse primer was designed based on the non-coding region immediately down-stream of the TGA stop codon.

Initial attempts to PCR amplify the EstEFH5 coding region using EstN/EstH and EstNT/EstH primer pairs and plasmid construct pTEND5 as a template failed. The successful amplification of the EstEFH5 coding region was achieved after the optimisation of amplification conditions, involving the addition of denaturants (2% DMSO and 0.2 M betaine) in a touch-down PCR protocol. The annealing temperature

was decreased from 65 °C to 55 °C (-1°C/cycle) followed by 20 cycles at an annealing temperature of 55 °C. PCR products of expected size were obtained (978 and 948 bp), corresponding to the full length and truncated EstEFH5 gene fragments, respectively (Fig. 5.16). The PCR products were blunt-end cloned into pMOS*Blue* vector to yield pMOSEFH5 (harbouring the full-length gene) and pMOSEFH6 (harbouring the truncated gene without the membrane anchor coding sequence).

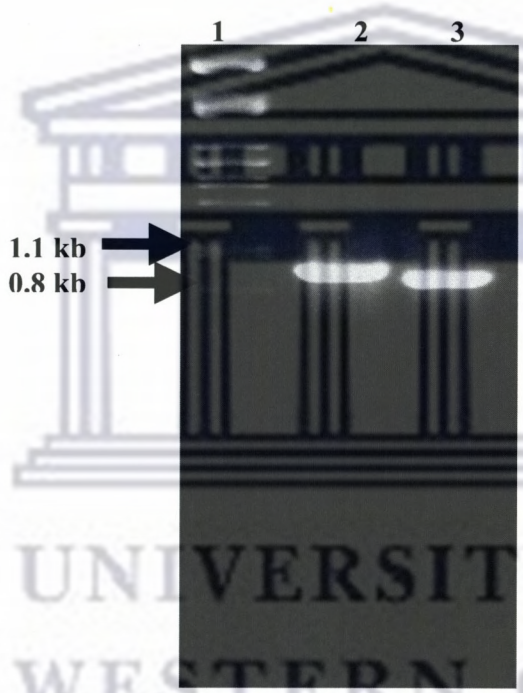


Figure 5.16: Agarose gel (1%) electrophoregram showing PCR amplification of *B. multivorans* UWC10. Lane 1: *PstI*-digested λDNA markers; Lane 2: full length EstEFH5 encoding gene; Lane 3: truncated EstEFH5 encoding gene

Following transformation and screening on tributyrin agar, active clones could only be identified with *E. coli* transformants harbouring the pMOSEFH5 plasmid construct. The absence of esterolytic activity with clones harbouring the pMOSEFH6 plasmid suggested

that the putative membrane anchor domain was important either for catalytic activity or proper folding of the protein. In order to confirm that the absence of the esterolytic activity in clones harbouring plasmid construct pMOSEFH6 was not due to the absence of the putative membrane anchor domain representative pMOSEFH5 and pMOSEFH6 clones were screened by means of engineered restriction sites. The genes were then cloned into corresponding sites of the pET22b (+) expression vector. This resulted in two expression constructs, pETEFH5 and pETEFH6, harbouring the full length and truncated genes, respectively. Following transformation of *E. coli* BL21 (DE3) and functional screening on tributyrin agar the activity was only detected with *E. coli* transformants harbouring the pETEFH5 expression construct (i.e., the full length constructs). Sequencing of the pETEFH5 and pETEFH6 expression constructs using vector-derived T7F and T7R promoter primers revealed that both genes had been correctly inserted within the pET22b (+) multiple cloning sites. Overviews of the full length and truncated EstEFH5 genes subcloning are shown in Fig. 17 and 18.

Expression analyses of the total protein fractions of *E. coli* BL21/pETEFH6 showed the presence of an additional protein band, which corresponded to the predicted 32 kDa size of the truncated esterase gene product (Fig. 5.19). This led to a conclusion that the putative membrane anchor domain was crucial for activity. Based on these observations, the pETEFH6 expression construct was further investigated.

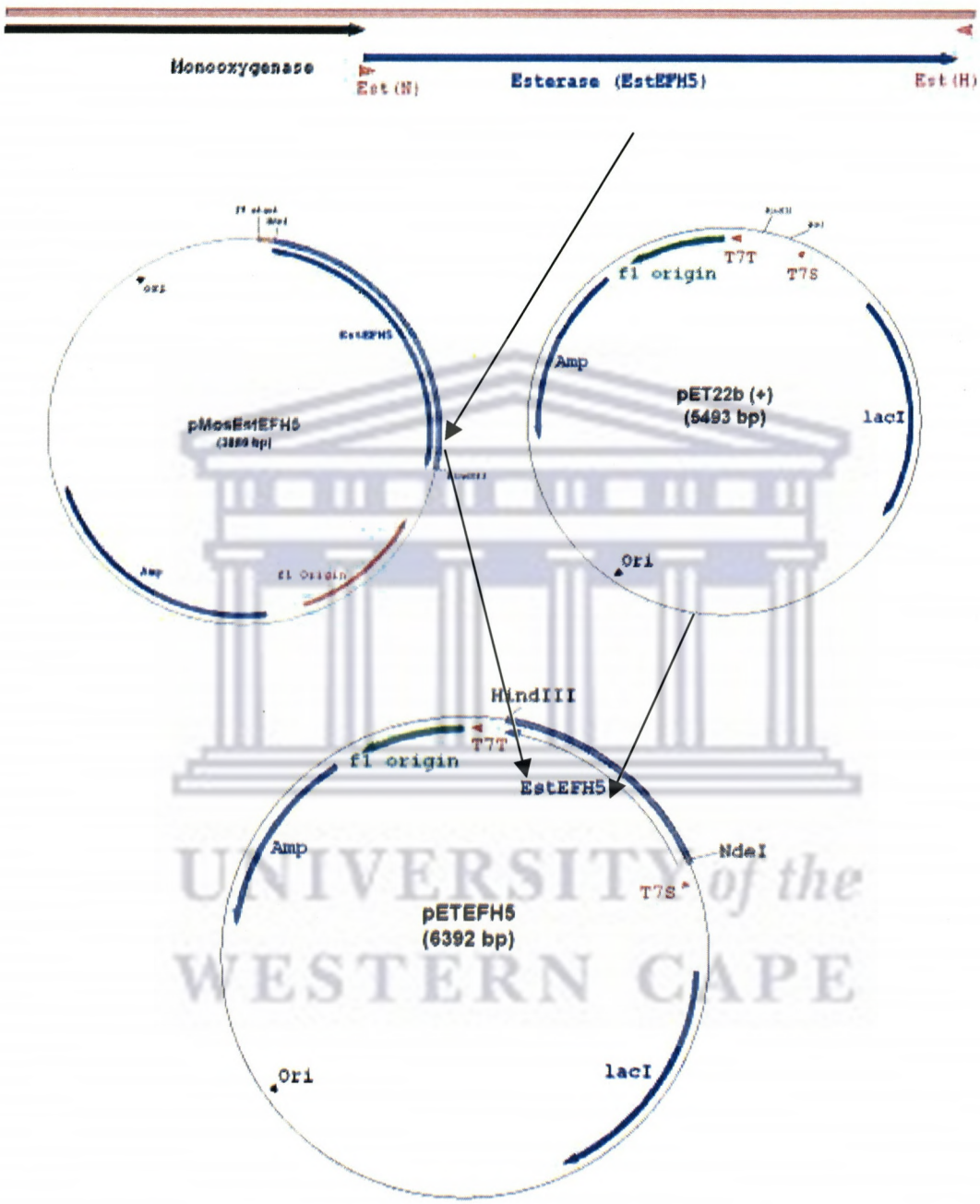


Figure 5.17: Schematic representation of the PCR cloning and subcloning of the full length esterase (EstEFH5) gene of *B. multivorans* into the pET22b (+) expression vector.

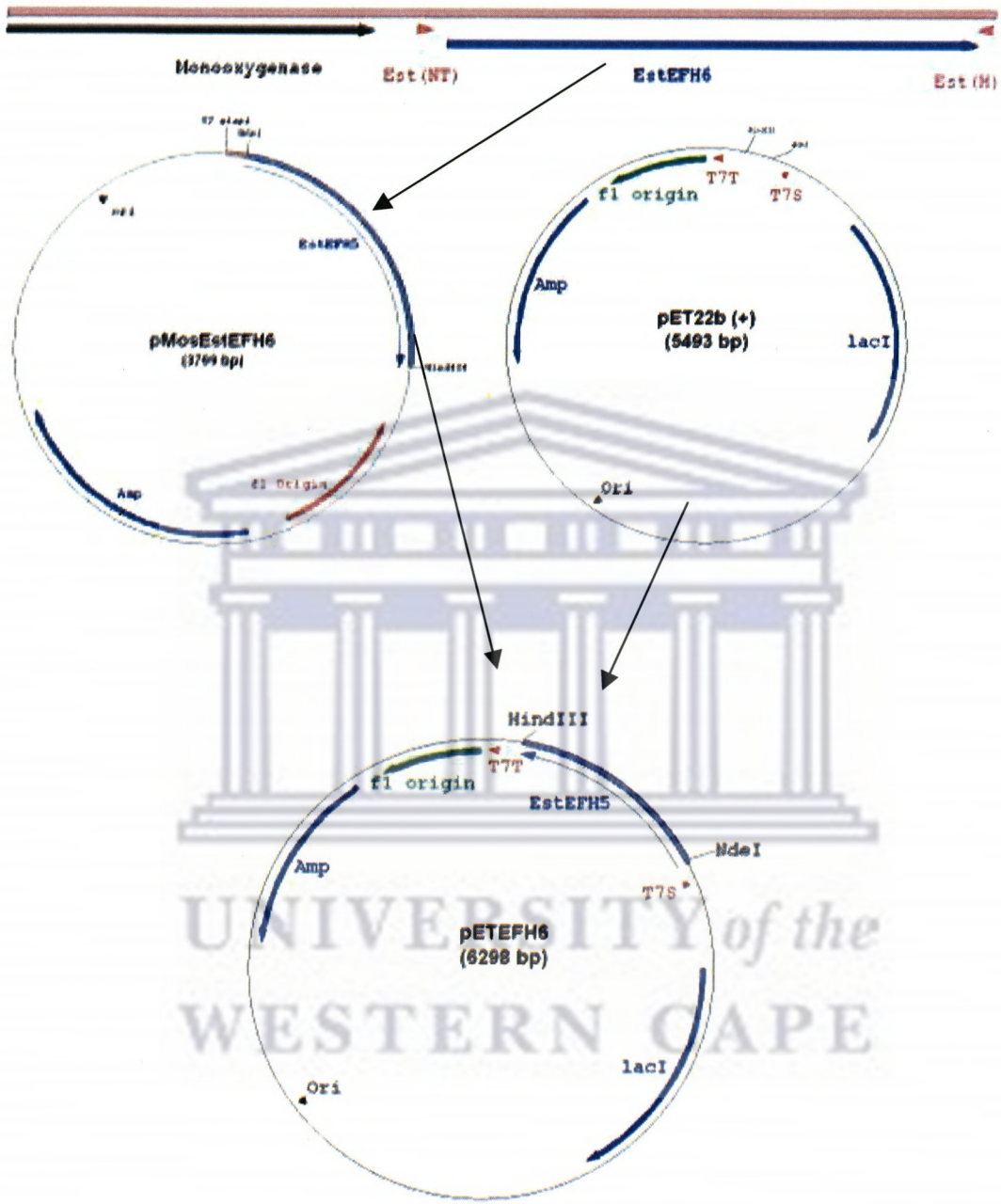


Figure 5.18: Schematic representation of the PCR cloning and subcloning of the truncated esterase (EstEFH6) gene of *B. multivorans* into the pET22b (+) expression vector.

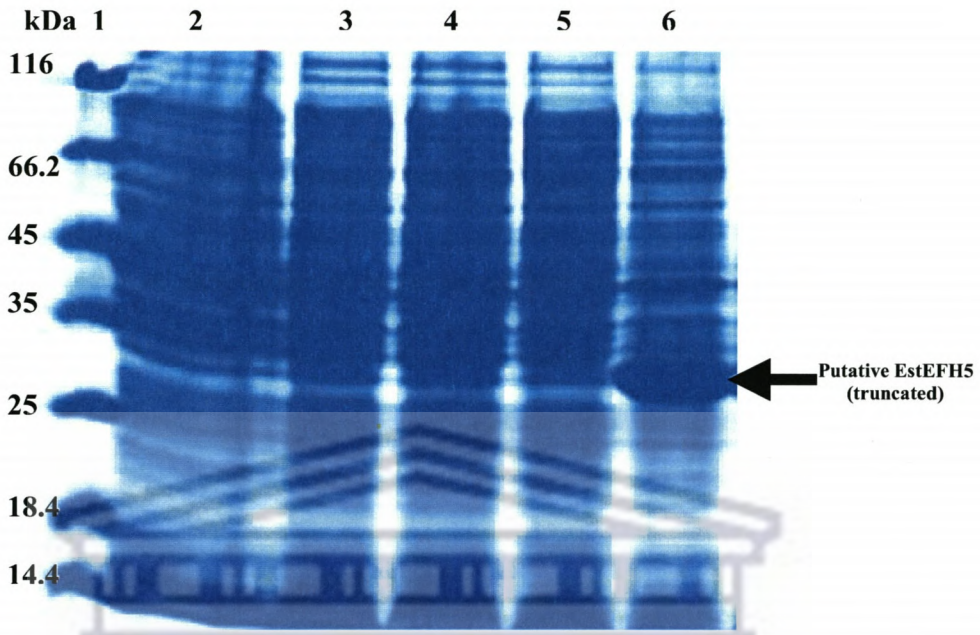


Figure 5.19: SDS-PAGE (12%) electrophoregram showing the analysis of the truncated EstEFH5 esterase production in *E. coli*. The fractions are total cell protein fractions after induction. Lane 1: Marker (kDa), Lane 2: *E. coli* only, Lane 3: *E. coli*/pET22b (Uninduced), Lane 4: *E. coli*/pET22b (induced), Lane 5: *E. coli*/pETEFH6 (Uninduced) and Lane 6: *E. coli*/pETEFH6 (Induced).

5.3.3 Expression of the full length EstEFH5

The expression construct pETFH5 was investigated in order to optimize expression conditions that would favour the formation of soluble EstEFH5 in *E. coli*. Induction under standard conditions (37 ^{o}C for 2.5 h at 0.4 mM IPTG), showed a specific esterase activity of approximately 0.4 U mg^{-1}. This value was very low, in comparison with the activity from the uninduced and induced cultures, which were between 0.1 and 0.3 U mg^{-1} respectively. SDS-PAGE analysis revealed that these conditions favoured the formation of insoluble recombinant EstEFH5 as shown by a protein band in the insoluble fraction at

around 35 kDa, corresponding to the predicted EstEFH5 size (Fig. 5.20). The formation of a large fraction of putative EstEFH5 in the insoluble fraction as evidenced by SDS-PAGE analysis, could account for the low esterase activity observed in the soluble fraction. Although no visible esterase band could be observed in the crude soluble (cytoplasmic) fraction after Coomassie staining, zymogram analysis showed a single esterase activity band (Fig. 5.20B). Furthermore, the soluble fraction contained enough esterase activity to give a zone of clearance on FAE plate assay (Fig. 5.21).

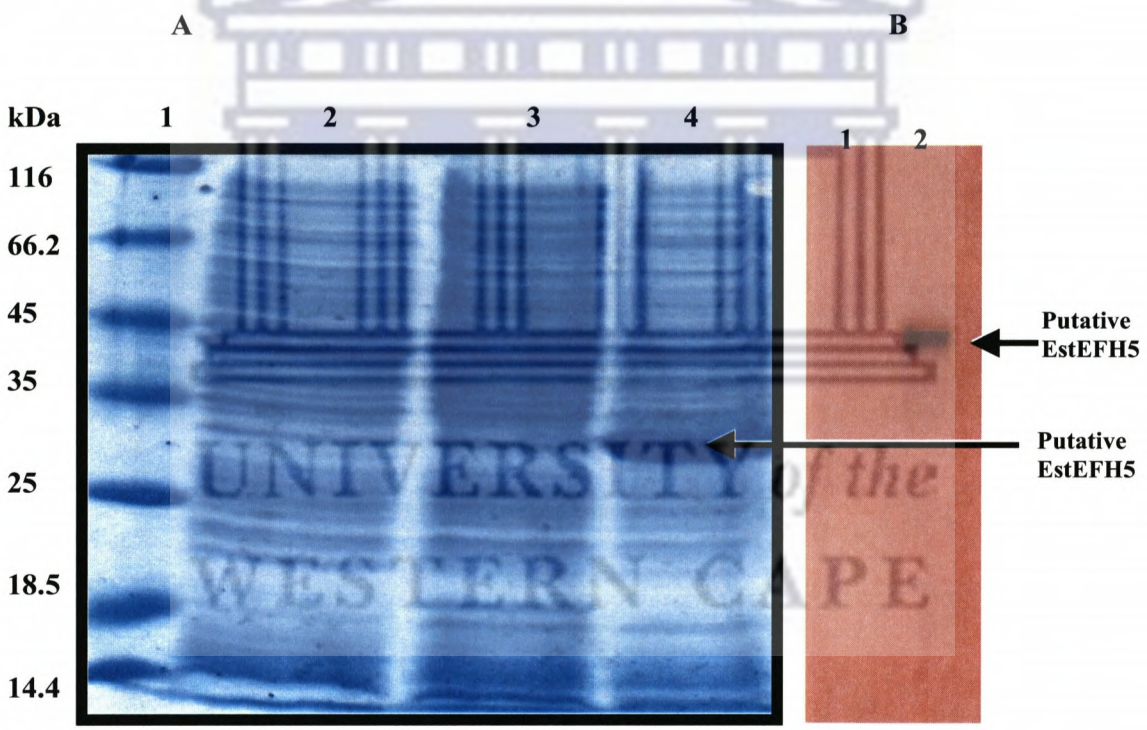


Figure 5.20: (A) SDS-PAGE (12%) analysis of insoluble esterase (EstEFH5) containing *E. coli* fractions. Lane 1: Marker (kDa), Lane 2: *E. coli*/pET22b (Induced), Lane 3: *E. coli*/ pETEFH5 (Uninduced), Lane 4: *E. coli*/pETEFH5 (Induced) (B) Non denaturing-PAGE analysis (1) soluble fraction of *E. coli*/pETEFH5 (Uninduced) (2) soluble fraction *E. coli*/pETEFH5 (induced).

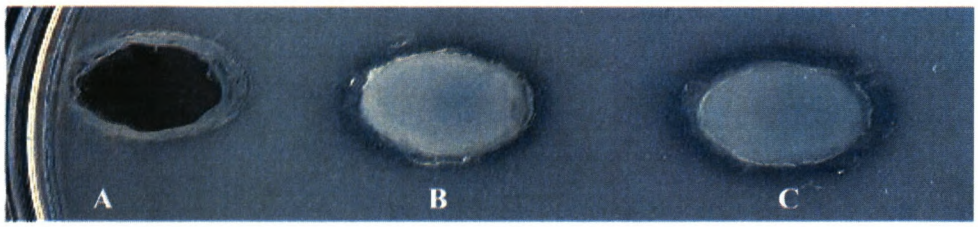


Figure 5.21: Ferulic acid esterase plate assay showing zone of clearance: (A) *E. coli*/pETEFH5 uninduced (B) soluble fraction *E. coli*/pETEFH5 induced containing 20 $\mu g.ml^{-1}$ protein, and (C) 50 $\mu g\ ml^{-1}$ protein.

Alternative culturing conditions, including lowering the IPTG concentration and temperature during expression were tested in attempt to optimise the formation of soluble EstEFH5 (Fig. 5.22). These conditions did not improve the specific activity of the soluble fraction as shown in Fig. 5.22. The addition of 3% v/v ethanol in the LB medium, as reported by Weickert *et al.* (1996), did not improve the solubility. Furthermore, attempts to use detergents to improve the solubility of the protein were also unsuccessful. Non-ionic detergents Triton x 100 and Tween 20 at 1% v/v completely inhibited esterase activity.

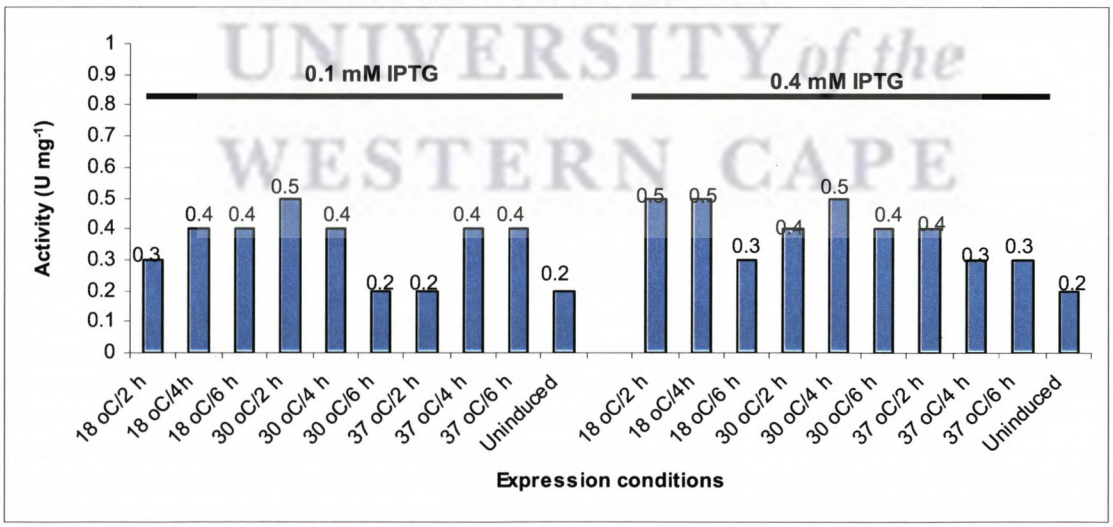


Figure 5.22: Influence of inducer concentration and induction temperature on the EstEFH5 activity in the crude extracts of *E. coli* BL21 (DE3) harbouring plasmid pETEFH5.

5.3.4 Expression of EstEFH5 in pMAL-p2x vector

Recently, a highly hydrophobic esterase from *B. cepacia* encoded by a GC rich template was successfully expressed in a soluble form using the pMAL-p2x expression vector (Kim *et al.*, 2004a). In order to enhance the solubility of the recombinant EstEFH5, an attempt was made to express the enzyme as a fusion protein using the pMAL-p2x expression vector. The multiple cloning site in this vector is positioned to allow translational fusion of the *E. coli* maltose binding protein (MBP), encoded by the *malE* gene, to the N-terminus of the cloned target protein. Transcription of the recombinant gene fusion is controlled by the inducible *tac* promoter (P_{tac}). Basal expression is minimized by the binding of the Lac repressor, encoded by the *lacI* gene, to the *lac* operator immediately downstream of P_{tac} (www. neb.com).

Two primers, pMALX and pMALH, were designed with *Xba*I and *Hind*III restriction sites, respectively. Following PCR, the EstEFH5 gene was fused to *malE* gene which encodes for a 42.5 kDa cytoplasmic maltose binding protein (MBP) to yield the pMALEFH5 expression vector, potentially yielding a fusion product of 77 kDa (42 + 34.5 kDa). SDS-PAGE analysis showed a protein band of the expected size (Fig. 5.23). Analysis of different protein fractions after induction suggested that the protein was still insoluble (Fig. 5.23, Lanes 7 and 9). Lowering the IPTG concentration and temperature during expression as previously shown in previous Section 5.33, did not increase the solubility of the protein.

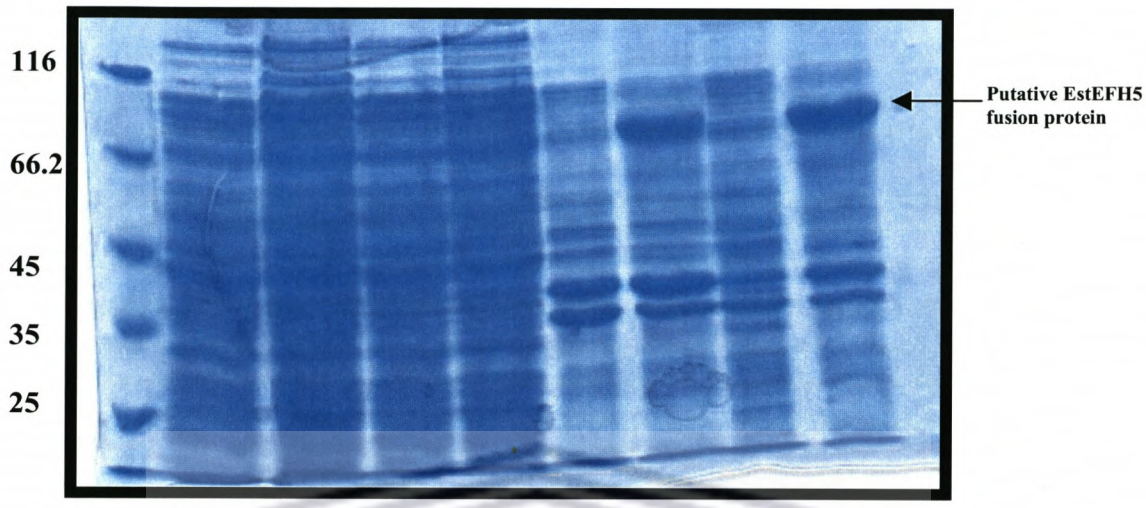


Figure 5.23: SDS-PAGE (12%) showing the expression of EstEFH5 in the pMAL-c2 vector. Lane 1: Marker (kDa), Lane 2: *E. coli* (BL21), Lane 3: *E. coli*/pMAL, Lane 4 and 5: *E. coli*/EstEFH5 soluble fractions uninduced and induced respectively, Lane 6 and 7: *E. coli*/EstEFH5 insoluble fractions uninduced and induced, respectively at 20 °C, Lane 8 and 9: *E. coli*/EstEFH5 insoluble fractions uninduced and induced respectively at 30 °C.

5.3.5 In vitro Refolding studies and purification

Attempts were made to recover a functional EstEFH5 from the insoluble fractions. The refolding strategy adopted in this study involved the denaturing of the insoluble proteins using 8 M urea followed by refolding and purification (see below). The full length EstEFH5 fused to the 6x His-tag at the C-terminus was used for these trials. To obtain the His tagged EstEFH5, the encoding gene was amplified by PCR using pETEFH5 as a template and the oligonucleotides (EstHis: 5'-ATTCGCAAGCTTGCAGCGCACCGTCATGTG-3') and Est (N): ATACATATGGCGCGTGACGCGCGCGTTTCCG) carrying *Nde*I and *Hind*III restriction sites, respectively. The PCR product was purified and cloned into the corresponding sites of pET22b (+), resulting in the pETEFH5$_{tag}$ expression plasmid. Since the reverse primer was against the non-coding sequence downstream the TGA stop

codon, the $EstEFH5_{tag}$ protein had a predicted molecular weight of 37 KDa, due to the additional 18 amino acid residue extension at the C-terminus including the vector encoded sequence (MTVRCKLAAALEHHHHHH).

In vitro refolding was achieved by loading urea (8M) denatured inclusion bodies onto a His select ™ Nickel column. EstEFH5 was allowed to refold by using a decreasing urea concentration linear gradient (8-0 M) over 48 column volumes. The refolded esterase was eluted with 250 mM imidazole, concentrated and the purity was checked by SDS-PAGE analysis and silver staining (Fig. 5.24A). The semi-purified EstEFH5 had a specific activity of 13.4 U mg^{-1}. Activity staining (Fig. 5.24B) resulted in a single band on the zymogram, confirming that no other esterases were present in the purified sample.

The purified protein was found to be very unstable following removal of glycerol from the enzyme solution after purification (Table 5.6). After imidazole removal, the enzyme was resuspended in Tris-HCl (20 mM pH 7.0) containing 15% v/v glycerol before being stored at 4 and -80 °C. Significant loss of activity within 2 days of storage, as shown in Table 5.6, prompted the suggestion that the enzyme was being degraded by proteases. Addition of a protease inhibitors cocktail at 1 µM concentration completely inhibited esterase activity. The loss of activity was suggested to be due to the binding of serine protease inhibitors to the EstEFH5 catalytic serine. As a result, the enzyme preparation was used for biochemical characterization immediately after purification.

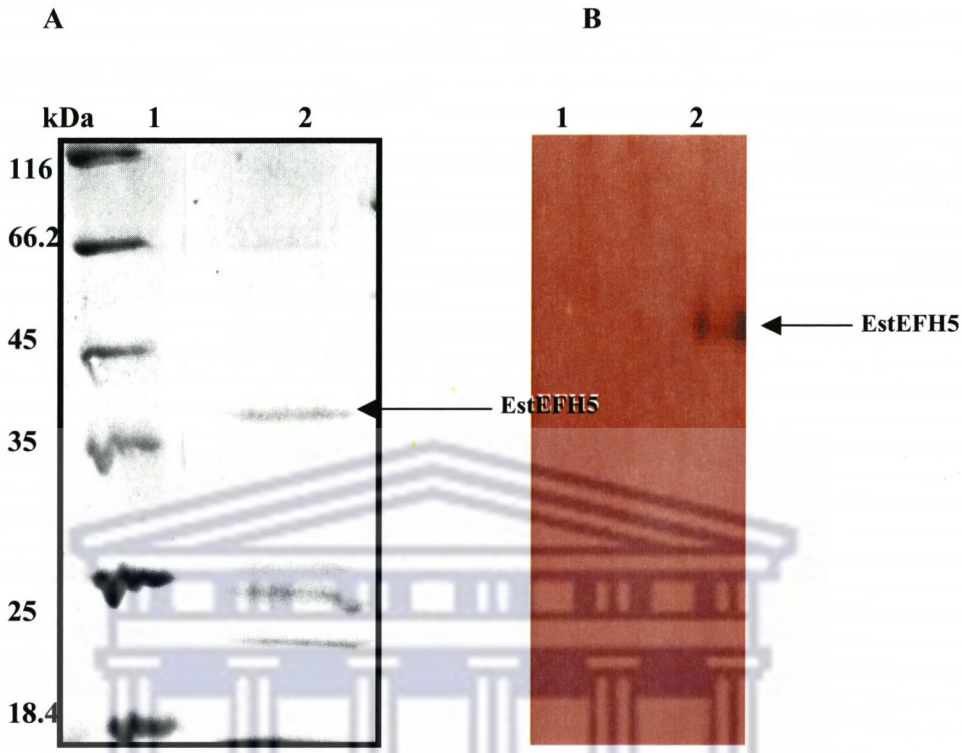


Figure 5.24: (A) Silver stained (12%) SDS-PAGE analysis of (1) Molecular weight marker (2) partially purified EestEFH5 (B) Activity stain assay of Lane 1: *E. coli* BL 21(soluble fraction) (2) partially purified EstEFH5.

Table 5.6: Esterase activity remaining after storage in the presence and absence of glycerol.

Time (Hours)	4 °C Activity (%)		-80 °C Activity (%)	
	A	B	A	B
12	30	60	36	90
24	BQL	43	12	90
36	BQL	12	BQL	42

Activity immediately after purification (13.4 U mg^{-1}) was taken as 100%

A=Storage without glycerol

B=Storage with glycerol (15% v/v)

5.3.6 Biochemical characterization of EstEFH5

5.3.6.1 Substrate specificity

The substrate specificity of the partially purified EstEFH5 against different fatty acids ester was determined using ρ-nitrophenyl esters of C2-C16 (Fig. 5.25). EstEFH5 selectivity was highest against shorter acyl chain lengths. EstEFH5 showed maximum activity (100%) against ρ-NP-C3 followed by ρ-NP-C2 at 74%. The hydrolytic activity of EstEFH5 decreased drastically against medium chain length esters (C6-C10). EstEFH5 was not active against ρ-nitrophenyl esters with *n*-acyl chain lengths of greater than 10 (C14 and C16).

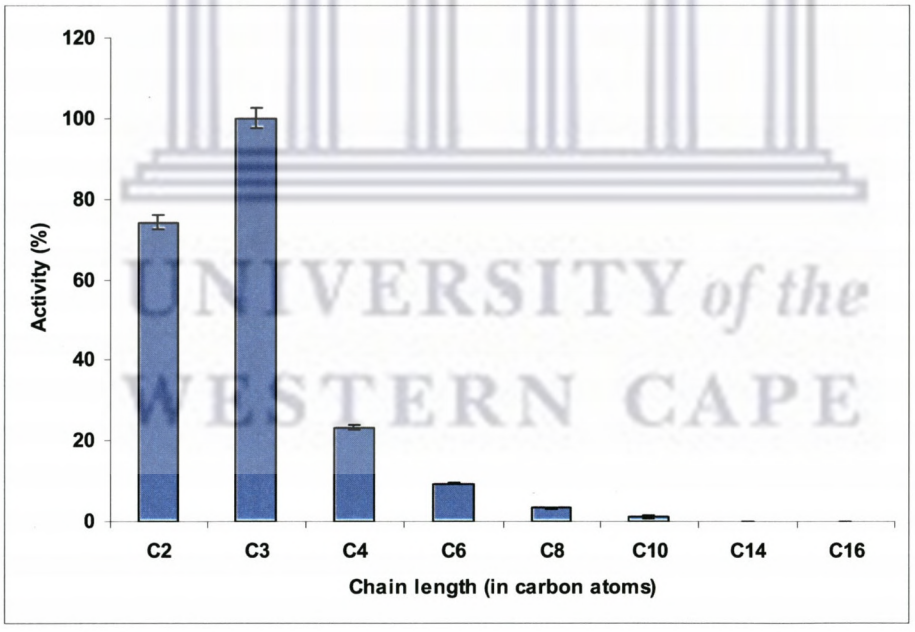


Figure 5.25: Hydrolysis of ρ-nitrophenyl fatty acid esters by EstEFH5 esterase (activity against ρ-NP-C3 was taken as 100%).

5.3.6.2 EstEFH5 reaction kinetics

Using two readily hydrolysed ρ-nitrophenyl ester substrates (C2 and C3), kinetic constants for EstEH5 were calculated (Table 5.7) using the Michaelis-Menten non-linear regression hyperbola plot in GraphPad Prism version 4.0 (Appendix 5.III). The data for the hydrolysis of ρ-NP-C2 and ρ-NP-C3 by EstEFH5 were also fitted on the Lineweaver-Burk plot for display purposes, in order to compare the non-linear regression data with those from the linear regression plot as shown in Appendix 5.IV. The K_M and V_{max} values for ρ-NP-C2 were approximately 1.7-fold lower than that of ρ-NP-C3. Catalytic turnover (k_{cat}) for ρ-NP-C2 was 1.6-fold higher than that of ρ-NP-C3. Specificity constant (k_{cat}/K_M) values for the two substrates suggested that ρ-NP-C3 was a marginally preferred substrate.

Table 5.7: Kinetic parameters for hydrolysis of ρ-nitrophenyl esters by EstEFH5 esterase

	ρ-NP-C2	ρ-NP-C3
K_M (mM)	9.2 ± 1.2	5.3± .8
V_{max} ($U.mg^{-1}$)	198.9 ± 13.2	125 ± 9.0
$k_{cat}(S^{-1})$	94.7 ± 6.3	59.5 ± 4.3
k_{cat}/K_M (s^{-1} mM)	10.3	11.2

Enzyme concentration in the assay= 3 μg ml^{-1}

K_{cat} was calculated assuming a molecular weight of 35 kDa as estimated SDS-PAGE analysis, and a single site per monomeric protein.

5.3.6.3 EstEFH5 activity against ethyl ferulate

EstEFH5 activity against ethyl ferulate was investigated using qualitative plate assay. The partial purified EstEFH5 after refolding was concentrated and immediately used for activity assay. After 12 h incubation at 25 °C, EstEFH5 showed a zone a clearance (Figure 5.26), indicating ethyl ferulate hydrolyzing activity of this enzyme. These observations further supported early library screening assay (Chapter 4, Section 4.2) which indicated that pTEND5 clone harboured a gene encoding ethyl ferulate hydrolysing activity.

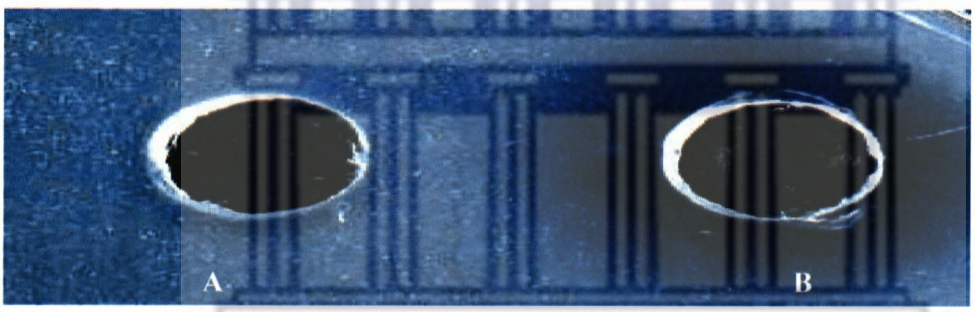


Figure 5.26: Hydrolysis of ethyl ferulate by EstEFH5 (A) Heat inactivated EstEFH5, (B) partially purified EstEFH5 (120 μg ml^{-1})

5.4 Lipase expression

E. coli transformant (pRASH14) showed lipolytic activity during library screening. Sequence analysis revealed that a lipase gene was encoded in an operon together with a LipB chaperone gene. This section reports the design of expression constructs that would allow the expression of lipase gene (where necessary with its cognate LipB chaperone gene) in *E. coli*.

5.4.1 Expression strategy

Promoter prediction showed the presence of N-terminal leader peptide sequence (40 a.a.) with a maximum cleavage site probability of 0.876 between amino acid positions 40 and 41 (^{40}Ala-^{41}Ala) of the LipA primary structure. Previous studies have also indicated that the production of lipases from proteobacterial sub-families I.1 and I.2 in heterologous hosts require co-expression with LipB (Jorgensen *et al.*, 1991, Quyen *et al.*, 1999, Quyen *et al.*, 2004).

On the assumption that expression of LipA would require co-expression with its cognate LipB but without the N-terminal leader peptide, primer sets were designed with engineered restriction sites to amplify each set of genes. The different expression constructs could be used to test the efficiency of lipase expression in *E. coli*. By engineering restriction sites into PCR primers, the LipB gene could be inserted in the correct orientation and reading frame just downstream of the RBS of pET22b (+).

5.4.2 Construction of LipA expression vectors

Five primers were constructed with engineered restriction sites (Table 5.8), designed to amplify each set of genes in the following way: Primer set LipAN/LipAH targeted the full length LipA gene; primer set LipBN/LipBH targeted the full length LipB gene; primer set LipAN/LipBH targeted the full length LipA-LipB genes; and LipSignal/LipBH targeted the LipB gene and its truncated cognate LipA gene without the leader peptide coding sequence. Restriction sites and stop codons were included in such a manner as to allow the addition of the C-terminal 6x His tags to the expressed proteins. The ATG start

codon of the LipA gene was used to create an *Nde*I site for the 5' (sense strand) insertion point whilst *Hind*III sites were added to the end of the 3' primers (anti-sense strand) of the LipA and LipB genes.

Table 5.8: Primers designed to amplify the *B. multivorans* UWC10 lipase operon (restriction sites are underlined)

Primer Name	Sequence(5'→3')	Restriction sites
LipAN	GAACATGCCATATGGCCAGATCGATGCGTTC	*Nde*I
LipAH	GACGAAGCTTCAGACGGCCCGCGAGTGCAAC	*Hind*III
LipBN	CGTCTGACATATGGCCGCCCGTGACGGGCAC	*Nde*I
LipBH	ATCAAGCTTCAACTGCGCGGTGCCCGCCGTAG	*Hind*III
Lip(Signal)	GGAATTCCATATGGCGACGCGCTATCCGATCGTC	*Nde*I

Following PCR, four PCR products of expected size (1095, 1119, 2103 and 2223 bp) were observed, corresponding to the LipA gene, LipB gene, LipA (without a leader sequence)-LipB genes and full length LipA-LipB genes, respectively (Fig. 5.27). The PCR products were purified and blunt-end cloned into the pMOS*Blue* vector yielding the following plasmids: pMOSLipA containing the full length lipase gene, pMOSLipB containing the full length LipB chaperone gene, pMOSLipAB containing the full length lipase-chaperone genes and pMOSLip$_{signal}$ containing the truncated lipase gene without the leader peptide and the LipB chaperone.

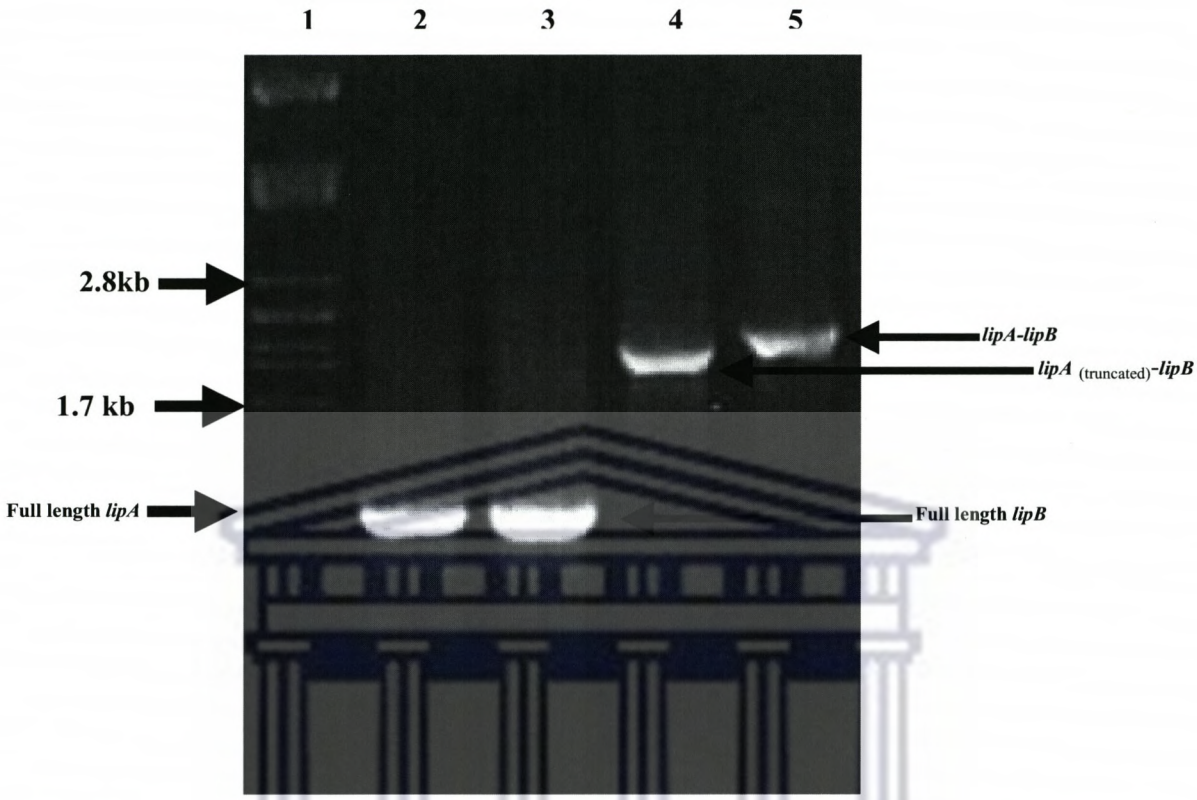


Figure 5.27: Agarose (1%) gel electrophoregram showing PCR amplification of *B multivorans* UWC10 lipase operon: Lane 1: *PstI*-digested λDNA markers; Lane 2: *lipA*; Lane 3: *lipB*; Lane 4: *lipA* without leader sequence and its *lipB*; Lane 5: full length *lipA-lipB*.

Following transformation and screening of the recombinant clones on tributryin agar plates, the only active clones identified were those harbouring the pMOSLipAB and $pMOSLip_{signal}$ plasmids. No lipolytic activity was observed in constructs where lipase was expressed without its chaperone. The inserts were excised by means of the engineered restriction sites, purified and cloned into the corresponding sites of pET22b (+) to yield pETLipAB and $pETLipase_{signal}$, carrying the full length lipase-chaperone, and

the truncated lipase and its full length chaperone, respectively. Time constraints precluded more detailed lipase expression studies.

5.5 Discussion

5.5.1 Expression and Purification of EstBL

After attempting different expression systems and conditions, a successful over-expression of EstBL was achieved with the pMS470Δ8 expression vector. The pMS470Δ8 expression vector was first described by Balzer *et al.* (1992) and has been used successfully to express a number of proteins encoded by high GC rich gene templates, including some esterase genes from the genera *Burkholderia* (Reiter *et al.*, 2000; Petersen *et al.*, 2001) and *Bacillus* (Glieder *et al.*, 2002). Expression studies revealed that inducer concentration and induction temperature were crucial factors in facilitating the production of soluble recombinant EstBL in *E. coli*. The highest specific activity (9.5 U mg^{-1}) could be detected in a soluble fraction after 14 h of growth at 18 °C, when induction with 0.1 mM IPTG was performed at a point of inoculation.

EstBL was purified to homogeneity, as judged by SDS-PAGE, using three chromatographic steps (Phenyl-Sepharose hydrophobic interaction, Q-Sepharose anion exchange and Superdex 75 gel exclusion chromatographies). It is noteworthy that Phenyl-Sepharose has been used as a primary chromatographic step to purify other EstBL homologues from *Brevi. linens* (Sakai *et al.*, 1999), *B. gladioli* (Petersen *et al.*, 2001) and *Ps. aeruginosa* (Ogino *et al.*, 2004). This chromatography step has been used successfully to purify Family VIII esterases because it takes advantage of the overall

hydrophobicity of these proteins. Although the purification of the *B. gladioli* EstB to homogeneity was achieved with two column chromatography steps (Petersen *et al.*, 2001), the purification to homogeneity of BDA, Lip8 and EstBL (this study) required a minimum of three column chromatography steps (Skai *et al.*, 1999; Ogino *et al.*, 1994).

The purified EstBL exhibited a single band on SDS-PAGE, determined to be approximately 42 kDa from a logarithmic plot of molecular weight versus mobility, using standard proteins. The estimated molecular weight coincided with the calculated molecular weight from the translated nucleotide sequence. The native molecular weight of EstBL was estimated to be around 42 kDa under non-denaturing PAGE analysis. As mentioned previously (Section 5.3.6), it is acknowledged that molecular weight estimation on native PAGE is generally inaccurate in predicting the true globular state of the protein. Size exclusion chromatography in the presence of molecular weight standard markers would have provided an accurate estimation. Nevertheless, the observations are consistent with data from previously purified recombinant esterases from Family VIII which exhibited monomeric states with molecular weights between 40 and 42 kDa. These include esterases from *Arthrobacter globiformis* (Nishizawa *et al.*, 1995), *Brevi. linens* (Sakai *et al.*, 1999), *B. gladioli* (Petersen *et al.*, 2001) and *Ps. aeruginosa* (Ogino *et al.*, 2004).

5.5.2 pH stability and thermostability of EstBL

There is a widely held view that thermostability of the enzyme generally reflects the growth temperatures of the parent organism (Cowan *et al.*, 1997). EstBL thermostability data reflected that of the mesophilic parental host (*B. multivorans*). EstBL was stable

between 20-30 °C, with more that 100% of activity retained after 30 min of incubation at pH 7.5 (Fig. 5.11). These observations were in accordance with the thermostability data from other family VIII esterases. For example, Lip8 from *Ps. aeruginosa* retained >90% of activity after 10 min of incubation between 20-30 °C at pH 8.0 (Ogino *et al.*, 2004), while EstB from *B. gladioli* (Petersen *et al.*, 2001) retained 80% of the activity at 25 °C after 20 min of incubation.

It has been suggested, with limited experimental evidence that proteins are more stable within the vicinity of their pI values (Pace *et al.*, 1990). pH stability profile showed that EstBL was stable between pH range of 7 and 8.5 (well below its predicted pI value of 5.5), with greater than 95% of activity retained after 30 min (Fig. 5.12). EstBL lost about 50% and 75% of activity in the acidic (pH 6.0) and alkaline ranges respectively. In contrast to EstBL, the pH-stability data of EstB from *B. gladioli* showed this enzyme was stable in wider pH ranges with more than 80% of activity retained between pH 6.0 and pH 9.0 (Petersen *et al.*, 2001). Since, EstBL was stable at neural pH, the observed loss of activity in the acidic and alkaline range could be attributed primarily to proton activation and deactivation of prototropic groups (within the active site) in this pH ranges.

5.5.3 Substrate specificity

The presence of the SxxK motif, which is highly conserved across the entire family of β-lactamases (Knox *et al.*, 1996), was also deduced in the primary structure of EstBL. Previous reports have shown the presence of esterolytic activity in β-lactamases (Kelly *et al.*, 1986) while an esterase with β-lactam hydrolyzing activity has also been reported

(Jone and Page, 1991). Based on these observations, the possibility that EstBL could be a bifunctional enzyme with β-lactam hydrolyzing activity was assessed. Tests with EstBL using β-lactam substrates revealed that EstBL did not possess any detectable β-lactamase activity. Similar results were observed with BDA esterase from *Brevi. linens* against penicillin derivatives (Sakai *et al.*, 1999) and EstB from *B. gladioli* against cephalosporin and penicillin derivatives (Petersen *et al.*, 2001).

There are number of properties that determine the substrate specificity of esterolytic enzymes, some of which include differences in the size and overall hydrophilicity or hydrophobicity of the substrate binding pocket (Jaeger *et al.*, 1999). The substrate specificity of EstBL was also studied using ρ-nitrophenyl and β-naphathyl esters of different acyl chain length. The increase in acyl chain length was marked with decrease in EstBL activity. The preference of EstBL for short chain length substrates suggested that the hydrophobic binding pocket of the EstBL can only accommodate a limited range of carbon atoms, a feature which is typical for many carboxylesterases (Khalameyzer *et al.*, 1999; Jaeger *et al.*, 1999; Petersen *et al.*, 2001; Bornscheuer, 2002; Chao *et al.*, 2003; Ogino *et al.*, 2004). These observations also suggested that EstBL is a "true" carboxyl esterase. Carboxylesterases have been differentiated from lipases based on the substrate preference, with esterases typically showing preference for shorter n-acyl chain lengths (Jaeger *et al.*, 1999). Specificity constant (k_{cat}/K_M) ratios for EstBL against ρ-NP esters (C2-C4) revealed that ρ-NP-C2 was the best substrate.

5.5.4 Effect of inhibitors

Enzyme activity may be affected by many extrinsic factors including organic solvents, divalent metal ions, detergents and non-ionic surfactants (Kademi *et al.*, 2000; Prim *et al.*, 2000; Persaresi *et al.*, 2005; Kim *et al.*, 2005). Analysis of the primary structure of EstBL revealed two cysteine residues (Cys283 and Cys384) which could potentially form a disulphide bond (Chapter 4, Section 4.5.3). However, EstBL did not show any loss of activity in the presence of DTT at 1 mM. These observations suggested that either no functional disulphide bond existed or that the bond was not readily accessible (i.e., not surface exposed) for reacting with the reducing agent. Sequence alignments of EstBL and other family VIII esterase revealed that the positions of the two cysteine residues were not conserved suggesting that they may not be functionally or structurally important.

The addition of divalent metal ions at 1 mM showed that EstBL was significantly inhibited by Ag^{2+}, Cu^{2+} and Fe^{2+}, but retained more than 95% of initial activity in the presence of Ca^{2+}, Li^{2+} and Mg^{2+}. The effect of divalent metal ions appears to be enzyme dependent. For example, Ca^{2+} and Mg^{2+} have been shown to activate an esterase from *Ps. aeruginosa* (Persaresi *et al.*, 2005), while Co^{2+} and Zn^{2+} significantly inhibited the activity. In general, these observations suggest that EstBL does not require metal ions as cofactors, which is a typical characteristic of esterases (Bornscheuer, 2002). EstBL retained more that 95% initial activity in the presence of the metal chelator EDTA (1 mM). This observation further supported the assumption that the catalytic action of EstBL might not involve metal ions. The two EstBL homologues (BDA and EstB) were not affected by EDTA at the same concentration (Sakai *et al.*, 1999; Petersen *et al.*,

2001), suggesting no essential role of metal ions in the catalytic mode of the family VIII esterases.

EstBL was significantly inhibited by PMSF, with more than 85% loss of activity at 1mM. Similar inhibitory effects have been reported with other Family VIII esterases. For example, about 97% of *B. gladioli* EstB activity was lost at 1 μM PMSF (Petersen *et al.*, 2001), while BDA esterase from *Brevi. linens* lost 54% of activity in 1 mM PMSF (Sakai *et al.*, 1999). The inhibition of EstBL by PMSF suggested a serine-dependent catalytic mechanism in this enzyme. PMSF inhibits many serine hydrolases by covalent linkage to the activated serine hydroxyl, thereby mimicking the first transition state in ester hydrolysis (Sweeny and Maxwell, 1999). This in turn prevents formation of the hydrogen bonding network which is critical for catalytic action of many (if not all) serine hydrolases.

Addition of various detergents revealed that EstBL was significantly inhibited in the presence of ionic detergents. More than 90% of the EstBL activity was lost in the presence of SDS (0.5%). Similar observations were noted with EstB from *B. gladioli* (Petersen *et al.*, 2001). In contrast to ionic-detergents, EstBL retained significant activity (>90%) in the presence of non ionic Tween 20 and Triton X-100 surfactants (0.5% v/v). The activity of a recently reported esterase (Est25) derived from a metagenomic library was shown to be enhanced in the presence of Tween and Triton surfactants (Kim *et al.*, 2005).

5.5.5 EstEFH5 expression and *in vitro* refolding

Several attempts were made to over-express EstEFH5 in a biologically active form. Initial expression attempts showed that EstEFH5 activity could be detected in the supernatant fraction, although SDS-PAGE analysis showed no corresponding protein band in the soluble fraction. However, a new major protein band was observed in the pellet fraction, suggesting that the protein was accumulating in an insoluble form.

From these observations, it was concluded that either the expressed protein was largely unfolded and inactive (presumably present as inclusion bodies) or was correctly folding but anchored to the membrane in active form. In order to address these questions, an expression construct lacking the 30 N-terminal amino acids corresponding to the putative membrane domain was designed. The removal of this domain resulted in complete loss of esterase activity, stressing the importance of this domain in maintaining the catalytic function or folding of this enzyme. However, the protein was still found in the membrane fraction after the removal of the putative membrane anchor domain. Attempts to use detergents to facilitate the recovery of the enzyme from the membrane were unsuccessful. All the detergents tested led to a complete inhibition of esterase activity.

In vitro refolding experiments were attempted to recover the EstEFH5 from the insoluble form. A 6x His tagged expression construct was designed to allow single step column refolding and purification using the His select™ Nickel column chromatographic procedure. *In vitro* refolding attempts based of previously published reports were unsuccessful (Clark., 1998; Quyen *et al.*, 1999; Chao *et al.*, 2003; Siritapetawee *et al.*,

2004). Furthermore series of *in vitro* refolding trials showed that glycerol was necessary to stabilise the refolding EstEFH5. Partially purified EstEFH5 was obtained after column refolding in the presence of glycerol. The semi-purified enzyme had a specific esterase activity of 13.4 ± 1.67 U mg^{-1} against ρ-nitrophenyl acetate as a substrate. The protein band of the predicted size could be observed on SDS-PAGE after silver staining.

5.5.6 EstEFH5 catalytic properties

Substrate specificity studies indicated that EstEFH5 was a typical carboxylesterase showing a strong preference for shorter acyl chain length substrates (Jaeger *et al.*, 1999). The specificity constant of EstEFH5 against ρ-NP-C3 and ρ-NP-C2 were not significantly different, suggested that the two substrates could be converted with similar efficiency. It must be emphasized that the enzyme solution used to characterize EstEFH5 was only partial pure. Although this might not significantly impact on K_M values, the calculated turnover number would be substantially below the true value.

Furthermore, partially purified EstEFH5 exhibited ethyl ferulate hydrolyzing activity on agar plate assay. However, instability of this enzyme after purification and time constraints prevented further detailed studies.

CHAPTER SIX

Homology modelling of the 3D structures of *Burkholderia multivorans* UWC10 lipase and esterase

6.1 Introduction

Homology modelling can be defined as the prediction of three dimensional structure of a target protein from the amino acid sequence (primary structure) of a homologous (template) protein for which an X-ray or NMR structure is available (Bowie *et al.*, 1991). The built protein model provides a wealth of information of how the protein functions with information at residue property level. This information may then be used for mutational studies or drug design.

Homology modelling based on threading (Domiques *et al.*, 2001) and comparative (Blundell *et al.*, 1987) approaches rely on detectable similarity spanning most of the modelled sequence and at least one known 3D-structure. Comparative or homology protein structure modelling builds a three dimensional model for a protein of unknown structure (the target) based on the one or more proteins of known structure (the templates) (Sali and Blundell, 1993). This is possible when there is (i) detectable similarity between the target sequence and the template structures (at least 20% a.a. identity) and (ii) availability of correct alignment between them.

All homology modelling methods consist of five sequential steps: the first step is to search for proteins with known 3D structures which are related to the target sequence. The second step is to choose the structures that would be used as templates. The third step is to align their sequences with the target sequence. The fourth step is to build a model for the target sequence given its alignment with the template structures. The last step is to evaluate the model using variety of criteria (Sali *et al.*, 1990).

The aim of this chapter was to use deduced primary structures (Chapter 4) to build 3D structural models, with an objective to identify residues that could be targeted for protein engineering.

6.2 LipA 3D structural model

6.2.1 Building of LipA 3D structural model

A homology-based model of LipA was constructed using a lipase template of high resolution crystal structure from *Chromobacterium viscosum* (PDB code: 1CVL), a member of the αβ hydrolase family to which LipA shares 78% sequence identity (Lang *et al.*, 1996). The model was generated by threading technique using GenTHREADER (Jones, 1999) and FUGUE (Shi *et al.*, 2001). A secondary structure-driven sequence alignment between LipA and 1CVL which was used to generate the model is shown in Figure 6.1. A total of 5 models were generated and the best model was chosen based on lowest energy (1444163.59=Modeller objective function) output from program, MODELLER (Sali, 1995).

The model validation using RAMPAGE (Lovell *et al.*, 2001) indicated that the final model had good geometry, with less than 1.9% of residues in the disallowed region, 97.2% residues in the most favoured region and only 0.9% in the outlier region (Figure 6.2).

```
_aln.pos          10        20        30        40        50        60
1CVL       ADTYAATRYPVILVHGLAGTDKFANVVDYWYGIQSDLQSHGAKVYVANLSGFQSDDGPNGRGEQLLAY
LipA       ADDYAATRYPIVLVHGLTGTDKYAGVLEYWYGIQEDLQRHGATVYVANLSGFQSDDGPNGRGEQLLAY
 _helix                                  9999999999                  999999999
 _beta             99999999                        99999

 _aln.p    70        80        90       100       110       120       130
1CVL       VKQVLAATGATKVNLIGHSQGGLTSRYVAAVAPQLVASVTTIGTPHRGSEFADFVQDVLKTDPTGLSS
LipA       VKQVLAATGATKVNLIGHSQGGLTSRYVAAVAPDLVASVTTIGTPHRGSEFADFVQSVLAYDPTGLSS
 _helix    99999999          999999999999              9999999999999       9
 _beta               99999999                  999999

 _aln.pos  140       150       160       170       180       190       200
1CVL       TVIAAFVNVFGTLVSSSHNTDQDALAALRTLTTAQTATYNRNFPSAGLGAPGSCQTGAATETVGGSQH
LipA       SAIAALVNVFGILTSSSHNTNQDALAALTTLTTAQAATYNRNFPSAGLGAPGSCTTGALTETVGGNTH
 _helix    99999999999999    9999999999    99999999999
 _beta                                                                  9999

 _aln.pos    210       220       230       240       250       260       270
1CVL       LLYSWGGTAIQPTS---TVTGATDTSTGT-LDVANVTDPSTLALLATGAVMINRASGQNDGLVSRCSS
LipA       LLYSWAGTAIQPTNAAFGVTGASDTSTIPVVDPANALDASTLALLGTGTVMINRGSGQNDGLVSKCSA
 _helix                                          9999999999999999      99999
 _beta     9999                  99999

 _aln.pos       280       290       300       310       320
1CVL       LFGQVISTSYHWNHLDEINQLLGVRGANAEDPVAVIRTHVNRLK-LQGV
LipA       LYGQVLGTNFKWNHLDEINQLLGVRGAYAEDPVAMIPHAREPVATLAGV
 _helix    9           9999999    -       99999999999999 9
 _beta
```

Figure 6.1: A secondary structure driven alignment between LipA and the structural template, 1CVL. The alignment was used to generate the LipA 3D model structure.

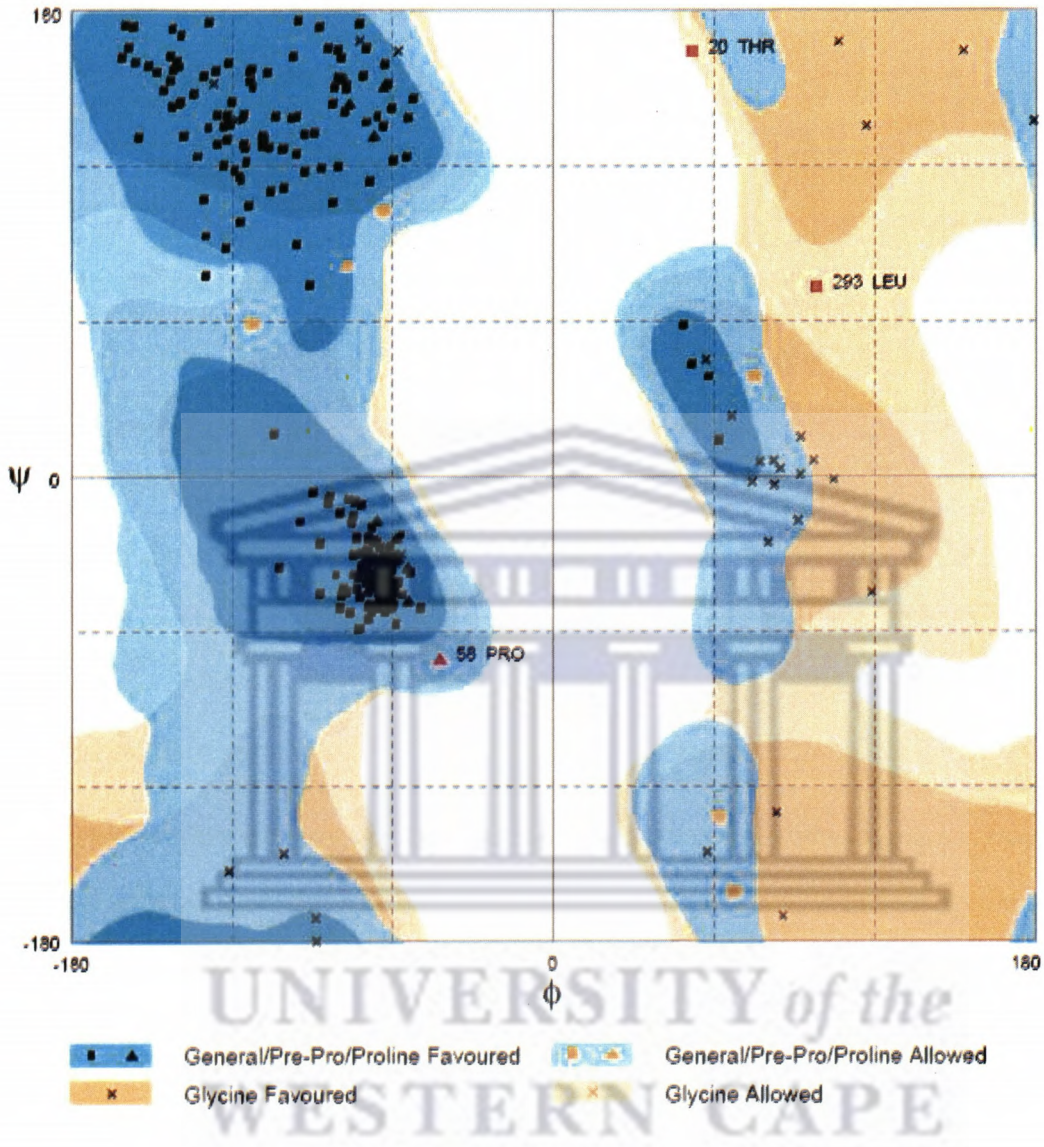


Figure 6.2: Ramachandran plot of the amino acids residues as predicted from the 3D model for *Burkholderia multivorans* lipase (LipA).

Analysis of the LipA model structure for the topological location of the residues falling within the outlier region revealed that these residues (Thr20, Pro58 and Leu 293) were all located in external loops. The general consensus is that protein structures are generally conserved more than are their sequences (Jones, 1999; Domiques *et al.*, 2001). It was therefore decided to ignore these residues on the assumption that they correspond to deletions and insertions of residues that do not disturb the essential core of the protein fold (Fiser *et al.*, 2000).

6.2.2 Topology description of LipA model

Figure 6.3 shows an overall LipA 3D structural model. Topology prediction of the LipA 3D model structure was performed using TOPS server (Michalopoulus *et al.*, 2004), available online at (tops.leeds.ac.uk). As expected the model structure is typical of protein in the α/β fold family. This family of proteins, which is believed to have evolved by divergent evolution, adopts a characteristic core of β-sheets connected by loops to α-helices despite a lack of significant sequence similarity (Ollis, 1992). The model predicts that LipA is composed of eight β-sheets formed by residues β1 (Ile11-Val14), β2 (Thr42-Val45), β3 (Val80-His85), β4 (Ser106-Ile110), β5 (The196-Val198), β6 (Asn202-Trp209), β7 (Thr214-Pro216), and β8 (Ala226-Asp228) and 13 α-helices formed by residues α1 (Ile33-His40), α2 (Arg61-Thr76), α3 (Glu88-V99), α4 (Pro101-Leu103), α5 (Asp118-Tyr129), α6 (Ser137-Leu149), α7 (Thr156-Leu167), α8 (Thr169-Asn178), α9 (Pro237-Leu241), α10 (Ala243-Arg258), α11 (Lys269-Ala272), α12 (His286-Ile290), and α13 (Pro304-Ala319).

The model reveals that the LipA centre is made of a large hydrophobic sheet consisting of the five parallel strands (β6, β4, β3, β2, and β1) and an anti-parallel strand, β5. The only difference in topological arrangement between the template and the LipA model was the fact that β6 strand in the template formed an anti-parallel strand. Furthermore, the central β-type hydrophobic core of the LipA is covered on one side by helices α1 and α13 and on the other side by helices α2, α3, α7 and α10 as previously shown by 1CVL structure. The other five α-helices (α4, α5, α6, α8, α9) and antiparallel β strands (β7 and β 8) are predicted to form the LipA substrate binding domain on the template structure.

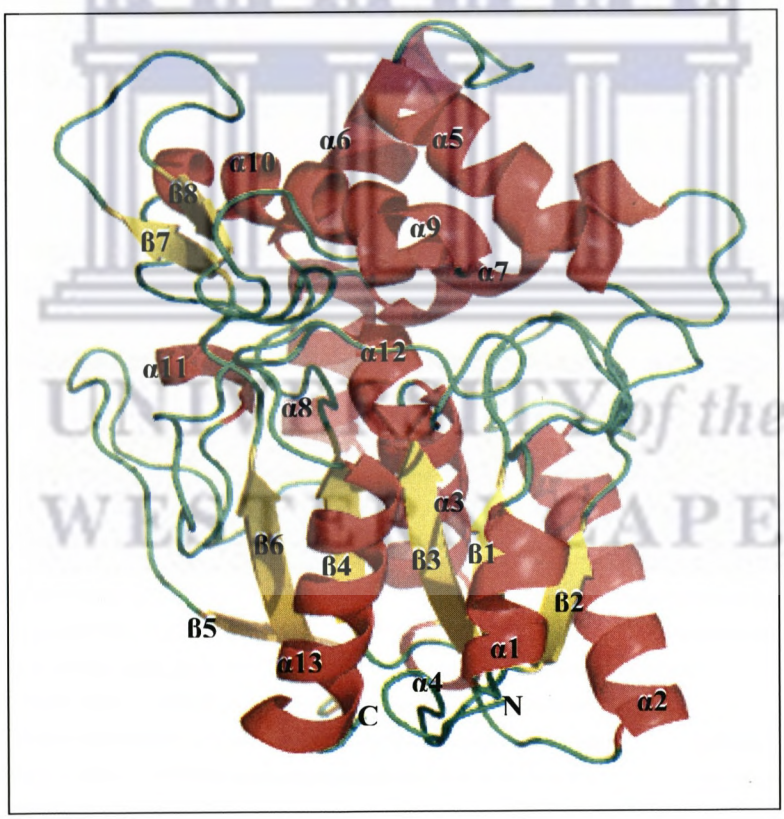


Figure 6.3: A cartoon representation of the overall LipA structural model.

6.2.3 Catalytic triad residues and the lid domain

The spatial arrangement of the catalytic triad is well conserved among the Group I proteobacterial lipases and LipA appears to be no exception. Based on the alignment with 1CVL 3D structure, it could be deduced that residues Ser87, His285 and Asp264 in the LipA model structure (Figure 6.4), form the catalytic triad. These residues have been shown to form hydrogen bonding arrangements that are characteristic for lipases, esterase and many other hydrolytic enzymes (Kraut, 1977): the $N^{\varepsilon 2}$ (His) is bound to O^{γ} (Ser), while the $N^{\delta 1}$ of the imidazolium group is hydrogen bonded to the $O^{\delta 1}$ of Asp (Lang *et al.*, 1996). This pairing is believed to modify the nucleophilicity of the His-bound active residue. The putative catalytic serine (S87) of LipA is located at the turn between the strand (β3) and the α-helix (α3) similar to other lipases. This secondary structure arrangement of a strand-loop-sheet is also referred to as a "nucleophilic elbow" (Figure 6.4). This "nucleophilic elbow" can only be when residues close to the catalytic serine have short side chains, explaining the presence of glycine residues around this region. Gly85 and Gly90 are located at the strand β3 and the α-helix α3 respectively in LipA model.

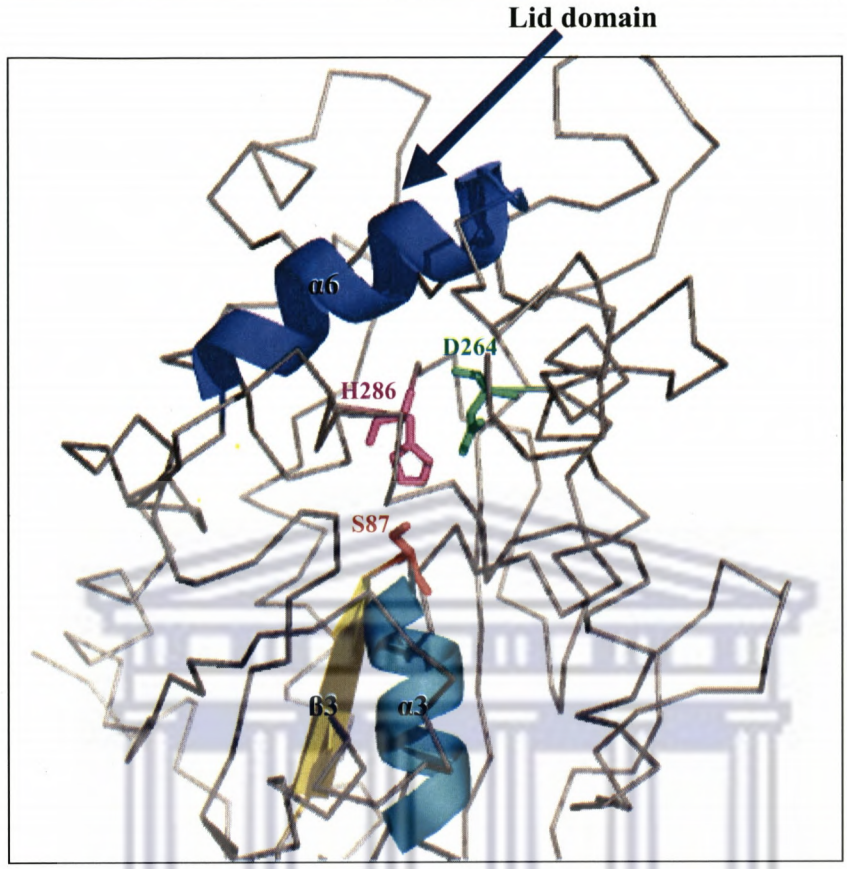


Figure 6.4: Stick representation of the catalytic triad residues of the LipA structural model. The lid domain is represented by α5-helix and is coloured blue.

The LipA model revealed that the catalytic triad residues were not surface-exposed but covered by lid-like α-helical structure (α6) (Figure 6.4). This helix have been termed the lid domain, that makes the active site inaccessible to the substrate unless is displaced by the interaction with aggregated substrates or lipid surfaces (Verger *et al.*, 1997). The only difference between the template and the LipA model was that α-helix (α5) covered the catalytic triad as opposed to α6 helix in the LipA model structure. Furthermore, in LipA α5-helix is the shortest while it is the longest in ICVL. Recent studies have shown that

mutations in the "lid" region affect chain length specificity and thermostability of *Pseudomonas fragile* lipase (Santarossa *et al.*, 2005).

6.2.4 The oxyanion hole

Lipases are also characterised by the presence of an oxyanion hole, an arrangement of amino acids, which plays an important role in the stabilization of the tetrahedral intermediate. The intermediate is transiently formed during the hydrolysis reaction through hydrogen bonding to main chain donors (Lang *et al.*, 1996). From the enzymology data of serine proteases and many other lipases, it is known that the negative charge of the tetrahedral intermediate must be stabilised by hydrogen bonding (Jaeger *et al.*, 1997). Based on the template structure, it could be predicted that in the LipA model, this region is also stabilised by a hydrogen bond at His15. The two other residues that might contribute to the formation of the stabilizing configuration include Leu17 and Gly51, as shown in the 1CVL structure (Lang *et al.*, 1996).

6.2.5 Disulphide bridges and calcium binding sites

The LipA 3D model also predicts that two cysteine residues (Cys190 and Cys270 located at the turn and α9-helix, respectively), could potentially form a disulphide bridge, based on MODELLER disulphide bridge restraints command (Figure 6.5). The predicted distance between the two cysteine residues was 1.9 Å which is in accordance with the data from the template and other lipase structures (Noble *et al.*, 1993; Lang *et al.*, 1996; Kim *et al.*, 1997a; Nardini *et al.*, 2000). Furthermore, these cysteine residues were

surface exposed, consistent with other observations from the 3D structures of the template structure (Lang et al., 1996) and other *Burkholderia* lipases (Noble *et al.*, 1993; Kim *et al.*, 1997a; Lang *et al.*, 1998). Enzymology studies showed that addition of DTT to the reaction significantly reduced lipase activity (Jaeger *et al.*, 1993). In *Ps. aeruginosa* lipase, cysteine residues forming disulphide bond were shown to be buried within the enzyme. The treatment of this enzyme with DTT did not significantly reduce activity (Nardini *et al.*, 2000).

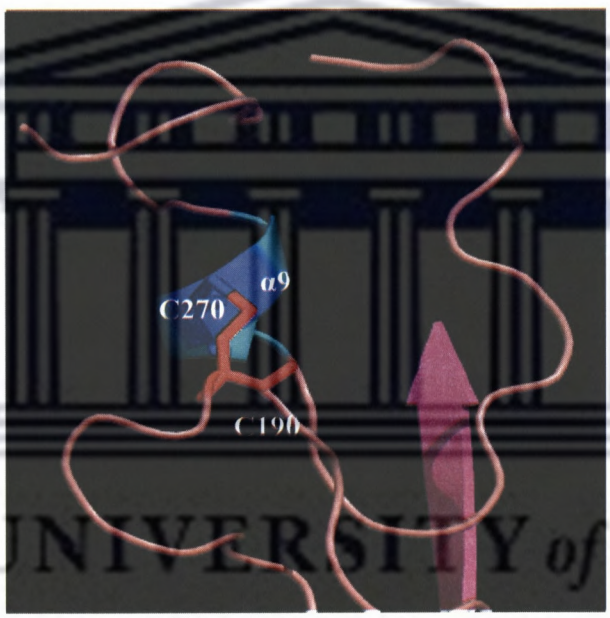


Figure 6.5: Schematic representation of the two the cystiene residues that could potentially form disulphide bridge in LipA 3D model. Cysteine residues are shown as sticks and coloured as follows: C190=red and C270 blue respectively.

Some lipases have been shown to be activated in the presence of calcium ions (Jaeger *et al.* 1997a). A role of calcium ions in stabilising lipase structures has been proposed (Noble *et al.*, 1993). The calcium binding pockets have been identified in all solved crystallographic structures from Group I proteobacterial lipases. Consequently, the LipA

3D model was analysed for the presence of calcium binding pockets. Guided by the template structure, aspartic residues (D241-D288) in the LipA model are located such that they might interact with the metal via their side chain carboxylate groups (Figure 6.6). Two other amino acids (H286 and R298) also appeared to have a suitable orientation of their backbone carbonyl groups so as to bind a metal ion. The distance between the residues forming the calcium binding pocket to the actual calcium ligand could not be predicted due to technical failure to integrate the calcium ligand into the model. But from the template and other lipase structures the distances to calcium ligand are within 2.0 to 3.0Å range (Noble *et al.*, 1993; Lang *et al.*, 1996; Kim *et al.*, 1997; Nardini *et al.*, 2000).

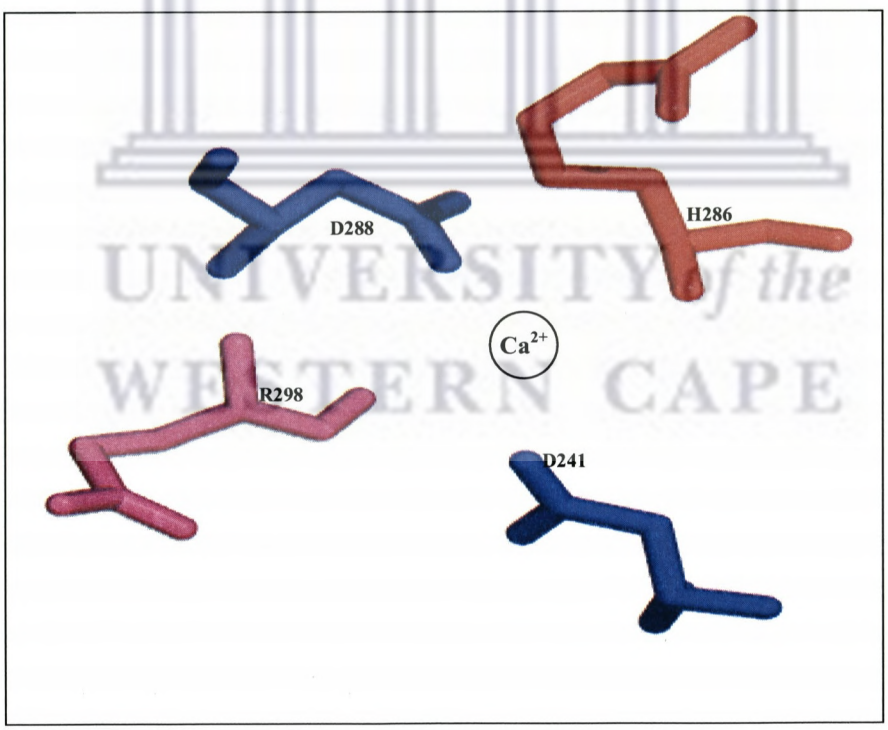


Figure 6.6: Stick representation of residues in forming a putative calcium binding pocket in the LipA 3D model.

6.3 EstBL 3D model

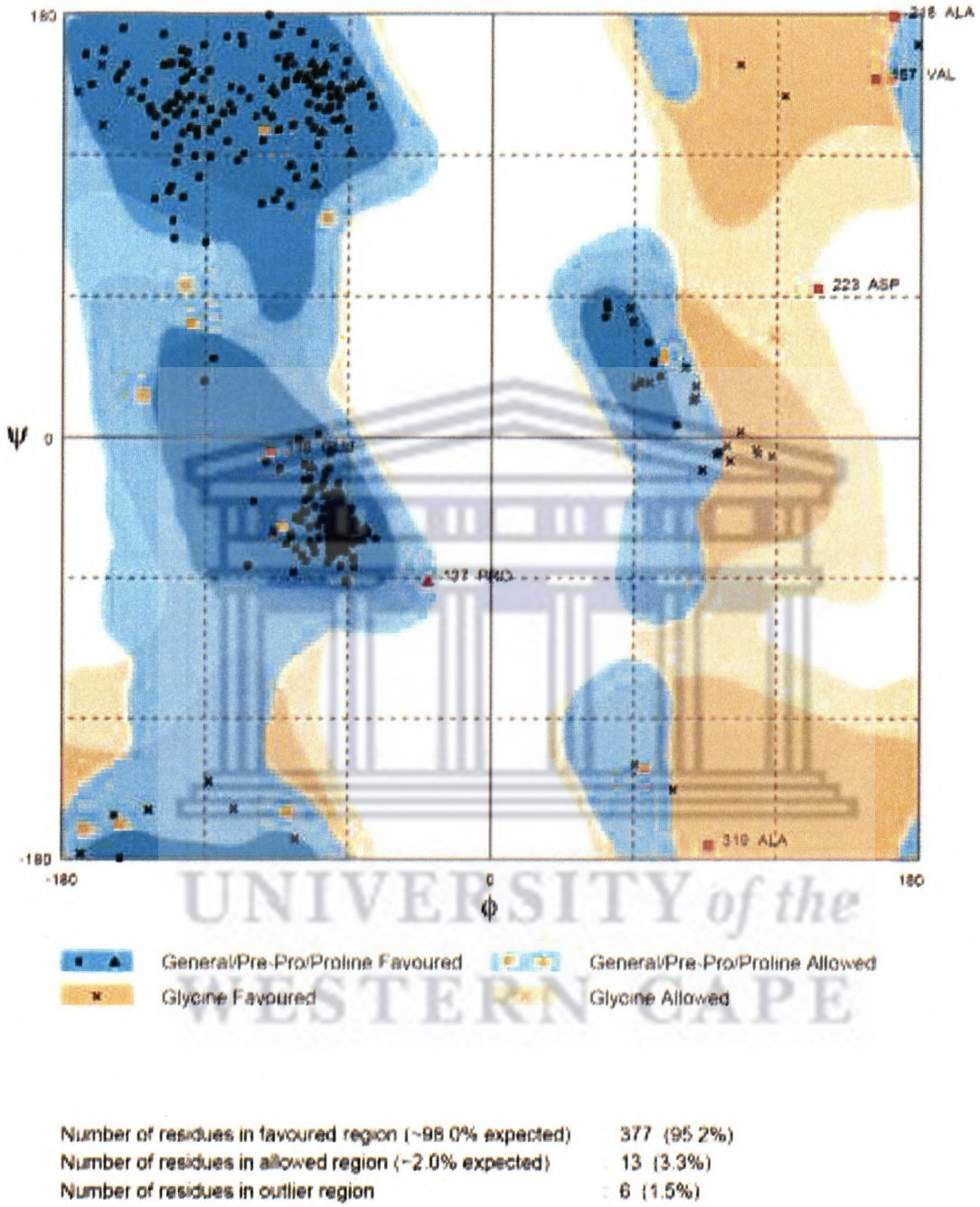
6.3.1 Model building and validation

The EstBL 3D model was constructed using an esterase from *Burkholderia gladioli* (PDB: 1CI8) structure as a template, with which EstBL shares 52% a.a. sequence identity (Wagner *et al.*, 2002). The alignment was generated (Figure 6.7) by the FUGUE and GeneTHREADER subjected to external modelling and validated as previously described in Section 6.2.1. A Ramachandran plot (Figure 6.8) showed good geometrical output for a high quality model, giving of 95% of residues in favoured regions, 3.3% in allowed regions and only 1.5% in outlier regions. All six residues in the outlier region were located in external loops.

Figure 6.9 shows the cartoon representation of the EstBL 3D model structure. Topological predictions based on TOPS suggest that EstBL model consists of a nine-stranded anti-parallel β-sheet. The identified strands were β1 (G36-R41), β2 (V46-A53), β3 (V244-P248), β4 (G255-F258), β5 (F322-F324), β6 (S329-F321), β7 (G344-G350), β8 (G352-D359) and β9 (G363-S370). The nine β-strands were surrounded by seventeen α-helices: α1 (V15-d30), α2 (S74-A90), α3 (V99-W102), α4 (L120-S125), α5 (P142-A146), α6 (L159-S168), α7 (L182-D196), α8 (L200-L212), α9 (A224-R226), α10 (R262-F264), α11 (A280-T293), α12 (R300-A307), α13 (A313-D315), α14 (P333-A337), α15 (R360-A362), α16 (Y374-L377) and α17 (R379-Y390). Seven of the β-strands (β2, β1, β9, β7, β6 and β5) are located at the central part of the molecule, with the α1 and α17 helices covering the "front" of the sheet.

```
_aln.pos          10        20        30        40        50        60
1ci8      ------------AASLAARLDAVFDQALRERRLVGAVAIVARHGEILYRRAQGLADREAGRPMREDT
EstBL     MSSLPVIARDSEPDVALHARLDDALERALADERIVGAVVMVARGGVLRYTRAAGLADREARTPMREDT
 _helix               9999999999999999999
 _beta                                        99999999    99999        99999 9

 _aln.p   70        80        90        100       110       120       130
1ci8      LFRLASVTKPIVALAVLRLVARGELALDAPVTRWLPEFRPRLADGSEPLVTIHHLLTHTSGLGYWLLE
EstBL     LFRLASVTKPIVTAAAMRLVAAGRIALDEPVARWLPAFRPTLRDGTPARITLRHLLSHTAGLGYRFLE
 _helix          99999999999999999                              99999999
 _beta     99999                                                          99999

 _aln.pos  140       150       160       170       180       190       200
1ci8      GA-GSVYDRLGISDGIDLRDFDLDENLRRLASAPLSFAPGSGWQYSLALDVLGAVVERATGQPLAAAV
EstBL     PDDDGPYARAGVSDGMDRTPVSLAENVRRIASVPLQFAPGTSGGYSLAIDVVGALIEAVDGRPLADAV
 _helix       9999999         999999999999          99999999999999  999999
 _beta

 _aln.pos    210       220       230       240       250       260       270
1ci8      DALVAQPLGMRDCGFVSAEPERFAVPYHDGQPEPVRMRDGIEVPLPEGHGAAVRFAPSRVFEPGAYPS
EstBL     AALVTTPLGMTDTAFVAPDATRFATPYVSTPGAPRRMADVDLVPVFDG-TIGIRFEPARAFDVDAWPS
 _helix   999999999
 _beta                          999999999  99999  999999     9999999

 _aln.pos      280       290       300       310       320       330       340
1ci8      GGAGMYGSADDVLRALEAIRANP-GFLPETLADAARRDQAGVGAETRGPGWGFGYLSAVLDDPAAAGT
EstBL     GGAGMVGTARDCVTLLDTLRTGRDGWLGRAWTDEMARAQPGAHDLRDAPGFGFGLGFSVLRDPVAAQS
 _helix         99999999999999     9999999999                        999999
 _beta       99999                                              99999  99999

 _aln.pos          350       360       370       380       390
1ci8      PQHAGTLQWGGVYGHSWFVDRALGLSVLLLTNTAYEGMSGPLTIALRDAVY-------A
EstBL     PESVGTWRGGGAYGHTWFVDRAAGLTVVALSNTLYEGLNGRIVTCVRDAVYGVGTGSAA
 _helix                                    999999999999999999
 _beta          99999     99999999  999999999
```

Figure 6.7: A secondary structure driven alignment between EstBL and the structural template, 1CI8. The alignment was used to generate the EstBL 3D model.



Number of residues in favoured region (~98.0% expected)	377 (95.2%)
Number of residues in allowed region (~2.0% expected)	13 (3.3%)
Number of residues in outlier region	6 (1.5%)

Figure 6.8: Ramachandran plot of the amino acids residues as predicted from the 3D model for *Burkholderia multivorans* esterase (EstBL).

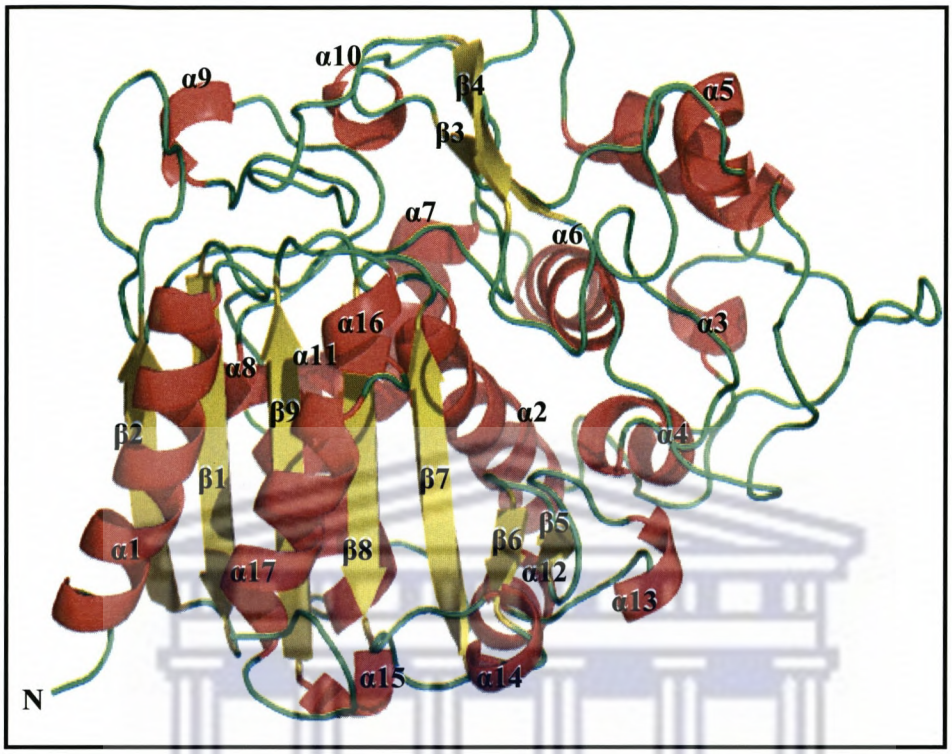


Figure 6.9: The cartoon representation of the *B. multivorans* esterase (EstBL) structural model.

6.3.2 Serine (S149) is unlikely to be putative catalytic serine in EstBL

In order to identify the putative catalytic serine in the EstBL model, a serine residue (Ser149) located within the classical GxSxG motif was considered. The model revealed that Ser149 was located at the centre of the turn/loop topology which was consistent with catalytic serines observed with many other esterases. However, some subtle differences were observed with regard to this topological location. This serine residue (Ser149) was located on an extended and twisted loop connecting two helices (α5 and α6) (Figure 6.10). This secondary structure arrangement (helix-turn-helix) is inconsistent with the structure arrangement of the "nucleophilic elbow" which normally consists of the

catalytic serine located in a short turn connecting a β-strand and α-helix (strand-turn-helix). Another observed deviation was that in the EstBL model, the surrounding glycine residues (Gly147 and Gly151), which are normally located for steric reasons on a β-strand and α-helix, respectively, at position ± 2 relative to catalytic serine, were instead both located on the same loop as Ser149. In addition, the Ser149 was surface exposed (Fig. 6.10), inconsistent with the locations of other catalytic serine residues which are normally located in the interior of the protein.

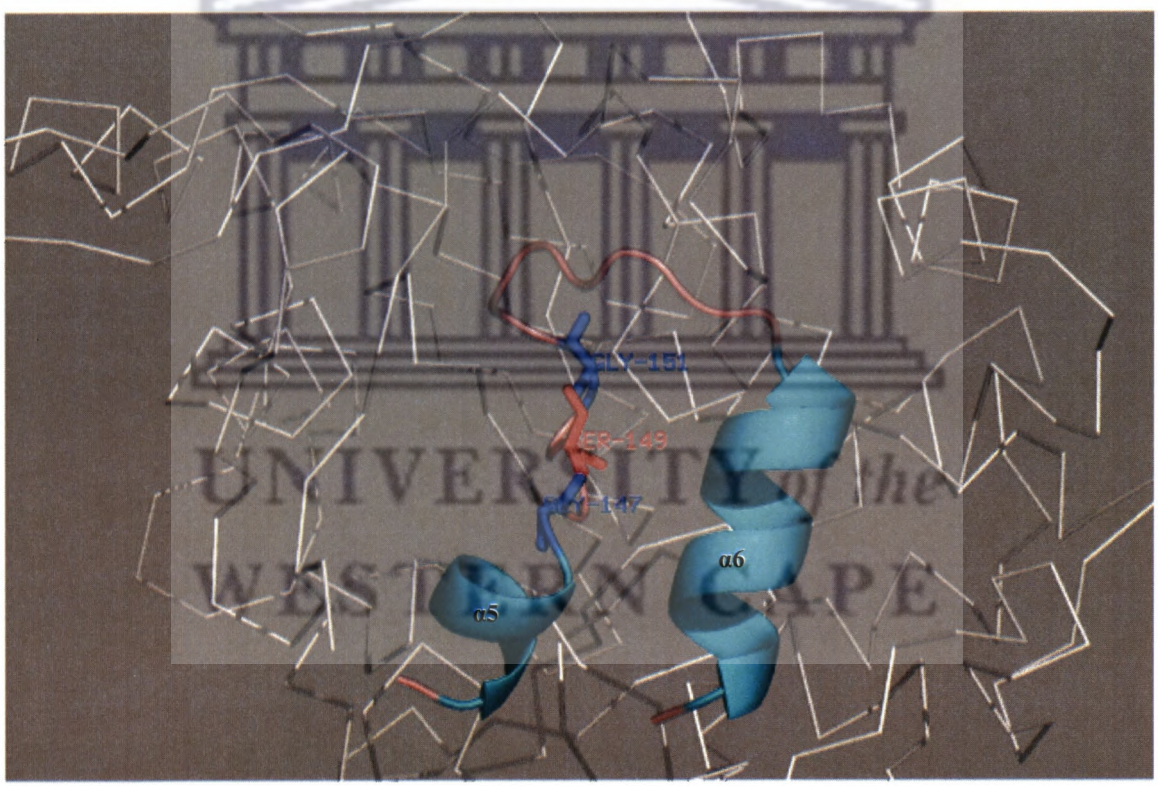


Figure 6.10: Topological location of Ser149 (stick; red) and surrounding glycine residues, G147 and G149 (stick; blue).

Furthermore, the catalytic serine within the GxSxG motif has consistently shown to be connected in a hydrogen bonding network with Asp(or Glu) and His residues, forming the Ser-Asp (or Glu)-His catalytic triad. All Asp (or Glu) and His residues within the EstBL model were individually analysed for potential hydrogen bonding with the putative catalytic serine (Ser149). None of the examined residues appeared to be close to Ser149, in a manner to allow hydrogen bond formation.

In addition, residues forming the oxyanion region could not be identified within the EstBL model. The topological and secondary structure arrangements of Ser149 led to a conclusion that this serine could not be the catalytic serine. This conclusion is further supported by site direct mutagenesis studies (Petersen *et al.*, 2001) which showed that the replacement of the corresponding serine (Ser149 in *B. gladioli* with alanine) resulted in only modest loss of activity (Petersen *et al.*, 2001).

6.3.3 Serine (Ser74) forms a putative catalytic serine in EstBL

Both the EstBL primary structure (Chapter four, Section 4.5.3) and that of the template contained a SxxK, motif which has been reported to the harbour catalytic serine in family VIII esterase (Sakai *et al.*, 1999; Petersen *et al.*, 2001), β-lactmases and some peptidases. It was therefore speculated that the esterolytic activity of EstBL which is part of the SxxK motif, exploits Ser74 for nucleophilic attack of the substrate, perhaps via a mechanism as observed in EstB from *B. gladioli* (Wagner *et al.*, 2002). An attempt was therefore made to analyse the EstBL 3D model for features that could support the role of Ser74 as the putative catalytic serine in EstBL.

The mechanism of ester hydrolysis involving Ser75 within the SxxK motif has been proposed for EstB from *B. gladioli* (Petersen *et al.*, 2001; Wagner *et al.*, 2002): In this mechanism, the Ser75 oxygen acts as a nucleophile and attacks the ester-carbonyl of the substrate, which is activated by hydrogen bonds between its carbonyl oxygen and the main-chain NH groups of residues Ser75 and Val351. The nucleophilicity of Ser75 is enhanced by Tyr181, which acts as a general base and is presumably stabilized as the phenolate group due to the proximity of the side chains of Lys78 and Typ348. The second step involves the attack of the acyl enzyme intermediate by water molecules. However, it was not clear from this proposed mechanism as to whether the water molecules involved in this deacylation step were the water molecules observed within the active site cleft of the native structure, which would then be occluded upon substrate binding, or were derived directly from solution.

The serine residues in the consensus SxxK motif act as a catalytic nucleophile in a number of esterases, peptidases and β-lactamases (Ser-75 in esterase from *B. gladioli*, Ser64 in *B. linens* esterase, Ser62 in DD-peptididase from *Streptomyces* R61 and Ser70 in *S. aureus* β-lactamase). The lysine residue in the consensus sequence of SxxK (Lys-78 in the esterase from *B. gladioli*, Lys-60 in DD-peptididase from *Streptomyces* R61 and Lys73 in β-lactamase from *S. aureus*) and tyrosine residue (Tyr-181 in esterase from *B. gladioli* and Tyr-159 in DD-peptididase from *Streptomyces* R61) are present near the active site in 3D structures of these enzymes. These residues (Ser74, Lys77 and Tyr 181 in the EstBL model) were present near the putative catalytic Ser74. Furthermore, both key residues (Ser74 and lys K77) were located in the same helix (α2), with Ser74 at the

beginning of the helix (α2) turn, consistent with the observations in other 3D structures. Due to similar spatial arrangements of the catalytic residues in EstBL model and the template, it could therefore suggested that the catalytic mechanism of EstBL could be similar to that proposed for the template (Petersen *et al.*, 2001; Wagner *et al.*, 2001).

6.3.4 Catalytic active site residues

In order to gain further understating of the residues forming the active site, an analysis of structural equivalent residues between EstBL, the esterase from *B. gladioli* and β-lactamase from *Enterobacter cloacae* was undertaken (Table 6.1). The EstBL model suggests that Lys77, Tyr132 and Tyr181 form one side of the catalytic cavity, while Trp346, Gly348 and Gly349 form the opposite wall of the catalytic cavity (Figure 6.11). Furthermore, His126 appear to be present near the active site. Equivalent residues, His-127 in esterase from *B. gladioli* and His108 in DD-peptidases from *Streptomyces* sp. R61, are also present near the active site serine in their 3D-structures, suggesting that they could form part of the active site.

Table 6.1: Equivalent residues between β-lactamase P99, esterase EstB and EstBL

P99[a]	Estb[a]	EstBL
Ser64	Ser75	Ser74
Lys67	Lys78	Lys77
Asn152	Tyr133	P88
Try150	Try181	His126
Lys315	Trp348	Gly348
Thr316	Gly349	Gly349
Gly317	Gly350	Gly350
Ser318	Val351	Trp346
		R347

([a] = data was taken from Wagner *et al.*, (2002))

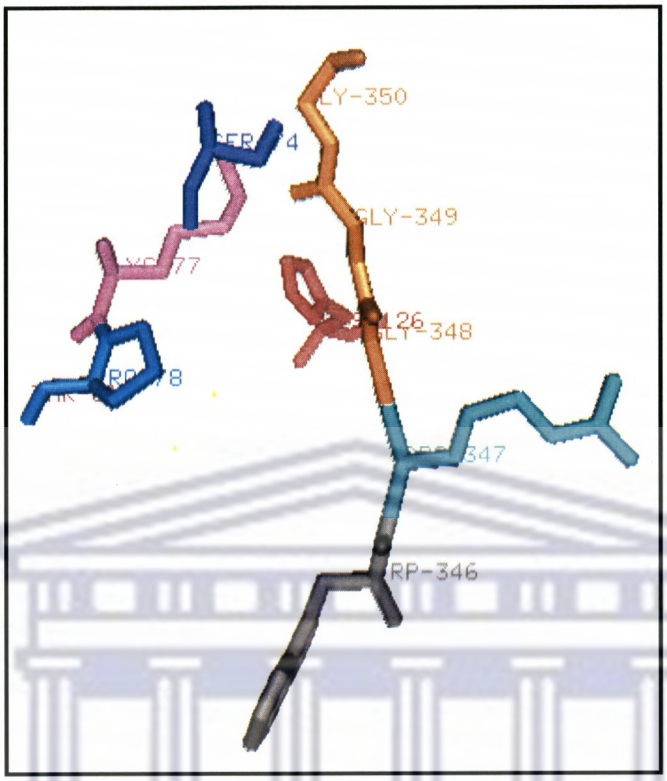


Figure 6.11: Residues predicted to form the active site of EstBL

6.2.5 Disulphide bridges-EstBL

The primary structure of EstBL revealed the presence of two cysteine residues (C283-C384) which were postulated to form disulphide bond. However, the EstBL model revealed that C283 and 284 were arranged in helix α5 and α7 respectively, and the distance predicted by the model (4Å) was not appropriate for these residues to form a disulphide bond (<3 Å). These observations were in agreement with the biochemical data (Chapter 5, Section 5.7.5), which showed no effect upon incubation of EstBL with the disulphide bond reducing agent.

6.4 Discussion

A 3D structural model of *B. multivorans* lipase (LipA) was consistent with the α/β hydrolase fold, which is found in two lipases from *Burkholderia* species with known 3D structures. Classical features for lipases, which include a catalytic triad, an oxyanion hole, a calcium binding pocket, a single disulphide bridge, and the lid domain, were predicted from the LipA 3D model structure, consistent with observations from solved structures (Noble *et al.*, 1993; Lang *et al.*, 1996; Kim *et al.*, 1997; Nardini *et al.*, 2000). These predictions are expected to be highly accurate since there is a large volume of high quality structural information available for these regions, based on the template and other Group I proteobacterial lipases with known 3D structures. This modelling experiment allows the identification of the residues that are structurally and functionally important. For instance, based on the model predictions, it is concluded that mutation of the putative catalytic triad residues (Ser87, His285, and Asp264) could results in loss of lipase activity. Furthermore, mutation of cysteine residues (C190-C270) could destabilize the structure, by deleting of critical disulphide bridge, as previously shown (Noble *et al.*, 1993).

The overall topology arrangement of the EstBL structural model was in good agreement with that of the template structure suggesting good accuracy of the model. The EstBL model revealed the α/β hydrolase fold, consistent with template structure (Wagner *et al.*, 2002). Before generating the EstBL model it was not clear which of the residues forms part of the EstBL active site. The EstBL 3D model provides a 3D framework for the rational design of site directed mutagenesis to identify residues that are crucial in

catalysis. Residues which are presumed to form part of active site in EstBL were identified in Section 6.3.4 and will be targeted for future site directed mutagenic studies. The catalytic role of serine within the SxxK motif in family VIII esterases has been demonstrated by site directed mutagenesis (Sakai *et al.*, 1999; Petersen *et al.*, 2001), however, the role of other residues involved in catalysis is still unclear. Although the structural data from EstB suggested the steric factors to account for the enzymes's selectivity for ester hydrolysis versus β-lactam cleavage (Wagner *et al.*, 2002), the detailed catalytic mechanisms of Family VIII esterase is still not known. More structural data is needed to elucidate the catalytic mechanism of these family VIII esterases.

CHAPTER SEVEN

General Discussion

Esterases and lipases are diverse groups of enzymes with high potential for industrial applications. These groups of enzymes are widely distributed in all domains of life. In view of the biocatalytic applications it is of importance to understand specificities and selectivities displayed by individual esterases and lipases. A vision for this project was to generate a set of esterases and lipases with defined catalytic features. This would form the basis for future development of biocatalytic processes for modification of ferulic acid derivatives.

In order to develop a biocatalytic process, a suitable catalyst must be identified (Ogawa and Shimizu, 1999; Burton *et al.*, 2002). The work reported in chapter three involves the isolation, screening and identification of novel esterase and lipase producing isolates. Bacterial isolates capable of growing on ethyl ferulate as a sole carbon source were screened from silage samples that were collected from a maize processing farm in Stellenbosch (South Africa). One strain, designed UWC10, was shown to produce high levels of esterolytic and lipolytic activities. Based on the phylogenetic analysis of the 16S rRNA gene, *rec*A gene RFLP pattern and species-specific primers, the UWC10 strain was identified as *B. multivorans* UWC10. Biotransformation studies with growing and resting cells also demonstrated that *B. multivorans* possessed ferulic acid decarboxylase activity.

Due to their high adaptation capacity and their catabolic potential to metabolize an enormous range of natural organic compounds, *Burkholderia* isolates play a crucial role in degradation processes of organic matter in nature (Larsen *et al.*, 1993). To the author's knowledge, this is the first report demonstrating the ability of *Burkholderia* isolates to grow on ethyl ferulate as a sole carbon source. Growth of *B. multivorans* in the presence of ethyl ferulate resulted in the production of both the ferulic and vinyl guaiacol in the media. The mechanisms of FA bioconversion are well document and include (i) non-oxidative decarboxylation, (ii) side-chain reduction, and (iii) coenzyme-A-independent deacetylation and coenzyme-A-dependent deacetylation (Priefert *et al.*, 2000). Based on the mechanisms of ferulic acid metabolism by *B. cepacia* (Andreoni *et al.*, 1984), *Ps. fluorescens* (Andreoni *et al.*, 1995) and *B. coagulans* (Karmakar *et al.*, 2000), a mechanistic model of ethyl ferulate hydrolysis by *B. multivorans* UWC10 was postulated. It is proposed that *B. multivorans* UWC10 hydrolyses the ester bond between the ethyl and ferulate groups, releasing ferulic acid. The released ferulic acid is decarboxylated to vinyl guaiacol by endogenous decarboxylase activity, as previously demonstrated (Andreoni *et al.*, 1984; Andreoni *et al.*, 1995; Karmakar *et al.*, 2000).

Chapter four describes the construction and screening of a *B. multiviorans* gene library for esterolytic and lipolytic activities. Screening of gene libraries offers several advantages over traditional wild type screening programmes. A major advantage of this molecular approach is the simultaneous access to the diversity of esterases/ lipases encoded in the genome of a target organism, as expression is carried out independent of

the natural regulatory system (Schlacher *et al.*, 1998). A further advantage is the relative simple separation of a single enzyme from the background of other esterase/lipase activities expressed in the particular donor organism, allowing rapid characterization of enzyme properties (Gledhill and Kell, 1998). Under such a strategy, target genes are immediately available, providing rapid access to enzyme improvement by protein engineering strategies such as directed evolution (Farinas *et al.*, 2001).

Screening of the *B. multivorans* gene library in *E. coli* for esterolytic and lipolytic activities led to the isolation of three recombinant clones, TEND5, pHOLA6 and pRASH14. Sequencing of the pRASH14 DNA insert led to identification of a lipase operon comprising of LipA gene located upstream of a LipB chaperone. The gene structure and organisation was consistent with the other lipase operons from subfamily I.2 (Arpigny and Jaeger, 1999). Lipase structural features included an N-terminal signal peptide, a catalytic triad (Ser, His, Asp), a single disulphide bond, a calcium binding site, oxyanion hole and a lid domain, all of which are characteristic of subfamilies I.1 and I.2 lipases (Jorgensen *et al.*, 1991; Frenken *et al.*, 1992; Kim *et al.*, 1996; Arpigny and Jaeger, 1999; Quyen *et al.*, 2004).

Analysis of pTEND5 insert DNA led to the identification of the gene encoding an esterase with homology to lactone-hydrolyzing esterases. The primary structure of EstEFH5 included a putative membrane anchor domain, a classical GxSxG signature motif, and putative oxyanion region feature which are also shared by lipases and other esterases (Arpigny and Jaeger, 1999; Khalameyzer *et al.*, 1999; Kim *et al.*, 2004b) This

class of esterases have been reported to be involved in the hydrolysis of lactones formed by Baeyer-Villiger monooxygenases (Onakunle *et al.*, 1997; Khalameyzer *et al.*, 1999). Interestingly, the truncated monoxygenase gene was found located upstream the EstEFH5 encoding gene, suggesting that a physiological role of EstEFH5 might involve hydrolysis of lactones in order to derive metabolic energy from these compounds. Functional classification of EstEFH5 based on gene oncology (GO) annotation further showed that this esterase belonged to the class of aromatic compound hydrolyzing enzymes, consistent with the activity of the pTEND5 enzyme against ethyl ferulate.

E. coli transformant pHOLA6 was also shown to encode esterolytic activity. Nucleotide sequencing revealed that a putative esterase gene (*estBL*) was located upstream of a permease gene. The primary structure showed two motifs (SxxK and GxSxG), either of which could potentially harbour a catalytic serine. Homology searches revealed that EstBL shared high sequence homology with family VIII esterases, peptidases and β-lactamases, all of which exploit serine residues within the SxxK motif as a nucleophile (Swaren *et al.*, 1995; Banerjee *et al.*, 1998; Sakai *et al.*, 1999; Petersen *et al.*, 2001). Based on the lipolytic enzyme classification scheme (Arpigny and Jaeger, 1999), EstBL belongs to family VIII esterases consisting of the poorly studied family of esterases, with only one three dimensional structure elucidated to date (Wagner *et al.*, 2002). Due to incorrect-annotation by genome sequencing programs and the high homology of these esterases with the N-terminus of β-lactamases, the CD-search protein domain annotations classifies most of the Family VIII esterases as members of class C β-lactamases (Marchler-Bauer and Bryant, 2004). It is generally agreed that the serine residue within

the SxxK motif is a catalytic serine in family VIII esterases, as shown by site directed mutagenesis studies (Sakai *et al.*, 1999; Petersen *et al.*, 2001). However, the other residues forming part of the hydrogen bonding network and the mechanism of reaction are still matters of speculation.

To the author's knowledge, this the first report on cloning and sequencing of esterase and lipase encoding genes from *B. multivorans* strains. From the cloning and sequencing data, *B. multivorans* UWC10 was shown to encode at least three functional lipolytic enzymes. The number of functional lipolytic genes in this organism may potentially be even greater, since narrow screening criteria were adopted during library screening procedures. This view is also supported by whole genome sequencing approaches of organisms from other *Burkholderia* including *B. pseudomallei* (Holden *et al.*, 2004) and *B. mallei* (Nierman *et al.* 2004), which have revealed a number of putative *lipolytic* genes. For example, in *B. pseudomallei* at least 10 putative lipase/esterase genes have been annotated (Holden *et al.*, 2004), while in *B. mallei*, more than 15 putative lipase/esterase genes have been identified. In *B. gladioli*, three esterase genes have been cloned and functionally expressed in *E. coli* (Schlacher *et al.*, 1998; Reiter *et al.*, 2000; Petersen *et al.*, 2001). The presence of multiple esterolytic/lipolytic encoding genes in a single genome is not surprising, given the versatility of *Burkholderia* isolates in the degradation of complex herbicides and pesticides (Yabuuchi *et al.*, 1992) and the roles of some lipolytic enzymes as virulence factors (Jaeger *et al.*, 1992).

The work presented in chapter five provided the first reports on functional expression of recombinant *B. multivorans* UWC10 esterolytic genes in *E. coli*. At the beginning of this study no evidence of the esterolytic or lipolytic in *B. multivorans* strain was available at either the molecular and biochemical level. Chapter five describes the development of expression vector systems that could allow functional expression of the cloned esterase and lipase genes in *E. coli*. A successful *E. coli* based expression system was developed which enabled the production of large quantities of biologically active recombinant EstBL. The recombinant EstBL was purified to homogeneity by a combination of chromatographic methods. A migration pattern on native PAGE suggested that EstBL was a monomer. The esterolytic activity of EstBL was also confirmed by zymogram assay.

Basic biochemical properties of the recombinant EstBL were also determined. Despite the presence of the β-lactamase (SxxK) motif within the EstBL primary structure, EstBL did not show any detectable activity on β-lactam substrates. Furthermore, substrate specificity against ρ-nitrophenyl- and β-naphthyl-esters revealed that EstBL showed a preference for shorter acyl chain length esters (ρ-NP-C2-C3). EstBL was inhibited by 1 mM PMSF, suggesting the involvement of a serine residue in the catalytic activity. Inhibition by metal ions Cu^{2+} and Ag^{2+} was consistent with this observation.

An *E. coli* based expression system was also developed for EstEFH5. Despite the high expression levels of the EstEFH5 enzyme, much of the protein was insoluble, possibly due to the highly hydrophobic nature of this protein. Attempts to optimize the expression

conditions, including changing the inducer concentration and induction temperature, were unsuccessful. EstEFH5 was recovered from the insoluble cell fraction by denaturation in 8M urea followed by *in vitro* refolding and purification using a His select ^{TM} Nickel column. The partially purified EsEFH5 was shown to be every unstable. Analysis of substrate specificity showed that EstEFH5 was a typical carboxylesterase, showing a strong preference for ρ-NP-C2 and ρ-NP-C3.

Lipase expression constructs were designed. The functional expression of *Burkholderia* lipase in *E. coli* typical requires co-expression with its downstream chaperone (Jorgensen *et al.*, 1991; Quyen *et al.*, 1999). However, time constraints precluded more detailed lipase expression studies. Rather than preventing off-pathway reactions such as aggregation, experimental evidence suggests that the downstream chaperone (LipB) provide steric information for a newly synthesised lipase to fold correctly (El Khattabi *et al.*, 2000). It postulated (Shibata *et al.*, 1998; Rosenau *et al.*, 2004), that during the secretion process, LipB is anchored to the inner membrane by the N-terminal hydrophobic segment with the large C-terminal domain exposed to the periplasm, thereby assisting the newly synthesised LipA cognate to fold into an active, protease-resistant conformation (Jaeger *et al.*, 1999).

Sequence data was used in the construction of 3D structural models. These structural models were considered to be of importance in identifying residues that could be targeted for future site directed mutagenesis studies. Sequence threading approaches (Jones *et al.*, 1999; Shi *et al.*, 2001), led to the identification of LipA and EstBL homologues sharing

78% and 52% sequence identity respectively, well above the "twilight zone" (Domingues *et al.*, 2001), where structural predictions are more reliable. Both LipA and EstBL 3D models structures revealed that the two enzymes adopt an α/β hydrolase fold, consistent with the template structures (Lang *et al.*, 1996; Wagner *et al.*, 2001).

The EstBL model structure suggested that Ser149, sited within the GxSxG motif is unlikely to be a catalytic serine, based on its topological arrangement. Serine (Ser149) was located on a long loop connecting two α-helices. This is inconsistent with previous studies, where the catalytic serine is typically located on the short loop connecting an α-helix and β strand which comprise a "nucleophilic elbow" (Kim *et al.*, 1997a; Bourne *et al.*, 2000; Bornscheuer *et al.*, 2002; Blair *et al.*, 2004). Another inconsistency was observed in the topological location of surrounding glycine residues (Gly147 and Gly151) on the same loop as the Ser149. For steric reasons surrounding glycine residues (corresponding to Gly149 and Gly150 in other lipase and esterase structures) are located on the respective α-helix and β strand, connected by a loop where the catalytic serine is normally located (Lang *et al.*, 1998; Jaeger *et al.*, 1999). Furthermore, the EstBL model also suggested that, Ser149 was surface exposed, as apposed to the observations from 3D structures were the catalytic serine had been consistently shown to be "buried" within the protein structure (Lang *et al.*, 1996; Kim *et al.*, 1997a; Dodson and Wlodawer, 1998; Wei *et al.*, 1999).

The EstBL model structure predicted that Ser74 could be a putative catalytic serine. The prediction was based on the fact that majority of residues proposed to form the active site in the template structure (Wagner *et al.*, 2002), were also observed (Chapter Six; Section 6.3.4). Topological locations of Ser74 and Lys77 were consistent with the observations from the template structure (Wagner *et al.*, 2002). The role of Lys77 has been proposed to be analogous to that of His in the catalytic triad (Dodson and Wlodawer, 1998), based on the observations from β-lactamase structures (Swaren *et al.*, 1995; Banerjee *et al.*, 1998). Although, Lys76 was close to the catalytic serine in the EstB template structure (Wagner *et al.*, 2002), its putative catalytic role has yet to be demonstrated.

Future work

Site directed mutagenesis and crystallization studies

A number of key questions on the catalytic function of EstBL remain unanswered. Future work should include the identification of the serine residue that acts as a nucleophile. A rational approach involving the replacement of potentially functionally residues by non-disruptive substitution could be adopted. The following mutations: S74A, S149A, and K77A would be appropriate targets based on the data obtained from homology modelling studies. Alanine would be an appropriate as substitute residue based on the following reasons: It lacks a bulky side chain thereby not imposing any steric and electrostatic effects, and it does not disrupt main chain conformation as occurs with glycine and praline (Ling and Robinson, 1997).

Mutagenic oligonucleotides would be designed with unique restriction site for screening purposes. Complementary mutagenic oligonucleotides would be used to direct replication of the EstBL encoding gene. These oligonucleotides could direct the synthesis of both strands of the plasmid DNA template in a linear (non-PCR based approach) and generate the proposed mutations: S74A, S149A, and K77A. The mutants would be screened based on the engineered restriction site(s) and the functional importance of specific amino acid residues would be assessed by comparing kinetic constants of the wild type protein with that of mutant proteins.

Only one crystal structure from family VIII esterases has been solved to date (Wagner *et al.*, 2002). Although the data from this study provide evidence of steric factors which may account for the enzymes's selectivity, the detailed esterolytic mechanism is still unknown. Crystallization trials are already underway, with an aim of obtaining more structural data in order to understand the substrate specificity and catalytic mechanism.

Expression optimization studies

Although, EstEFH5 was successfully refolded from insoluble protein, the enzyme was not purified to homogeneity and could not be characterized in detail. This was mainly due to the fact that most of the expression studies attempted to failed produce soluble biologically active enzyme. Since EstEFH5 showed activity against the model substrate (ethyl ferulate), part of the future work should involve development of an *E. coli* based expression system that would enable over-expression of this enzyme in a soluble biologically active state. The approach would involve *de novo* synthesis of a ± 200 pb

region to reduce the high GC content at the 5' and of the gene, which encodes the putative N-terminal membrane domain. A set of overlapping oligonucleotide pairs would be designed to modify the GC rich condons of EstEFH5 according to the codon usage of *E. coli* (because most of the esterase and lipase genes from *Bacillus* species have been efficiently expressed in *E. coli*). This strategy has been previously successfully attempted in optimizing the expression of esterase (EstC) encoding gene from *B. gladioli* (Reiter *et al.*, 2000) and lipase encoding gene from *B. cepacia* (Quyen *et al.*, 1999).