Functional characterisation of a novel ferulic acid esterase from Malawian hot spring metagenome

by

Rhulani Ngobeni

Submitted in fulfilment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology at the University of the Western Cape

UNIVERSITY *of the*
WESTERN CAPE

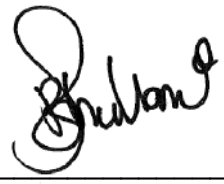
Supervisor: Assoc. Prof. I.M. Tuffin

Co-Supervisors: Dr. S. Easton, Dr. R. Bauer and Prof. D.A. Cowan

NOVEMBER 2011

Declaration of Originality

I, the undersigned, hereby declare that “Functional characterisation of a novel ferulic acid esterase from Malawian hot spring metagenome” is my own original work that has not been previously in its entirety or in part submitted for any degree or examination at any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.



_____

Rhulani Ngobeni

23 February 2011

Date

Abstract

There has been a decline in the global fossil fuel reserves, due to an increasing demand for petroleum. Biofuels can be used as an alternative source of energy whereby biomass is converted to liquid fuels such as bioethanol. There is considerable interest in lipolytic enzymes because of their broad substrate range and for this purpose these enzymes have potential in a variety of biotechnological application. Lipolytic enzymes include esterases (E.C 3.1.1.1) and lipases (E.C 3.1.1.3). Esterases preferentially hydrolyse short chain ($<C_{10}$) ester-containing molecules that are partly soluble in water, while lipases hydrolyse a broad range of substrates preferably water-insoluble fatty acyl molecules ($>C_{10}$). The aim of this study was to express, purify and characterise the lipolytic enzyme present on a fosmid, Try 11, previously isolated from a metagenomic library of a Malawian hot spring, which conferred activity on tributyrin and ethyl ferulate. Bioinformatic analysis of the fosmid insert sequence predicted an open reading frame consisting of 951 bp, designated RHgene34, encoding a 317 amino acid protein with 41 % similarity to the α/β hydrolase fold-3 domain protein of *Burkholderia* sp. The RHgene34 protein contains conserved motifs of esterases/lipases, such as HGGG (residues 95-98), GxSxG (residues 167 - 171) and the putative catalytic triad composed of Ser157, Asp255 and His285. The gene was cloned and expressed in pET21a(+), and transformed into *Escherichia coli* Rosetta. *p*-Nitrophenol (*p*-Np) fatty acyl esters of different carbon chain lengths were used for kinetic characterisation of RHgene34. Kinetic analysis revealed that RHgene34 had a broad range activity on the *p*-Np esters, from C_2 - C_{14}. RHgene34 operates optimally at 45 ºC, pH 9.0 and has a half-life of 30 mins at 45 ºC. This study demonstrates that functional screening combined with the sequence analysis is a useful approach for isolating novel enzymes from a metagenome.

Keywords*: Biofuels; Lipolytic enzyme; Esterase; Lipase; Esters*

This thesis is dedicated to Teboho for the unprecedented support and eternal love.

“Wisdom is not a product of schooling but of the lifelong attempt to acquire it.”

- Albert Einstein

Acknowledgements

Working in the exciting and quickly changing field of biofuels has been a thrill. I wish to express my grateful thanks to my principal supervisor ASSOC. PROF. MARLA TUFFIN for giving me a chance to work with her, for your trust in me and also I would like to thank her supervision, encouragement, criticism and patience to guide me during my thesis studies as well as the funding. PROF. DONALD COWAN, for affording me the opportunity to attain my MSc. To DR. ROLENE BAUER, thank you for your valuable comments and help during my thesis studies. Also I would like to thank DR. SAMANTHA EASTON, for being such an efficient supervisor: your hectic deadlines and quick reviews of the thesis chapters, thank you for your valuable guidance, support and enthusiasm. In addition, would like to thank DR. INONGE MULAKO for her supervision during my first year of masters and candid conversations.

The open and friendly atmosphere at the IMBM Institute has made for an excellent time. I would like to thank all colleagues and fellow students for two very good years. In particular my good friends TIMNA JANUARY, MUNAKA MATSHAYA and FREEDOM TSHABUSE; thank you for all the help, discussions and fun – sharing space with you has been great. I also wish to thank LONNIE VAN ZYL, WILLIAM MAVENGERE and DOMINIQUE ANDERSON for their invaluable help, I am gravely indebted; lab technicians NONHLE GEZA and ORATILWE TSHUKUDU for helping me out with reagents and chemicals. DR. HEIDE GOODMAN, for her words of encouragement and her influence in Residence Admin, UWC finance and HR.

I would like to thank my best friend, TEBOHO MOTSWARE, for his love, support, prayers and listening when I needed to unload even when he did not understand the terminology. Thank you for believing in my capabilities, pushing me to be a better version of myself and never losing hope in me. To MY MOTHER for her endless support and encouragement without whom I would not have made it this far: you're the best. Finally, my profound gratitude goes to the almighty GOD who gave me strength, courage and His immeasurable blessings and grace.

Preface

This thesis is presented as a compilation of six chapters. The first five chapters comprise the thesis followed by a reference chapter.

CHAPTER 1	**Literature Review**
	Ferulic Acid Esterase for Bioethanol Production
CHAPTER 2	**Materials and Methods**
CHAPTER 3	**Screening and Sequence Analysis**
	Results and Discussion
CHAPTER 4	**Functional Characterisation**
	Results and Discussion
CHAPTER 5	**General Discussion and Conclusion**
CHAPTER 6	**References**

Contents

Declaration II

Abstract III

Dedication IV

Acknowledgements V

Preface VI

List of Figures XI

List of Tables XV

Abbreviations XVI

CHAPTER 1: LITERATURE REVIEW 1

1.1: Energy Crisis and Biofuels 1

1.2: Bioethanol 1

1.2.1: Lignocellulosic Biomass and Main Components 4

1.2.2: Esters and Ester Linkages 7

1.3: Bioethanol Processes 7

1.3.1: Pretreatment 7

1.3.2: Enzymatic Hydrolysis 9

1.3.2.1: Cellulases 9

1.3.2.2: Hemicellulases 9

1.3.2.3: Ligninases 10

1.3.3: Production of Bioethanol through Microbial Fermentation 11

1.4: The role of Thermo-stable Enzymes in Fermentation 11

1.5: Ferulic Acid Esterase (FAE) 12

1.6: α/β Hydrolase Fold Family Protein 13

1.7: Mode of Action of α/β Hydrolase Fold 14

1.8: Classification of FAE 15

1.9: FAE's from Microbial Sources 17

1.10: Industrial Application of Ferulic Acid Esterase 18

Project Aims and Objective 20

CHAPTER 2: MATERIALS AND METHODS 21

2.1: Chemicals and Enzymes 21

2.2: Media 21

2.3: Bacterial Strains, Plasmids and Growth Conditions 22

2.4: DNA Purification 23

2.5: Transposon Mutagenesis 24

2.6: Bioinformatics and Analysis 24

2.7: Polymerase Chain Reaction (PCR) of RHgene34 25

2.8: Cloning of RHgene34 26

2.9: Preparation of competent *E. coli* cells 27

2.9.1: Preparation of Electrocompetant Cells 27

2.9.2: Preparation of Competent *E. coli* Cells by $CaCl_2$ Treatment 27

2.10: Transformation of *E. coli* cells 28

2.10.1: Electroporation 28

2.10.2: Chemical Transformation 28

2.11: Preparation of Cell-Free Extract for RHgene34 Purification 28

2.11.1: Enzymatic Lysis 29

2.11.2: Mechanical Disruption (Sonication) 29

2.12: Protein Expression and Purification 29

2.13: SDS-PAGE Analysis 30

2.14: Enzymes Assays 31

2.14.1: Substrate Specificity 32

2.14.2: Determining the Effect of pH on Enzyme Activity 32

2.14.3: Determining the Effect of Temperature on Enzyme Activity 32

2.14.4: Determining Catalytic Efficiency 33

2.15: Fast Performance Liquid Chromatography (FPLC) Determination of Quaternary Structure 33

2.16: High Performance Liquid Chromatography (HPLC) Analysis 33

CHAPTER 3: SCREENING AND SEQUENCE ANALYSIS 35

3.1: Metagenomic Fosmid Library Verification 35

3.2: End-Sequencing 38

3.3: Transposon Mutagenesis 40

3.4: Sequence Analysis 40

3.5: Phylogenetic Analysis 52

3.6: Homology Modelling 54

CHAPTER 4: ENZYME CHARACTERISATION **63**

4.1: Cloning of the Lipolytic Gene RHgene34 63

4.2: Expression of RHgene34 67

4.3: Enzymatic Characterisation of the RHgene34 Gene Product 69

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSION **78**

CHAPTER 6: REFERENCES **85**

List of Figures

CHAPTER 1

Figure 1.1: Different sources of biomass for bioethanol production. First generation bioethanol production uses conventional feedstocks such as maize, wheat, sweet sorghum and sugar cane while second generation bioethanol uses lignocellulosic feedstocks such as wood chips, fast growing grass like switch grass, forest/agricultural waste and straw (*Adapted from* European Renewable Energy Council, 2007).

Figure 1.2: Lignocellulose model showing cellulose, lignin and hemicellulose molecular structures (Holtzapple, 2003).

Figure 1.3: Pretreatment renders lignocellulosic biomass easier to hydrolyse using enzymes (*Adapted from* Wyman and Yang, 2009).

Figure 1.4: The synergistic action of a variety of hemicellulases responsible for complete degradation of xylan (Shallom and Shoham, 2003).

Figure 1.5: Schematic illustration of the α/β hydrolase fold. α-Helices are shown as red cylinders and β-sheets as blue arrows. Orange circles indicate the topological position of active site residues (nucleophile after β5, Asp/Glu after β7 and His in the loop between β8 and αF) (Bornscheuer, 2002).

Figure 1.6: Mechanism of ester bond hydrolysis by esterase enzymes (Jaeger *et al*., 1994).

Figure 1.7: Neighbour-joining phylogenetic tree of selected lipolytic sequences, discovered by functional screening of metagenomic libraries based on conserved sequence motifs of bacterial lipolytic enzymes (Tuffin *et al*., 2009).

CHAPTER 3

Figure 3.1: Screening a metagenomic library for tributyrin degrading enzymes (A). The two clones identified were Try 11 and Try 12, shown here depicting zones of clearing, after 2 days of 37 ºC incubation (B).

Figure 3.2: Try 11 clone exhibiting ferulic acid esterase activity on ethyl ferulate agar after 2 days incubation at 37 ºC.

Figure 3.3: Restriction enzyme digestion profiles of clones Try 11 and Try 12, using *Eco*RI and *Hind*III. Lanes 2 and 3: Try 11 and Try 12, respectively. The 8.2 kb fosmid vector

backbone is indicated by a red arrow. Lambda DNA digested with the *Pst*I restriction enzyme was used as a molecular weight marker (Lane 1).

Figure 3.4: Arrangement of open reading frames identified in insert of Try 11. Arrows indicate the location and orientation of predicted open reading frames. Details of putative protein function and GenBank accession numbers are given in Table 4.2.

Figure 3.5: The nucleotide sequence of the RHgene34 lipolytic gene from the metagenomic library. The predicted amino acid sequence of RHgene34 is given above the nucleotide sequence in the standard one-letter code. The putative promoter regions (-35 and -10 regions) and the ribosomal bind site (RBS) is highlighted in boldface and underlined. RHgene34 is 951 nucleotides in length with an ATG start codon. The asterisk denotes the stop codon (TAG) and the protein encoded a polypeptide with a molecular mass of 34 kDa.

Figure 3.6: Multiple sequence alignment of the RHgene34 protein sequence with close relatives indentified from UNIPROT BLAST. Boxes indicate sequence similarity with a threshold value of 90 %. The putative catalytic triad residues composed of Ser157 (S), Asp255 (D) and His285 (H) are indicated with shaded triangles. The family IV most conserved motif, HGGG, is indicated by a shaded circle. The GxSxG lipolytic motif (PROSITE Accession No. PS00120) is indicated with a shaded star. Accession numbers denote the following; D2UAR1: Putative esterase/lipase/thioesterase family protein [*Xanthomonas albilineans*], B5WN60: α/β Hydrolase fold-3 [*Burkholderia* sp. H160], F0BFQ2: Esterase/lipase [*Xanthomonas vesicatoria*], B8Y564: Lipolytic enzyme [Uncultured bacterium], E6VMR1: Putative lipase/esterase [*Rhodopseudomonas palustris* DX-1]. Percent identity of RHgene34 to the protein sequences is indicated in the Table. The sequences were aligned using ClustalW (Thompson *et al*., 1994).

Figure 3.7: Prediction of N-terminal signal peptide cleavage site in polypeptide RHgene34.

Figure 3.8: Phylogenetic analysis of RHgene34 and 31 selected lipolytic enzymes, representing 9 different families. Lipolytic families were defined by Lee *et al*., (2006) and Arpigny and Jaeger (1999). The phylogenetic tree was constructed using CLC Genomics Workbench 3.0 software with the neighbour-joining method. The numbers at nodes indicate the bootstrap percentage of 1000 resamples. RHgene34, denoted by shaded circle, is shown to belong to family IV lipolytic enzymes.

Figure 3.9: Secondary structure for the amino acid sequence obtained for RHgene34. Figure generated by PSIPRED (McGuffin *et al*., 2000; Jones, 1999).

Figure 3.10: Secondary structure topology for the amino acid sequence obtained for RHgene34 using PDBsum (Morris *et al*., 1992). The elements of the secondary structure shown correspond to the canonical α/β hydrolase fold.

Figure 3.11: Homology model of the (A) RHgene34 protein built by the Swiss Model server using PDB structures 3ainD; residues 27 to 315 (B) and 2wirA; residues 3 to 313 (C) used as templates. The polypeptide spans across the colour spectrum from blue (N-terminal) to red (C-terminal).

Figure 3.12: RHgene34 homology model built by the 3D JIGSAW model server. The polypeptide spans across the colour spectrum from blue (N-terminal) to red (C-terminal).

Figure 3.13: Ramachandran plot for model of RHgene34 built by the Swiss model server, showing number of residues in favoured, allowed and disallowed region.

Figure 4.1: Restriction enzyme digestion of clone RH-pET. Lane 1: DNA molecular marker, lambda-*Pst*I digested DNA. Lane 2: *Nde*I and *Hind*III digested recombinant RH-pET plasmid construct. Lane 3: Uncut recombinant RH-pET plasmid.

Figure 4.2: RH-pET clone harbouring RHgene34 gene with lipolytic activity (A) and control 7-6G clone with no lipolytic activity (B) on tributyrin agar indicator plates.

Figure 4.3: PCR amplification of RHgene34 using gene specific primers (Table 2.2) for confirmation of cloning into the pET vectors. Lane 1: DNA molecular marker Lambda *Pst*I digested DNA; Lane 2: PCR amplified gene of RHgene34; Lane 3 and 4: negative control.

Figure 4.4: RH-pET clone displaying FA activity on ethyl ferulate plates (B). The control (*E. coli* Rosetta pET21a) with no lipolytic activity (A).

Figure 4.5: RH-pET clone displaying lipase activity (A) and control 7-6G clone (B) on olive oil-Rhodamine B lipase agar plates viewed under UV light.

Figure 4.6: SDS-PAGE analysis of cell extracts of RH-pET in *E. coli* Rosetta(DE3)pLysS. The protein band corresponding to a size of 34 kDa is indicated. Lane 1: protein molecular weight marker (#SM0671 Fermentas); Lane 2 and 6: IPTG induced total protein extract; Lanes 3 and 7: soluble fraction after overnight induction at 25 ºC and 30 ºC with IPTG, respectively; lane 4 and 8: uninduced total protein extract soluble; lane 5 and 9: soluble fraction of RHgene34 after overnight induction at 25 ºC and 30 ºC with IPTG.

Figure 4.7: SDS-PAGE analysis of His-Tag purification of *E. coli* Rosetta(DE3)pLysS RH-pET. Lanes 1: protein molecular weight marker #S8445 (Sigma), lane 2: Total cell extract of RHgene34; Lane 3: Flow through elute; Lane 4: Elute from binding buffer; Lane 5: Elute from

washing buffer; Lane 6: Eluted RHgene34 protein (MW 34 kDa). Lane 7: Elute from strip buffer.

Figure 4.8: FPLC standard curve of the different retention times of albumin, cytochrome C and carbonic anhydrase. *E. coli* Rosetta(DE3)pLysS RH-pET was determined using the straight line equation determined by the line of best-fit.

Figure 4.9: Substrate specificity of the purified RHgene34 enzyme toward *p*-nitrophenyl esters of varying chain lengths. Relative activity was shown as the percentage of the activity towards *p*-Np decanoate (C_{10}).

Figure 4.10: Reverse phase chromatography of reactions performed with purified RHgene34 on methyl *p*-coumarate (A), methyl caffeate (B), methyl ferulate (C) and methyl sinapate (D).

Figure 4.11: Effect of pH on purified RHgene34 protein activity using *p*-NP esterified with fatty acid of ten carbon chain length as the substrate. The activity was determined different in various pH buffers at 25 °C. Maximum activity at pH 9.0 was taken as 100 %.

Figure 4.12: Effect of temperature on purified RHgene34 protein activity using *p*-NP esterified with fatty acid of ten carbon chain length as the substrate. The activity was determined at different temperatures at pH 7.5 in 100 mM sodium phosphate, 100 mM NaCl buffer. Maximum activity at 45 ºC was taken as 100 %.

Figure 4.13: The thermal inactivation profile of RHgene34 at 25 °C (♦), 35 °C (▲), 45 °C (■) and 55 ºC (▼).

Figure 4.14: Kinetic parameters of purified RHgene34 gene product towards three *p*-Np ester substrates. **A**: Michaelis-Menten direct linear plot of RHgene34 towards C_2. **B**: Michaelis-Menten direct linear plot of RHgene34 towards C_3. **C**: Michaelis-Menten direct linear plot of RHgene34 towards C_8. Kinetic data were fitted using GraphPad Prism software (San Diego, USA).

Figure 4.15: Kinetic parameters of purified RHgene34 gene product towards three *p*-Np ester substrates. **A**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **B**: logarithmic Hill plot towards C_{10}. **C**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **D**: logarithmic Hill plot towards C_{12}. **E**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **F**: logarithmic Hill plot towards and C_{14}. Kinetic data were fitted using GraphPad Prism software (San Diego, USA).

List of Tables

CHAPTER 1

Table 1.1: Production of bioethanol on a global scale for 2007 (Sanchez and Cardona, 2008).

Table 1.2: Chemical composition of lignocellulosic biomass from different second generation sources (Sun and Cheng, 2002).

Table 1.3: Physiochemical characteristics of purified ferulic acid esterases isolated from different microbial and fungal sources.

CHAPTER 2

Table 2.1: Strains and plasmids used in this study.

Table 2.2: Primers used in this study.

CHAPTER 3

Table 3.1: Nucleotide end-sequences of fosmid clone Try 11. The nucleotide identity of the closest match is indicated.

Table 3.2: Predicted ORFs in fosmid Try 11.

Table 3.3: Rare codons and their frequency in the nucleotide sequence obtained for RHgene34 predicted by rare codon calculator for expression in *E. coli*.

CHAPTER 4

Table 4.1: Kinetic parameters of the RHgene34 enzyme using Michaelis-Menten *p*-Np esters.

Table 4.2: Kinetic parameters of the RHgene34 enzyme using non Michaelis-Menten *p*-Np esters.

Abbreviations

A	absorbance
ºC	Degrees Centigrade
APS	Ammonium persulphate
ATP	Adenosine triphosphate
Bp	Base pair
BSA	Bovine serum albumin
CAPS	3-Cyclohexylamino-1-propanesulphonic acid
CAM	Chloramphenicol
C-terminus	Carboxy-terminus
x *g*	Centrifugal force as a multiple of gravitational acceleration
Da	Dalton
dH_2O	Demineralised water
DNA	Deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EC	Enzyme Commission
et al	*et alia* (and others)
EtBr	Ethidium bromide
EDTA	Ethylene diamine tetra acetic acid
h	Hill coefficient
hr	hour
IPTG	Isopropyl-β-D-thiogalactopyranoside
$K_{0.5}$	Hill constant (apparent K_m)
K_{cat}	Catalytic turnover
kDa	kilo Dalton
K_m	Michaelis-Menten constant
l	litre

LB	Luria-Burtani
LBA	Luria-Burtani agar
µg	microgram
µl	microlitre
ml	millilitre
mg	miligram
mM	milimolar
min	minutes
ms^{-1}	meters per second
mol	molar
MW	Molecular Weight
MES	2-(*N*-morpholino)ethanesulphonic acid
ng	nanogram
nm	nanometer
MCS	Multiple cloning site
N-terminus	Amino-terminus
TEMED	N,N,N',N'-Tetramethylethylenediamine
ORF	Open reading frame
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
p-Np	*para*-nitrophenyl
rpm	revolutions per minute
S	substrate
SDS	Sodium dodecyl sulphate
SOB	Super optimal broth
SOC	Super optimal broth with catabolic repression
sec (s)	seconds

TAE	Tris acetic EDTA
TE	Tris EDTA
Tris	Tris-hydroxymethyl-aminomethane
Tris-HCl	Tris(hydroxymethyl)methylamine hydrochloride
U	units
UV	Ultraviolet
V_{max}	Maximum velocity

CHAPTER 1

Ferulic Acid Esterase for Bioethanol Production

1.1 Energy Crisis and Biofuels

Due to increasing populations and ever-expanding industrial economics, fossil fuel reserves are declining (Zecca and Chiari, 2010). Fossil fuels are non-renewable and may be responsible for changing the world's climate (McMillan, 1996). Fossil fuels, which include coal, crude oil and natural gas, currently provide 80 % of the world's energy, 58 % of which is used for transport, heating and other industrial processes (Lynd, 1996; McMillan, 1996). Fossil fuels are a non-renewable resource, which take millions of years to form, therefore the current rate of consumption is unsustainable (Zecca and Chiari, 2010; Demirbas, 2007).

Biofuels is the name given to any solid, liquid or gas fuel, which is derived from plant material, usually referred to as biomass (Demirbas, 2009). Biofuels can, in principle, be synthesised from any biological organic source; however, the most common sources are plants, such as plant oils, sugar beets, cereals, organic waste and the processing of biomass (Jegannathan *et al*., 2009). The fermentation of biological matter produces secondary energy carriers, such as sucrose and glucose, which can subsequently be used as fuel. Biofuels are a viable substitute for fossil fuels because they reduce greenhouse gas emissions and also contribute to sustainable development by minimising the global dependence on fossil fuels (Gary *et al*., 2006). There are different types of commercially available biofuels such as biomethanol, biobutanol, biodiesel and bioethanol (Balat *et al*., 2008).

1.2 Bioethanol

Bioethanol is separated according to the type of the source material used for its production (first generation and second generation) (Figure 1.1).

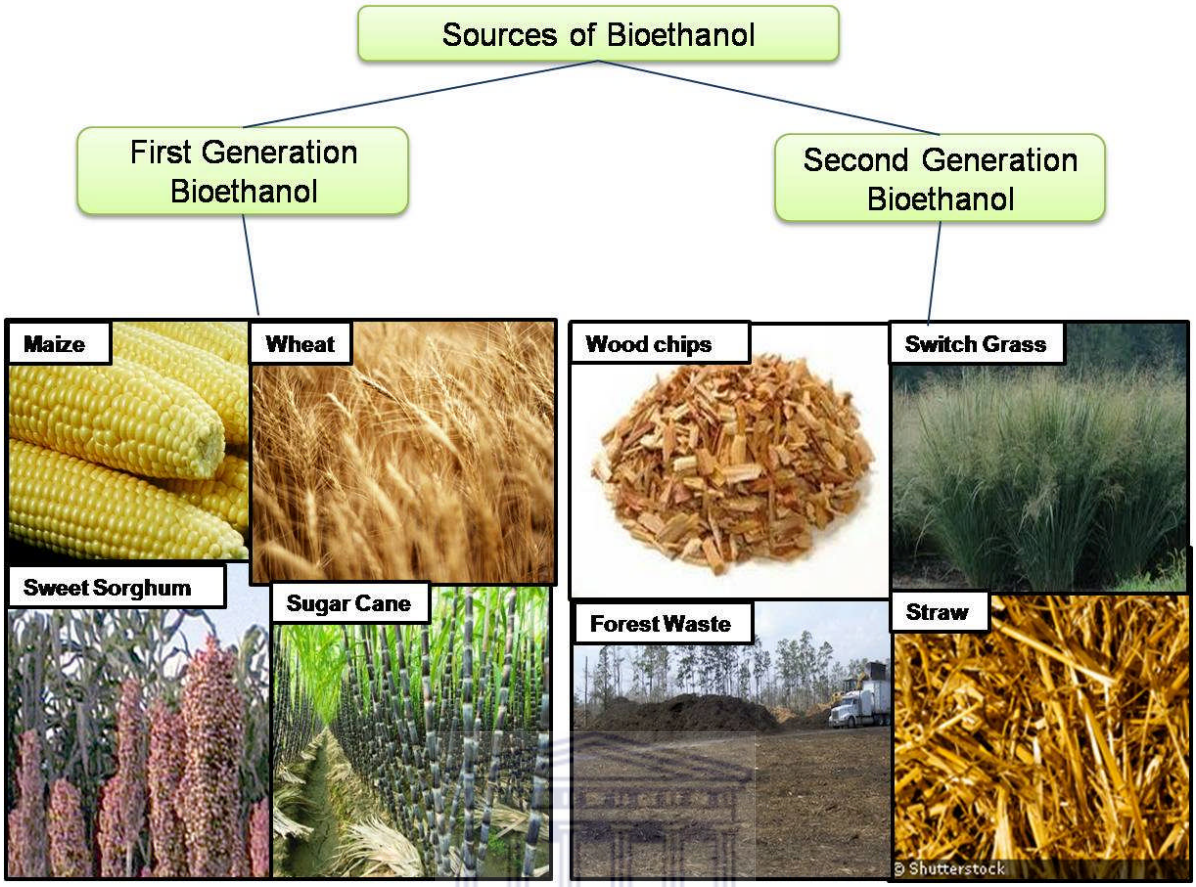


Figure 1.1: Different sources of biomass for bioethanol production. First generation bioethanol production uses conventional feedstocks such as maize, wheat, sweet sorghum and sugar cane while second generation bioethanol uses lignocellulosic feedstocks such as wood chips, fast growing grass like switch grass, forest/agricultural waste and straw (*Adapted from* European Renewable Energy Council, 2007).

Fuels that are produced from starchy (high in simple sugars) biomass sources like maize, sugarcane and wheat, produce first generation bioethanol (Figure 1.1). An advantage of first generation bioethanol is the high sugar which facilitates the fermentation process (Cherubini, 2010). The annual production of first generation biofuels is almost 50 billion litres; however the source material which drives such production is causing serious debate (Tan *et al*., 2010). The diversion of food crops, water and land resources for biofuel production, when food security is already a recognised problem (Srinivasan, 2009), has forced the development of second generation biofuel from agricultural waste, i.e. lignocellulosic biomass (Naik *et al*., 2010). Second generation biofuel utilises the residual non-food parts of current

crops, such as stems, leaves and husks, that are left behind once the food crop has been extracted, as well as other crops that are not used for food purposes such as switch grass and cereals which bear little grain (Lynd, 1996). Clearly, second generation biofuels are preferable to the first, however second generation bioethanol production requires a hydrolytic pre-treatment to break down the polymeric cellulose, hemicellulose and lignin biomass components into fermentable sugars.

The commercial production of bioethanol has been ongoing since the 1800's, albeit with fluctuations in popularity (Antoni, 2007). According to Galbe (2002) and Zaldivar (2001), the recent fuel crisis has attracted renewed interest in bioethanol production. In a developing country like South Africa, the production of bioethanol is expensive which is mainly due to investment in feedstocks and downstream processes (Balat *et al*., 2008). The difficulties are clear when we consider biofuel production in South Africa in comparison with other countries (Table 1.1) (Sanchez and Cardona, 2008).

For a developing country bioethanol offers other advantages such as economic growth and job creation (Mabee and Saddler, 2010). Biofuels are biodegradable and safer to handle than fossil fuels, making spills less dangerous and much easier to remediate compared to oil. During biofuel combustion, there is lower carbon output and fewer toxins are produced, making them a cleaner alternative, which preserves atmospheric quality and lowers air pollution (Naik *et al*., 2010). Fossil fuels are non-renewable and their continued use is unsustainable, whereas biofuels can be manufactured from virtually any plant material, making it an efficient recycling process (Kantarelis and Zabaniotou, 2009). Biofuels can be significantly less expensive than gasoline and other fossil fuels, particularly as the technology becomes more advanced. As worldwide demand for oil increases and oil resources deplete, oil and gasoline prices are expected to increase (González-Garcia *et al*., 2010).

Although there are clearly many advantages to the production and use of bioethanol, there are also some disadvantages to consider. Biofuels have a lower energy output than traditional fuels, therefore requiring higher consumption levels in order to produce the same amount of energy. In order to refine current biofuels to increase energy outputs, and to build the necessary manufacturing plants to increase biofuel production, a high initial investment is required (González-Garcia *et al*., 2010).

Table 1.1: Production of bioethanol on a global scale for 2007 (Sanchez and Cardona, 2008).

Ranking	Country	Ethanol (in million litres)
1	United States of America	18 376
2	Brazil	16 998
3	China	3 849
4	India	1 900
5	France	950
6	Germany	765
7	Russia	647
8	Canada	579
9	Spain	462
10	South Africa	386

Biofuels are not widely available for consumer purchase and most vehicles are not equipped to run on biofuel products. The current limited availability of biofuels reduces their feasibility as alternative energy sources. Several studies have been conducted to analyze the carbon footprint of biofuels (Naik, *et al*., 2010). While they may be cleaner to burn, there are strong indications that the process to produce the fuel, including the machinery necessary to transport the crops or plants from source to production plant, has high carbon emissions (Osmont *et al*., 2010; Vassilev *et al*., 2010).

However, there is a need to find alternative renewable energy sources. Biofuels are generally agreed to be a viable alternative for the 21st century.

1.2.1 Lignocellulosic Biomass and Main Components

Lignocellulosic biomass is the most abundant material in the world. Its sources range from trees to agricultural residues (Mabee and Saddler, 2010). Lignocellulose provides structure to plants and is found in roots, stalks and leaves. It is made up of three basic polymers: cellulose ($C_6H_{10}O_8$), hemicellulose ($C_5H_8O_4$) and lignin (Balat

et al., 2009**;** Cosgrove, 2001); and together they constitute the plant cell wall. The combination of hemicellulose and lignin provides a protective sheath impervious to microbial attack around the cellulose, which must be modified or removed before the hydrolysis of cellulose can occur (Humelinck *et al*., 2005). Cellulose microfibrils are linked together by hemicellulosic tethers to form the cellulose-hemicellulose network, which is embedded in the pectin matrix (Figure 1.2) and the composition differ between biomass sources (Table 1.2).

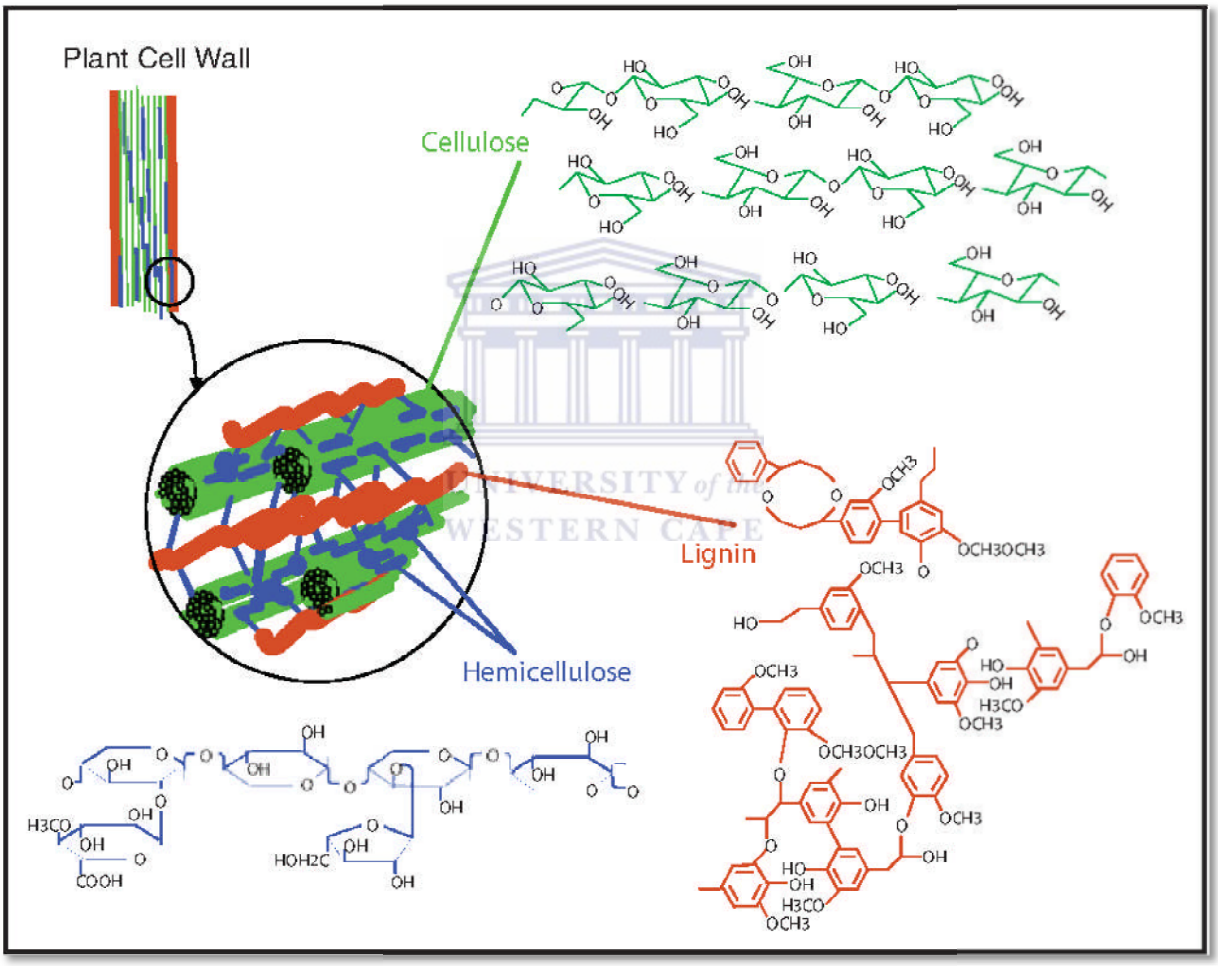


Figure 1.2: Lignocellulose model showing cellulose, lignin and hemicellulose molecular structures (Holtzapple, 2003).

Cellulose is a homogenous linear polymer of anhydroglucose linked by β-1,4 glycosidic bonds and is the most abundant polymer on earth (Zaldivar *et al*., 2001). The secondary and tertiary conformation of cellulose, together with hemicellulose,

lignin, protein and mineral elements, makes cellulose resistant to enzymatic digestion and gives plants structural strength (Zhu *et al*., 2008; Chang and Holtzapple, 2000; Fan *et al*., 1980).

Table 1.2: Chemical composition of lignocellulosic biomass from different second generation sources (Sun and Cheng, 2002).

Lignocellulosic material	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Hardwood stems	40-55	24-40	18-25
Softwood stems	45-50	25-35	25-35
Wheat straw	30	50	15
Leaves	15-20	80-85	0
Switch grass	45	31.4	12.0

Second to cellulose in abundance is hemicellulose, which is composed of highly branched and linear heterogeneous polysaccharides that are connected via hydrogen bonds to the cellulose microfibrils in the plant cell wall, cross-linking them into a strong network (Shallom and Shoham, 2003). Hemicellulose is composed of polysaccharides of pentoses (D-xylose and L-arabinose), hexoses (D-galactose, L-galactose, D-mannose, L-rhamnose, L-fucose) and a number of sugar acids (D-glucuronic acid). It has a backbone composed of 1,4-linked β-D-hexosyl residues branched at O-2 and/or O-3, with L-arabinofuranose, acetate or 4-O-methyl glucuronic acid moieties. Xylan is the most abundant and important hemicellulose in cell walls (Mosier, *et al*., 2005; Zaldivar *et al*., 2001; Fillingham *et al*., 1999).

Hemicellulose is chemically bound to lignin and serves as an interface between lignin and cellulose. The degradation of hemicellulose requires the synergistic action of many enzymes because of the diverse composition of the plant cell wall (Gírio *et al*., 2010; Mosier, *et al*., 2005), and because hemicellulose is randomly fortified with ester bonds (Shallom and Shoham, 2003).

Lignin is a complex aromatic, three-dimensional polymer composed of linked six-carbon phenolic rings with various carbon chains and other chemical functionalities

(Chang *et al*., 2001). Since it is not easy to dissolve lignin without destroying it, the precise chemical structure has not been established. Lignin is non-crystalline and generally contains three monomeric alcohols (*trans*-*p*-coniferyl alcohol, *trans*-*p*-sinapyl and *trans*-*p*-coumaryl) derived from *p*-cinnamic acid. Lignin is linked to both hemicellulose and cellulose, forming a physical seal around the components, which prevents entry of solutions and enzymes (Zaldivar *et al*., 2001). Plant cell walls also contain small amounts of structural gylcoproteins (hydroxyproline-rich extensions), phenolic esters (mainly ferulic acids and *p*-coumaric acids), and ionically and covalently bound minerals like calcium and boron.

1.2.2 Esters and Ester Linkages

Hydroxycinnamic acids are phenolic acids that are directly esterified to lignin surfaces (Jiyama *et al*., 1994). Lignocellulosic plant cell wall contains a significant amount of hydroxycinnamic esters of which the most abundant is ferulic acid. These acids are linked at different positions on the arabinose sugar of the hemicellulose polymer, and are cross-linked through diferulate bridges to heteroxylan and lignin polymers (Bunzel *et al*., 2005). The functional roles of ester linkages in plant cells walls have been intensively studied and they are implicated in regulating cellular expansion and plant defence. The cross-linking of ferulic acids are important structural components as they reduce the degradability of cell walls by restricting the accessibility to carbohydrates which reduces enzymatic reactivity (Bunzel *et al*., 2005; Crepin *et al*., 2003; Rumbold, *et al*., 2003).

1.3 Bioethanol Processes

1.3.1 Pretreatment

Conversion of raw feedstock to a valuable product (e.g. bioethanol) requires that carbohydrates are first removed from the lignocellulosic complex (Archadi and Sellstedt, 2008). Due to the resistant structure of cellulose and natural composite structures of lignocellulosics, pretreatment technologies are required to increase

efficacy of enzymatic hydrolysis irrespective of the type of biomass used, to make the carbohydrates available for fermentation (Figure 1.3).

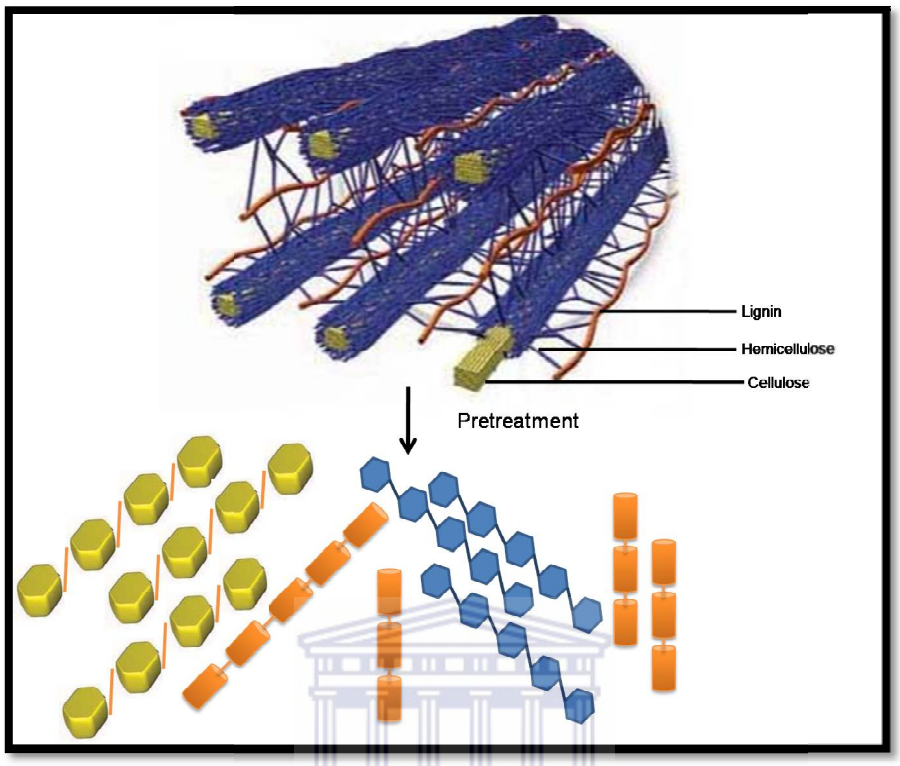


Figure 1.3: Pretreatment renders lignocellulosic biomass easier to hydrolyse using enzymes (*Adapted from* Wyman and Yang, 2009).

There are various methods available to fractionate lignocellulosic biomass. Steam explosion involves high steam pressure which vaporises water within the biomass, thus increasing surface area and facilitating decomposition. Similarly, ammonium fibre explosion involves contact with ammonia under high temperature and pressure conditions which disrupts the biomass. Both steam explosion and ammonia explosion are considered as suitable chemical pretreatments (Mosier, *et al*., 2005). A mechanical pretreatment method, such as grinding, creates increased surface area for enzymes to penetrate the biomass. This can be done by a hammer mill, knife mill, inline homogeniser or with air by a vortex grinder. Radiation, although unlikely to be used commercially, subjects the biomass to gamma rays, electron beams or microwaves (Wyman *et al*., 2005). Fungi such as white rot, brown rot or soft rot can also be used as biological pretreatments, as they naturally attack lignin, rendering the biomass more digestible (Himmel and Bayer, 2009).

1.3.2 Enzymatic Hydrolysis

After pretreatment and mechanical separation, lignocellulosic biomass must be hydrolysed to break the hemicellulose down into simple sugars. Although hydrolysis technologies such as concentrated acid and dilute acid have long industrial histories, recent focus has shifted to enzymatic hydrolysis as a promising method for reducing costs while improving yields. Enzymes catalyse hydrolysis at a multitude of sites therefore a small volume of enzyme can catalyse a large amount of substrate, keeping overall costs low (Vintila *et al*., 2009). Three main enzyme classes are required for complete lignocellulosic breakdown: cellulases, hemicellulases and ligninases (Lopez *et al*., 2002).

1.3.2.1 Cellulases

Cellulases are composed of a complex mixture of enzymes with different specificities to hydrolyse glycosidic bonds. They are responsible for the hydrolysis of cellulose and can be divided into three major enzyme activity classes (Robinovich, 2002; Goyal, 1991). There are endoglucanases or endo-1,4-β-glucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21). Endoglucanases are proposed to initiate randomly at multiple internal sites in the amorphous regions of the cellulose fibre, opening up sites for subsequent attack by the cellobiohydrolases (Wood, 1991). Cellobiohydrolases remove monomers and dimers from the end of the glucose chain. In general, endoglucanases and cellobiohydrolases work in synergy to hydrolyse cellulose. The β-glucosidases hydrolyse glucose dimers and in some cases cello-oligosaccharides to glucose (Robinovich, 2002).

1.3.2.2 Hemicellulases

The complete degradation of hemicellulose requires two groups of enzymes. Firstly, endo-xylanase and β-1,4-xylosidase cleave the xylan backbone and secondly, a variety of accessory enzymes remove side chains and break the cross-links between xylan and other polymers (Figure 1.4) (Topakas *et al*., 2007). These accessory

enzymes are crucial to breakdown because many hemicellulases, such as endo-xylanases, do not cleave glycosidic bonds between xylose molecules that are substituted by other side chains. These side chains must be cleaved off the xylan backbone by a concert of accessory enzymes before complete hydrolysis can occur (Saha, 2003; Prates *et al*., 2001). This process is slow due to the inaccessible and rigid structure of the plant cell wall and the limited commercial availability of efficient cellulolytic and hemicellulolytic enzymes. There are two types of catalytic modules in hemicellulases: glycoside hydrolases (GHs) and carbohydrate esterases (CEs). GHs target glycosidic bonds while the CEs target the ester linkages of acetate or ferulic acid side groups. Accessory enzymes include α-L-arabinofuranosidase, α-glucuronidase, acetyl xylan esterase (AXE) (Wong, 2006) and, the object of this study, ferulic acid esterase (FAE).

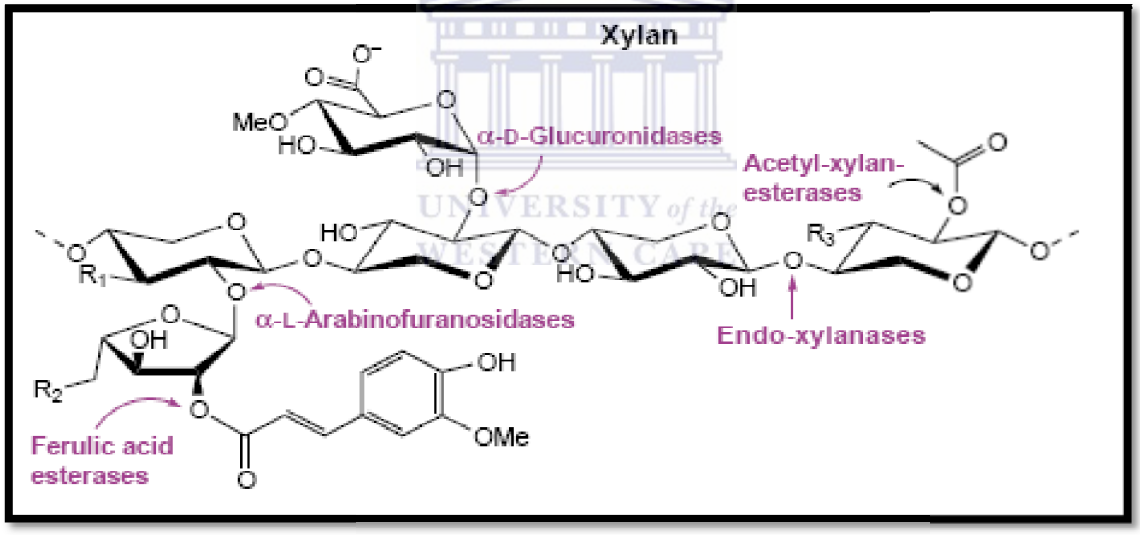


Figure 1.4: The synergistic action of a variety of hemicellulases responsible for complete degradation of xylan (Shallom and Shoham, 2003).

1.3.2.3 Ligninases

Microorganisms break down lignin aerobically through the use of a family of extracellular enzymes collectively termed “ligninases”. Two families of ligninolytic enzymes are widely considered to play a key role in the enzymatic degradation: phenol oxidases such as laccase, and peroxidases such as lignin peroxidase (LiP)

and manganese peroxidase (MnP) (Krause *et al*., 2003; Malherbe and Cloete, 2002). Studies on ligninolytic enzymes are still insuffiecient (de Souza *et al*., 2005).

1.3.3 Production of Bioethanol through Microbial Fermentation

Fermentation is a biological process in which bacterial enzymes catalyse the reduction of simple sugars into lower molecular weight materials such as ethanol. An enormous variety of bacteria, yeasts and fungi are able to ferment 6-carbon sugars (Sanchez and Cardona, 2008; Brown, 2003). Hydrolysates from enzyme degradation contain a cocktail of hexoses and pentoses including glucose, galactose, mannose, D-xylose and L-arabinose. However, not all microorganisms can use hexoses and pentoses for growth, making complete biomass conversion inefficient. Bacteria such as *Escherichia coli*, *Klebsiella oxytoca* and *Zymononas mobilis* have been modified using genetic engineering to produce ethanol from hexose and pentose sugars. Most bioethanol produced by industrial fermentation use the yeast *Saccharomyces cerevisiae*, due to its rapid sugar utilisation and tolerance to ethanol (Becker and Boles, 2003).

1.4 The role of Thermo-stable Enzymes in Fermentation

One of the major advantages of enzymes over industrial catalysts is their potential for high activity at low temperature. Mesophilic enzymes are important in protecting labile substance from damaging reactions (Wood *et al*., 1995). Thermostable enzymes, which originate mostly from thermophilic organisms, are industrially important due to stability at high temperatures (Haki and Rakshit, 2003). Thermostable enzymes have optimum activity temperatures of 45 ºC to 85 ºC (Satpal, 2010), and biotechnological process take advantage of these high temperatures to decrease the risk of contamination by mesophilic organisms which have optimum growth temperatures of 20 ºC to 35 ºC (Demirijian *et al*., 2001). Elevated temperatures influence high specific activity of enzyme, which results in decreased enzyme loading and high stability for reuse and extended shelf-life. Additionally, their rate of reaction is higher because of decreased viscosity and an increased diffusion coefficient of substrates. Due to the increased solubility of

substrates, yield also increase. These enzymes also allow processes with improved integration in terms of heat recovery and recycling of process streams (Haki and Rakshit, 2003; van den Burg, 2003).

In terms of current applications, the most industrially important thermophilic enzymes are esterases, which are mainly used in fine chemical applications for the production of pure compounds (Demirijian *et al*., 2001). Libraries of thermophilic esterases have been constructed together with screening methods. The use of esterases in industry has developed quite rapidly compared with that of other enzyme classes and quick characterisation tools are available for enzyme discovery and activity fingerprinting (Demirijian *et al*., 2001).

Thermophiles have found ways to adapt to extreme environments. They resist protein denaturation by using chaperons which are specialised proteins involved in refolding proteins to their native form and restoring native function (Kumar and Nussinov, 2001). In addition, saturated fatty acids constitute the plasma-membrane of thermophiles. The fatty acids provide an environment which is hydrophobic in nature and maintains the cell wall to enable life at high temperatures. The DNA of thermophiles has a reverse DNA gyrase that makes positive super coils in the DNA strands, which consequently raises the melting temperature of the DNA (Haki and Rakshit, 2003). Increased interactions such as electrostatic, disulphide bridges and hydrophobic interactions, which non-thermotolerant organisms also have, are used by thermophiles to tolerate high temperatures (Haki and Rakshit, 2003; Kumar and Nussinov, 2001). Therefore, thermophiles are a valuable source of thermostable enzymes for use in bioethanol production.

1.5 Ferulic Acid Esterase (FAE)

Ferulic acid esterases (EC 3.1.1.73) are hemicellulases belonging to a sub-class of the carboxylic ester hydrolase family (EC 3.1.1) (Fazary and Ju, 2008; Topakas *et al*., 2007; Polizeli *et al*., 2005). Together, these enzymes are responsible for cleaving the ester bonds between the arabinose substitutions on the polysaccharide main chain of xylans and the ferulic acid moiety (Koseki *et al*., 2009; Aurilia *et al*., 2008; Panagiotou *et al*., 2007; Shallom and Shoham, 2003). Esterases prefer to hydrolyse

short chain (<C_{10}) ester-containing molecules that are partly soluble in water as opposed to lipases which hydrolyse long chain (>C_{10}) molecules and are insoluble in water (Levisson *et al*., 2009). Esterases (EC 3.1.1.1), lipases (EC 3.1.1.3) and various types of phospholipases are classified as lipolytic enzymes (Rhee *et al*., 2005).

FAEs act synergistically with other hemicellulases, such as xylanase, to degrade plant cell walls. A wide range of bacteria and fungi have been reported to produce FAEs depending on the substrates used for growth (Panagiotou *et al*., 2007; Garcia *et al*., 1998). Due to the varied applications of FAEs, it is essential that they function well under different environmental conditions such as pH and temperature (Table1.3). FAEs with different specificities are required due to the diversity of substrates, which may be important in determining the optimal synergy between FAEs and other hemicellulases. When used in a one step fermentation process, all enzymes would need to be compatible with regards to pH, temperature optima and stability profile. For this reason, the discovery of new FAEs with novel characteristic properties still remains an important research area (Shin and Chen, 2006).

1.6 α/β Hydrolase Fold Family Protein

The three dimensional structure of esterases exhibit a definite order of α-helices and β-sheets, known as the α/β hydrolase fold (Figure 1.5). This is a typical example of a tertiary fold adopted by proteins that do not necessarily have similar sequences nor work on similar substrates, but share structural similarity, a conserved arrangement of the Ser-His-Asp catalytic triad and co-factor independent activity (Shallom and Shoham, 2003; Nardini and Dijkstra, 1999). The conical α/β hydrolase fold consists of 8 β-sheets with the second antiparallel to others. The parallel strands β3 and β8 are connected by α-helices packed on either side of the central parallel β-sheet (Aurilia *et al*., 2008; Jaeger *et al*., 1994). The α/β hydrolase fold family includes proteases, lipases, dehydrogenases, peroxidases and more importantly esterases, making it one of the most versatile and widespread protein folds known.

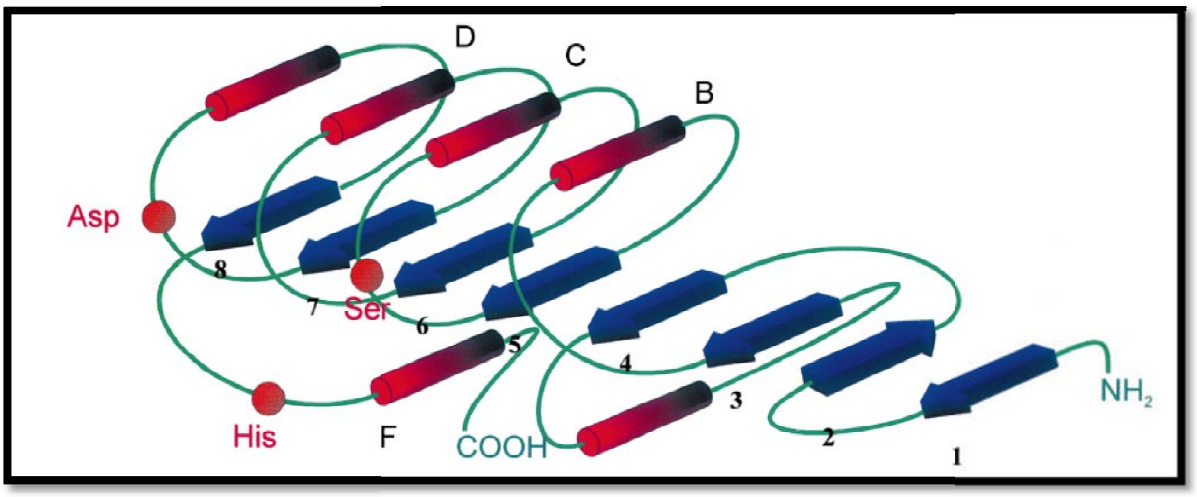


Figure 1.5: Schematic illustration of the α/β hydrolase fold. α-Helices are shown as red cylinders and β-sheets as blue arrows. Orange circles indicate the topological position of active site residues (nucleophile after β5, Asp/Glu after β7 and His in the loop between β8 and αF) (Bornscheuer, 2002).

The catalytic triad is composed of a nucleophile, located in a sharp turn called the 'nucleophile elbow', making it readily accessible by substrates and hydrolytic water molecules. There is a consensus sequence that makes the nucleophile elbow identifiable; Sm-X-Nu-X-Sm (Sm = small residue, e.g Gly, X = any residue and Nu = nucleophile) (Levisson *et al*., 2009; Nardini and Dijkstra, 1999). The binding site for substrates is located inside a pocket on top of the central β-sheet that is typical of this fold. Substrate specificity is determined by size and the shape of the substrate-binding cleft (Levisson *et al*., 2009; Aurilia *et al*., 2008).

1.7 Mode of Action of α/β Hydrolase Fold

The reaction mechanism of hydrolysis of esterase consists of four steps (Figure 1.6) (Aurilia *et al*., 2008; Bornscheuer, 2002; Jaeger *et al*., 1994): (i) binding of substrate to the active site serine results in formation of a transient tetrahedral intermediate, stabilised by interaction with the NH-groups of the catalytic His and Asp residues, (ii) the histidine residue donates a proton and the alcohol component of the substrate is released, (iii) nucleophilic attack by water on the carbonyl C atom of the covalent intermediate occurs (deacylation step), (iv) the negatively charged tetrahedral

intermediate is now stabilised by interaction with the oxyanion hole. Histidine donates a proton to the oxygen atom of the active site serine and this releases the acyl component of the substrate, releasing the free enzyme.

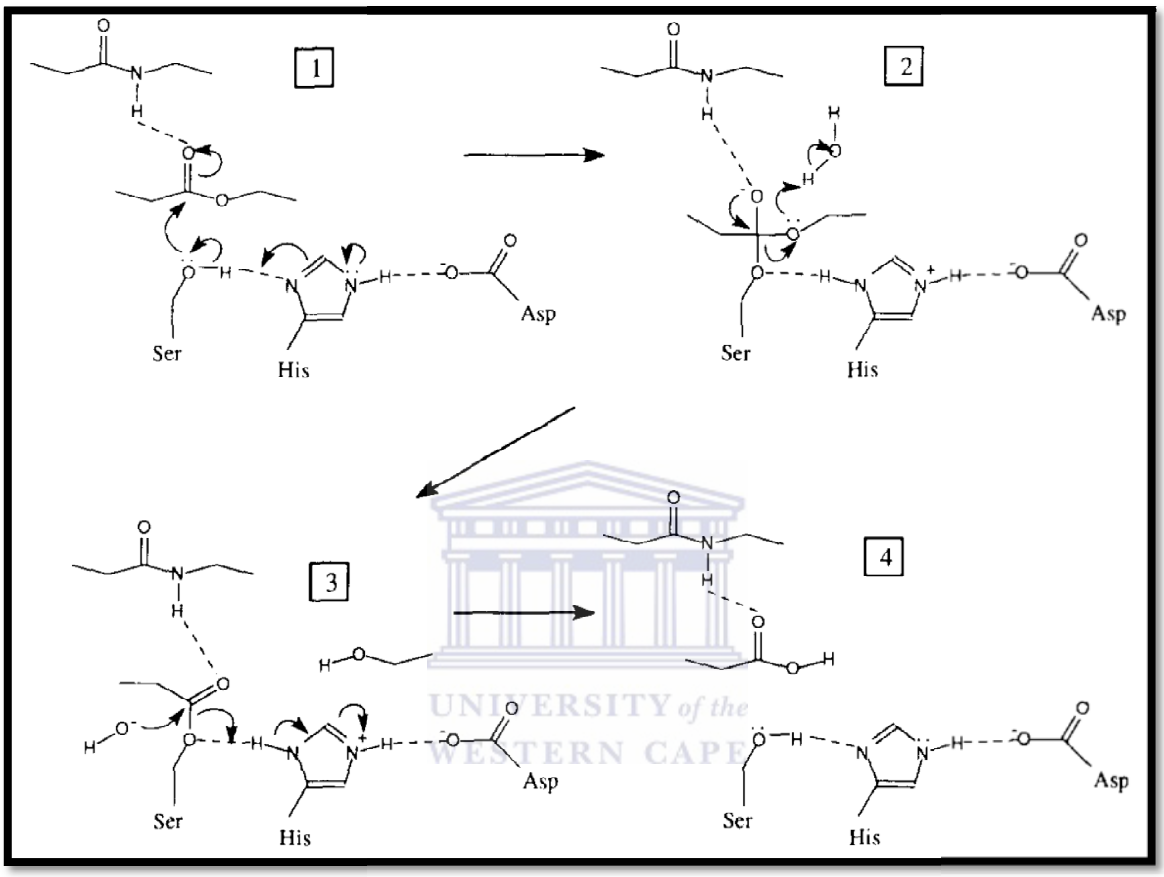


Figure 1.6: Mechanism of ester bond hydrolysis by esterase enzymes (Jaeger *et al*., 1994).

1.8 Classification of FAE

Enzymes are generally classified and named according to the type of reaction catalysed (Enzyme Commission). FAE have been classified as type A or type B, depending on their specificity for aromatic substrates or their ability to release diferulic acids from esterified substrates ([a]Crepin *et al*., 2004). Recently, a more detailed categorization based on substrate consumption and supported by primary sequence identity has been proposed which consists of four subclasses: type A, B, C and D (Topakas *et al*., 2007). FAEs that are active on methyl ferulate, methyl *p*-coumarate and methyl sinapate are classified as type A. They prefer phenolic

moieties of substrate containing methoxy substitutions especially at the *meta*-position(s). Such FAEs have amino acid sequences very similar to lipases and are able to hydrolyse synthetic ferulate dehydrodimers. An example that falls into this group is *Aspergillus niger* FAE-A (Topakas *et al*., 2007). Type B FAEs show complementary activity to type A esterases. Type B esterases also prefer substrates with one or two hydroxyl substitutions such as methyl ferulate, methyl *p*-coumarate and methyl caffeate, but not methyl sinapate. Type A esterases prefer hydrophobic substrates with bulky substituents on the benzene ring. Type B enzymes do not release diferulic acid and show sequence similarities to 1-acetylxylan esterase (carboxylic esterase family). Examples for type B include *Penicillium funiculosum* FAE-B and *Neurospora crassa* FAE-1. Types C and D act on all four synthetic hydroxycinnamic acid methyl esters (ferulic, *p*-coumaric, caffeic and sinapinic acid), meaning that they possess broad substrate specificity, although with a difference in the ability to release 5-5' diferulic acid ([a]Crepin *et al*., 2004, [b]Crepin et al., 2004). Type C enzymes do not liberate diferulic acids from model and complex substrate, whereas type D is capable of hydrolysing dimers. Type C and D show sequence similarities to chlorogenate esterase and xylanase, respectively. Examples for type D are *Piromyces equi* EstA and *Celluvibrio japonicus* esterase D (Faulds, 2010; Koseki *et al*., 2009; Levisson *et al*., 2009; Aurilia *et al*., 2008; Topakas *et al*., 2007; [a]Crepin *et al*., 2004). The phylogenetic tree displaying FAE activity (Figure 1.7) and related sequences suggests an evolutionary relationship between FAE, acetyl xylan esterases and certain lipases (Topakas *et al*., 2007; [a]Crepin *et al*., 2004).

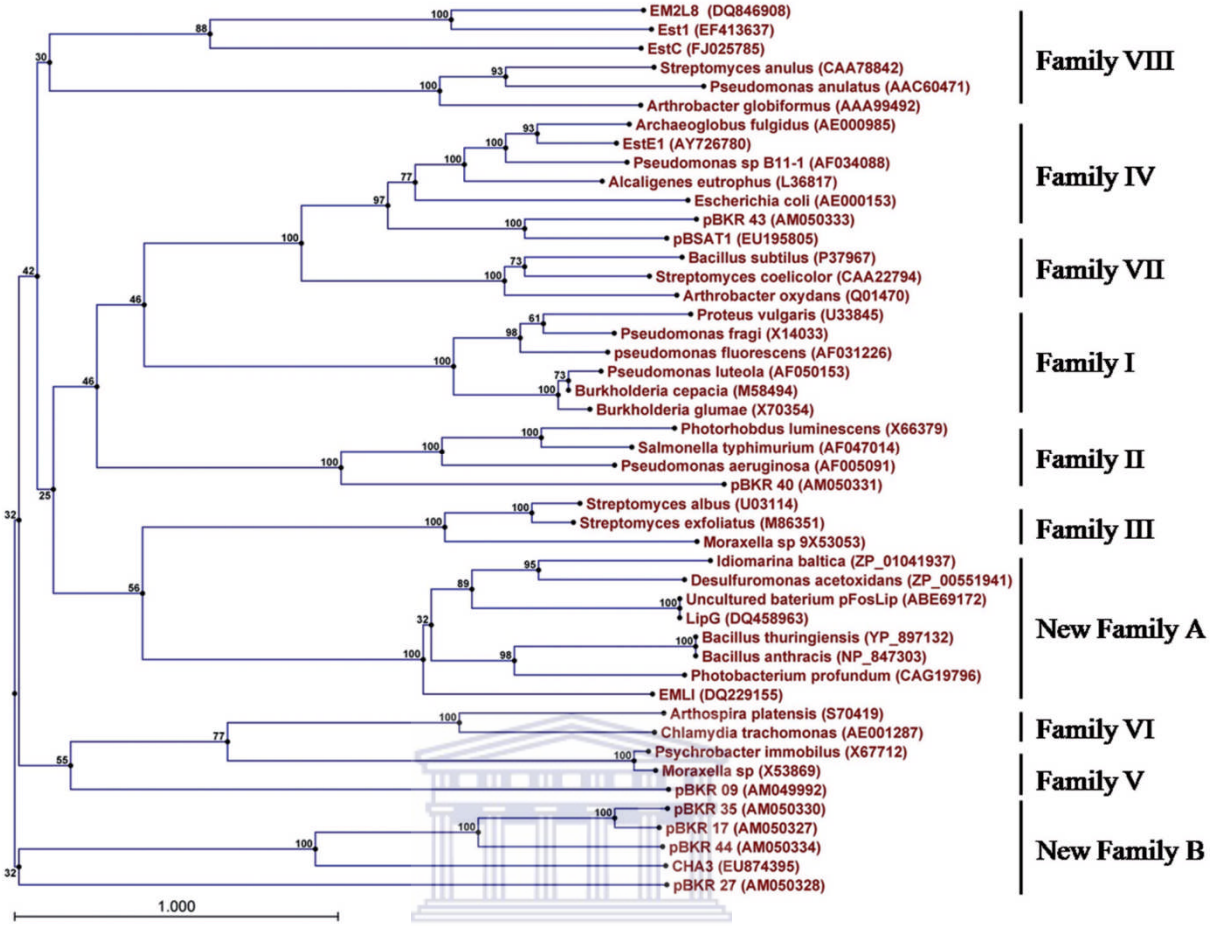


Figure 1.7: Neighbour-joining phylogenetic tree of selected lipolytic sequences, discovered by functional screening of metagenomic libraries based on conserved sequence motifs of bacterial lipolytic enzymes (Tuffin *et al*., 2009).

1.9 FAE's from Microbial Sources

FAEs produced by bacteria are generally secreted into the culture medium. The first reported ferulic acid esterase was from cellulolytic and xylanolytic systems of *Schizophyllum commune* (MacKenzie and Bilous, 1988). The enzyme hydrolysed ester bonds of ferulic acid from the crude hemicellulose preparation of wheat bran. However, a limited number of ferulic acid esterases have been biochemically characterised and the first ferulic acid esterases purified were from *Aspergillus oryzae* and *Streptomyces olivochromogenes*. Both these enzymes had low molecular weights of about 30 and 29 kDa, respectively, and had similar pH optima ranging from 4.5 to 6.0 and 5.5 using methyl ferulate as a substrate (Topakas *et al*., 2007; Christov and Prior, 1993). Up until now, most FAEs have been isolated from

fungal sources rather than bacterial sources. Other microbial FAE sources together with their characteristic properties are listed in Table 1.3 below.

Table 1.3: Physiochemical characteristics of purified ferulic acid esterases isolated from different microbial and fungal sources.

Organism	FAE type	Molecular weight (kDa)	Optimal temperature (ºC)	Optimal pH	pI	Reference
Microbial Sources						
Clostridium thermocellum	-	45k	60	6.0	5.8	Blum *et al*., 2000
Fervidobacterium nodosum Rt17-B1	-	27.5	75	8.5	-	Yu *et al*., 2010
Streptomyces olivochromogenes	Type A	29	30	5.5	7.9	Faulds and Williamson, 1991
Vibrio fischeri	-	37	30	7.5	-	Ranjitha *et al*., 2009
Fungal Sources						
Aspergillus awamori	-	35	45	5.0	3.8	Koseki *et al*., 1998
Aureobasidium pullulans	Type B	210	60	6.7	6.5	Rumbold *et al*., 2003
Fusarium proliferatum NRRL 26517	Type B	31	50	6.5-7.5	-	Shin and Chen, 2006
Fusarium oxysporum	Type C	62	65	6.0	-	Moukouli *et al*., 2008
Sporotrichum thermophile	Type B	33	55-60	6.0	3.5	Topakas *et al*., 2004

1.10 Industrial Application of Ferulic Acid Esterase

Recently, interest has increased in the variety of potential uses for FAEs in the agro-food and energy industries. As already discussed, FAEs play an important role in

biomass degradation and form part of an enzyme system that acts in concert with a variety of other hemicellulase and cellulase enzymes, increasing overall sugar yield following breakdown (Christov and Prior, 1993).

FAE have also been implicated in the paper and pulp industry, where they solubilise lignin-polysaccharide complexes in paper processing. Together with glucanases and oxidases, FAEs also improve bread-making quality and related cereal processing (Topakas *et al*., 2007; Damirjian *et al*., 2001). Ferulic acids, as products of FAE, contribute to the importance of these enzymes as they have a potential application in medicine. Ferulic acid and its derivatives are very strong antioxidants and have gel-forming properties which act as potential protective agents against skin damage and wound treatment (Wong, 2006; Wang *et al*., 2004), as well as the antimicrobial and photo-protectant properties they possess (Aurilia *et al*., 2008; Topakas *et al*., 2007). Besides being exploited as a hydrolase, FAE was recently shown to be an excellent catalyst in sugar-phenolic ester production, and could potentially also be used to functionalise sugar polymers by adding phenolic derivatives onto the natural biopolymers (Shin and Chen, 2006).

Project Aims

The focus of this project was to characterise a novel thermophilic hemicellulytic gene and enzyme which could ultimately be applied towards the degradation of various lignocellulose feedstocks to fermentable products for application in biofuels production. In this study, a ferulic acid esterase was isolated from a metagenomic fosmid library constructed with DNA extracted from a Malawian hot spring soil. It is envisaged that this enzyme could be applied on an industrial scale in the biological degradation of plant biomass, thereby enhancing the efficiency of the hydrolytic process.

The specific objectives of this study are summarised below:

- To confirm ferulic acid esterase activity of a fosmid clone, using activity-based screening.
- Obtain gene and protein sequence using Roche 454 sequencer and analyse using bioinformatic tools.
- Design primers and amplify the gene of interest, for cloning and expressing gene into a bacterial expression vector.
- Purify bacterial protein such that it can be used in enzymatic assays.
- Characterise the ferulic acid esterase in terms of pH and temperature optimal range, and determine substrate specificity and enzyme kinetic data.

Chapter 2

Materials and Methods

2.1 Chemicals and Enzymes

The *para*-nitrophenyl (*p*-NP)-fatty acyl ester substrates; *p*-Np-acetate (C_2), *p*-Np-propionate (C_3), *p*-Np-octanoate (C_8), *p*-Np-decanoate (C_{10}), *p*-Np-laurate (C_{12}), *p*-Np-myristate (C_{14}) in addition Triton X-100 were purchased from Sigma (South Africa). Methyl esters of ferulic (MFA), *p*-coumaric (M*p*CA), caffeic (MCA) and sinapic (MSA) acids were purchased from Apin Chemicals (South Africa). Total protein assay solution was purchased from Bio-Rad (South Africa). All other chemicals and solvents used were of analytical grade and commercially available. Biolabs supplied culture media. DNA and PCR purification kit was purchased from Giagen and GE Healthcare (South Africa), respectively. DNA size markers and all DNA modifying enzymes (polymerase and restriction endonuclease) were purchased from Fermentas Life Science (South Africa). The 0.2 nm nitrocellulose filter was purchased at Millipore. Oligonucleotide primers for polymerase chain reaction (PCR) were synthesised by Inqaba Biotec (South Africa).

2.2 Media

Luria-Burtani (LB) broth consisted of 1 % [w/v] tryptone, 0.5 % [w/v] yeast extract and 0.5 % [w/v] NaCl. The LB agar consisted of LB broth with the addition of 1.5 % [w/v] bacteriological agar. Tributyrin agar consisted of 1 % [w/v] tryptone, 0.5 % [w/v] yeast extract, 0.5 % [w/v] NaCl, 1.5 % [w/v] bacteriological agar, 1 % [v/v] tributyrin and 1 % [w/v] gum arabic. After dissolving the ingredients together, the mixture was sonicated for 10X 30 sec interval. SOB broth consisted of 2 % [w/v] tryptone, 0.5 % [w/v] yeast extract, 0.05 % [w/v] NaCl, 0.02 % KCl. All components were mixed together with distilled water and pH was adjusted with 1M NaOH to pH 7.0. Media was autoclaved at 121 ºC for 20 min. SOC was prepared from SOB broth with the addition of filter sterilised 0.5 % [w/v] 2M $MgCl_2$ and 2 % [w/v] 1M glucose. Ethyl ferulate (EF) agar consisted of 1 % [w/v] tryptone, 0.5 % [w/v] yeast extract and 0.5

% [w/v] NaCl and 1.5 % [w/v] bacteriological agar. Once the media was autoclaved and cooled to 55 ºC, 0.4 % [w/v] of filter sterilised Ethyl 4-hydroxy-3-methoxycinnamate (EF) (dissolved in dimethyl sulphoxide (DMSO)) was added before pouring plates. Olive oil-Rhodamine B lipase agar consisted of 2 % [w/v] olive oil and 0.2 % [w/v] gum arabic dissolved in water and sonicated for 5 cycles at 30 seconds each. This solution was then mixed with filter sterilised 2 % [w/v] MOPS in a separate bottle adjusted to pH 7.0 with NaOH. Three percent agar was added and then autoclaved. Once autoclaved, 2 % [w/v] Rhodamine B dye was added.

All media types were cooled to 55 ºC after autoclaving, and when necessary, the appropriate filter sterilised antibiotic and inducer was aseptically added before pouring plates. All agar plates were allowed to solidify and dry at 25 ºC for 3 hours before use.

2.3 Bacterial Strains, Plasmids and Growth Conditions

Final concentration of antibiotics was: (unless otherwise stated) 12.5 µg/ml chloramphenicol (cam), 100 µg/ml ampicillin (amp) and 50 µg/ml kanamycin (kan). Chloramphenicol was prepared with 100 % ethanol, ampicillin and kanamycin with distilled water. Arabinose (0.01 %) and Isopropyl-β-D-thiogalactopyranoside (IPTG) (0.5 mM or 1 mM) were prepared using distilled water and filter purified.

Bacterial strains were grown in broth or on solid media supplemented with the appropriate antibiotic. Unless otherwise stated, cultures were incubated at 37 ºC. Strains grown in broth were agitated at 150 to 225 rpm. All clones were stored in two forms: (i) as colonies on agar plates and (ii) as glycerol stocks at -80 ºC. The plates were stored at 4 ºC and were maintained for sub-culturing. The glycerol stocks were prepared by adding 500µl of 50 % glycerol to 500µl of cell culture. The DNA was stored in 1X TE buffer at 4 ºC for short term storage and at -20 ºC for long term storage.

Table 2.1: Strains and plasmids used in this study

Characteristics		Source
Bacterial strains	**Genotype**	
E. coli		
EPI-300	F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ^- *rps*L *nup*G *trf*A *ton*A *dhfr*	Epicentre Biotechnology (USA)
Genehog	F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *fhu*A::IS2 (confers phage T1 resistance)	Invitrogen Life Technologies (USA)
BL21(DE3)pLysS	F^-, *omp*T, *hsd*S_B (r_B- m_B-), *dcm*, *gal*, (DE3), pLysS, Cam®	Invitrogen Life Technologies (USA)
Rosetta(DE3)pLysS	F^-, *omp*T, *hsd*S_B (r_B- m_B-), *gal*, *dcm*, (DE3), pLysSRARE, (Cam^R)	Novagen (USA)
Plasmids/vectors		
pCC1FOS	$Chloramphenicol^R$ 12.5 µg/ml	Novagen (USA)
pUC19	$Ampicillin^R$ 100 µg/ml	Novagen (USA)
pJET1.2	$Ampicillin^R$ 100 µg/ml	Novagen (USA)
pET21a(+)	$Ampicillin^R$ 100 µg/ml	Novagen (USA)

2.4 DNA Purification

DNA from the clone (Try 11) showing lipolytic activity on tributyrin agar was purified using the alkaline lysis protocol (Sambrook *et al*., 1989). DNA concentration was measured using a NanoDrop Spectrophotometer as recommended by the manufacturer (Thermo Scientific). The fosmid DNA was digested with *Eco*RI and *Hind*III and then loaded onto a 1 % agarose gel in 1X TAE buffer overnight at 50 V. The DNA fragments were calculated according to their migration in the gel as compared to that of the DNA molecular marker, Lambda DNA which is restricted with

*Pst*I endonuclease. Visualisation of DNA was carried out using the Alphalmager 2000 digital imaging system (Alpha Innotech, San Leadro, CA).

2.5 Transposon Mutagenesis

An attempt was made to obtain the clone insert's sequence using mutations at random locations by *in vitro* transposon mutagenesis according to manufacturer's instructions (GPS®- Mutagenesis system, New England Biolabs, UK).

2.6 Bioinformatics and Analysis

In an effort to identify the origin of the insert DNA, Try 11 clone was end-sequenced with the T7 forward and pCC1Fos reverse primer [Table 2.2] flanking the insert. Sequencing reactions were carried out by the University of Cape Town sequencing service using the Roche 454® GS-FLX sequencing platform of Inqaba Biotechnical Industries.

The fosmid was sequenced using the Roche 454® sequencer. Sequence assembly, contig editing and the end-sequences were aligned against 454 sequences using the CLC Genomics Workbench software. SoftBerry (Salamov and Solvyev, 2000), was used to predict open-reading frames (ORFs) and proteins in the sequence which was compared to other proteins in the National Centre for Biotechnology Information (NCBI) database using BLASTp (Altschul *et al*., 1997). Once the final nucleotide contig was constructed, ExPASY (Gasteiger *et al*., 2005) was used to translate the nucleotide into protein. Multiple sequence alignments using ClustalW (Larkin *et al*., 2007) were used to determine conserved regions in the gene. Pfam was used to find matches to the predicted protein family domains (Finn *et al*., 2008). The PROSITE motif search was also used to identify possible matches based on conserved motifs found in the protein (Hulo *et al*., 2007). PSIPRED was used to predict the secondary structure of the protein (McGuffin *et al*., 2000). TMBETA-NET was used to discriminate outer membrane proteins and predict transmemebrane β-strands in an outer membrane protein from the amino acid sequence (Gromiha *et al*., 2005). The signal peptide was predicted using the SignalP 3.0 server (Emanuelsson *et al*.,

2007). Rare Codons Caltor was used to predict rare codon content. The translated nucleotide sequence was used for homology modelling using the Swiss-protein modeller program (Schwede *et al*., 2003) and Interactive 3D-JIGSAW (Bates *et al*., 2001). RAMPAGE was used to assess the accuracy of the model by generating a Ramachandran plot (Lovell *et al*., 2003). The model was superimposed onto the template using the PyMol program. Phylogenetic analyses were conducted using MEGA version 4 (Tamura *et al*., 2007) and were used in the construction of phylogenetic trees based on neighbour-joining analysis. Protein sequences of various enzymes or subunits were extracted from the National Centre for Biotechnology Information (Washington, D.C).

2.7 Polymerase Chain Reaction (PCR) of RHgene34

Primers (RhuF2 and LaniR2) [Table 2.2] were designed with restriction cut sites *Nde*I and *Hind*III, respectively. The primers were used to amplify the gene from the original fosmid using PCR. The PCR reaction mixtures were placed in a Gene Amp® Master Cycler Gradient Eppendorf machine and the following cycling conditions were used: initial denaturation at 95 ºC for 1 min, followed by 35 cycles of denaturation at 95 ºC for 30 sec, annealing at 59 ºC for 1.5 min and extension at 72 ºC for 1 min. A final elongation step of a single cycle was performed at 72 ºC for 5 min and the holding temperature was at 15 ºC for 5 min.

The target DNA was amplified using 0.2 ml thin walled tubes in either Gene Amp® Master Cycler Gradient or ThermoHybaid PCR sprint machines. PCR reactions (20-50 µl) contained 50-100 ng fosmid DNA, 1X PCR buffer (from 10X buffer consisting of 200 mM Tris [pH 8.8], 100 mM KCl, 100 mM $[NH_4]_2SO_4$, 2 mM $MgSO_4$ and 1 % [w/v] Triton-X-100), 0.5 µM of RhuF2 and LaniR2 primers [Table 2.2], 0.2 mM dNTPs mixture (dATP, dCTP, dGTP and dTTP) and 0.5 µl labtaq DNA polymerase. For control purposes, forward primer, reverse primer and a negative control (a reaction mixture containing all reagents except template) was routinely included. Reactions were made up to the appropriate final volume using ddH_2O.

Gel electrophoresis of the resulting PCR products showed that a fragment corresponding to the expected size was successfully amplified.

Table 2.2: Primers used in this study.

Primers	Sequence (5'-3')	Specific annealing temperatures (ºC)	Source
T7 forward	GGATGTGCTGCAAGGCGATTAAGTTGG	N/A	Novagen (USA)
pCC1FOS reverse	TACGCCAAGCTATTTAGGTGGTGAGA	N/A	Novagen (USA)
RhuF2	GGCAT**ATG**CCATATATTTCCACCGAAG	67.7	This study
LaniR2	GGAAGCTTTGGCTTTTGAATTAATTG	64.1	This study
pJET1.2 forward	CGACTCACTATAGGGAGAGCGGC	N/A	Fermentas (USA)
pJET1.2 reverse	AAGAACATCGATTTTCCATGGCAG	N/A	Fermentas (USA)
pET21 T7 Promoter	TAATACGACTCACTATAGGG	N/A	Novagen (USA)
pET21 T7 Terminator	GCTAGTTATTGCTCAGCGG	N/A	Novagen (USA)

The restriction sequences for *Nde*I and *Hind*III are underlined. The start codon is in bold.

2.8 Cloning of RHgene34

The PCR product was cloned into the CloneJET™ PCR Cloning Kit (Fermentas) according to manufacturer's instructions. Restriction enzymes, *Nde*I and *Hind*III, were used to digest selected clones. Clones showing the expected pattern were sequenced. An *E. coli* pET21a(+), was used to achieve expression of RHgene34 protein. pET21a(+) contains a C-terminal His tag, which will be useful for nickel affinity chromatography purification

2.9 Preparation of competent *E. coli* cells

2.9.1 Preparation of Electrocompetant Cells

Electrocompetent EPI-300 *E. coli* cells were prepared by inoculating a single freshly streaked colony into 5 ml SOB broth and culturing overnight at 37 ºC at 250 rpm. Half of the overnight culture was inoculated into a 1 l sterile flask containing 250 ml SOB broth and the cells were cultured at 37 ºC (250 rpm) for 45 - 60 min until OD_{600nm} was between 0.5 - 0.7. Cells were kept on ice from this point on and transferred to chilled Corning bottles. The cells were harvested at 4000 x *g* for 15 min at 4 ºC. As much supernatant as possible was discarded and the cells were gently resuspended in 1 l of ice-cold sterile 10 % [v/v] gylcerol and harvested as before. The cells were centrifuged again then resuspended in 0.5 l ice-cold sterile 10 % [v/v] glycerol and centrifuged as before. Pellets were re-suspended in a total of 250 ml ice-cold 10 % [v/v] glycerol and harvested as before. Ice-cold 10 % [v/v] glycerol (3 – 4 ml) was used to resuspend the cells pellets. The cell suspension was kept on ice and aliquoted (100 µl) into 1.5 ml Eppendorf tubes and stored at -80 ºC until needed. To test the efficiency of the electrocompetent cells, 1 µl of pUC19 vector DNA (120 ng/µl) was electroporated with 50 µl EPI-300 cells.

2.9.2 Preparation of Competent *E. coli* Cells by $CaCl_2$ Treatment

Glycerol stocks of appropriate *E. coli* cultures were streaked onto the surface of an LB agar plate. The plate was incubated for 24 hrs at 37 ºC. Pre-culturing was performed by transferring a single colony into 5 ml LB medium. The culture was incubated overnight at 37 ºC in a shaking incubator and 500 µl of the overnight culture was inoculated into 100 ml SOB medium in a 1 l flask. The culture was incubated at 37 ºC until an optical density ($OD_{600\ nm}$) of 0.3 - 0.6 was attained. The flask was placed on ice and the culture was split into 4 equal volumes. Cells were kept on ice in all subsequent steps. The cultures were centrifuged at 4 ºC for 5 min at 5000 rpm. The supernatant was discarded and the pellet was resuspended in 100 ml ice cold 0.1 M $CaCl_2$ and held on ice for 1 min. Cells were collected as before and resuspended in 50 ml of ice cold 0.1 M $CaCl_2$ and held on ice for 90 min. The

cultures were centrifuged at 4 ºC at 5000 rpm for 5 min and placed on ice. The supernatant was discarded and the pellet was resuspended in 10 ml ice cold 0.1 M $CaCl_2$. A volume of 10 ml of ice-cold sterile glycerol was added, the cells were resuspended and aliquots were stored at -80 ºC (Sambrook *et al*., 1989).

2.10 Transformation of *E.coli* cells

2.10.1 Electroporation

E. coli EPI-300 electrocompetent cells (50 µl) were transformed by the addition of 1 µl fosmid DNA then gently mixed. The mixture was cooled on ice for 5 min then transferred into a pre-cooled 0.1 cm sterile electroporation cuvette (Bio-Rad Laboratories, CA, USA). Electroporation was performed under the following conditions: 1.8 kV, 25 µF and 200 Ω. After electroporation, 1 ml of SOC broth, pre-warmed to 37 ºC, was immediately added to the cuvette and mixed. The cells were transferred to a 2 ml sterile Eppendorf tube and incubated at 37 ºC for 30 – 60 min with agitation. Cells were plated in aliquots of 200, 100 and 50 µl onto LB-agar plates.

2.10.2 Chemical Transformation

Plasmid DNA (5 µl) was added directly to 50 µl of competent cells, incubated on ice for 30 min and heat-shocked at 45 ºC for 30 sec. Cells were incubated on ice for a further 2 min and 950 µl of LB broth was added and the cells incubated for 50 min at 37 ºC with agitation. The recovered cells were plated on media.

2.11 Preparation of Cell-Free Extract for RHgene34 Purification

Random transformants with zones of hydrolytic activity were selected for small scale expression. Colonies were grown overnight in 5 ml LB medium. Five hundred microlitre of this culture was used to inoculate 12 ml LB medium until OD_{600nm} reached 0.6 - 0.8. The culture was divided into two 6 ml volumes; one was induced

with 1 mM IPTG, the other uninduced served as the negative control. The tubes were incubated with shaking (150 rpm) at 37°C. Cultures were sampled at 2 hr intervals for up to 8 hrs and overnight for analysis on SDS-PAGE gels. Each sample (1 ml) was harvested by centrifugation at 6000 x g for 10 min. The pellet was stored at -20 ºC and the supernatant at 4 ºC. The volume of sample to be loaded on a 12 % SDS-PAGE gel was calculated using the following equation:

$$\text{Volume (µl)} = \frac{180}{OD_{600nm} \times \text{concentration factor}}$$

2.11.1 Enzymatic Lysis

Frozen cell pellets were resuspended in BugBuster extraction (5 ml/g of pelleted cells) (Novagen, USA) containing benzonase nuclease (1 µl/ml) (Novagen, USA) and gently agitated at room temperature for 30 min. The lysed cells were centrifuged at 10,000 x g for 10 min and the supernatant was transferred to a sterile tube and stored at 4 °C.

2.11.2 Mechanical Disruption (Sonication)

Frozen cell pellets were resuspended in 1:50 volume of PBS buffer and sonicated on ice in cycles of 30 sec pulse and 30 sec pause for 5 min per 50 ml of culture volume. The lysed cells were centrifuged at 10,000 x g for 10 min and the supernatant was transferred to a sterile tube and stored at 4 °C. The cell-free extract was analysed by SDS-PAGE.

2.12 Protein Expression and Purification

Cell cultures were incubated to an OD_{600nm} of 0.7 at 37 ºC. The culture was centrifuged at 13 000 x *g* and the pellets were resuspended in Bugbuster [5 ml for every 1 g of pelleted cells] (Novagen, USA) and Benzonase nuclease [1 U/ml Bugbuster] (Novagen, USA) and incubated at room temperature for 30 min with

gentle agitation. The lysed cells were centrifuged at 13 000 x *g* for 5 min and the supernatant was transferred to a sterile Eppendorf tube.

His-Bind resin (Novagen, USA) was completely resuspended by gentle inversion. Three millilitres of the slurry was transferred to a purification column and packed by gravitational flow to a final bed volume of 1.5 ml. To charge and equilibrate the column the following sequence of washes were allowed to flow through the column by gravity flow:

4.5 ml sterile deionised water

7.5 ml 1 x charge buffer (8 x = 400 mM $NiSO_4$)

4.5 ml 1 x binding buffer (8 x = 4 M NaCl, 160 mM Tris-HCl, 40 mM imidazole [pH 7.9])

After draining of the binding buffer, prepared cell extract was added to the column and allowed to flow through. The column was washed with 15 ml 1 x binding buffer, 9 ml 1 x wash buffer (8 x = 4 mM NaCl, 480 mM imidazole, 160 mM Tris-HCl [pH 7.9]), 9 ml 1 x elute buffer (4 x = 4 M imidazole, 2 M NaCl, 80 mM Tris-HCl [pH 7.9]) and finally 9 ml 1 x strip buffer (4 x = 2 M NaCl, 400 mM EDTA, 80 mM Tris-HCl [pH 7.9]). After use, the column was washed with sterile demineralised water and stored in 1 ml 20 % ethanol at 4 ºC.

All fractions were collected. The eluted fraction was transferred to a dialysis cassette (Thermo Fisher Scientific) and dialysed for 6 hours in 1 l dialysis buffer (50 mM Tris-HCl [pH 7.0], 10 % glycerol), then overnight against 1 l of fresh buffer. The recovered fraction was stored at 4 ºC and used in subsequent enzyme assays with dialysis buffer as a control.

2.13 SDS-PAGE Analysis

Vertical SDS-PAGE gels were cast with a 12 % stacking gel (1.5M Tris-HCl [pH 8.8], 30 % [w/v] acrylamide, 0.8 % [w/v] bis-acrylamide, 20 % [w/v] SDS, 10 % [w/v] ammonium persulphate [Sigma], 0.1 % [v/v] TEMED [Fluka]) and a 4 % stacking gel (0.5 M Tris-HCl [pH 6.8], 30 % [w/v] acrylamide, 0.8 % [w/v] bis-acrylamide, 20 %

[w/v] SDS, 10 % [w/v] ammonium persulphate , 0.1 % [v/v] TEMED). Samples were mixed with an equal volume of 2 x loading dye (80 mM Tris-HCl [pH 6.8], 10 % [v/v] mercaptoethanol, 2 % [v/v] SDS, 10 % [v/v] glycerine, 0.2 % [w/v] bromophenol blue), vortexed and heated to 95 ºC for 5 - 10 min.

Samples of 10 µl were loaded on gels and electrophoresed at 60 V in 1 x running buffer (0.25 mM Tris-HCl, 2 M glycine, 1 % [w/v] SDS) for 30 min through the stacking gel. Electrophoresis continued through the separating gel at 100 V for approximately 2 hours. The gel was stained with Coomassie brilliant blue stain (0.25 % [w/v] Coomassie blue R250, 50 % [v/v] methanol, 10 % [v/v] acetic acid) for 45 min and de-stained overnight with SDS destain (50 % [v/v] methanol, 10 % [v/v] acetic acid) (Sambrook *et al*., 1989). The size of the proteins was determined according to their migration in the gel as compared to that of the protein ladder used (PageRuler™ stained protein ladder [Fermentas]).

2.14 Enzyme Assays

The His-tag purified protein was used in enzyme assays. Dialysis buffer was used as control. Nine hundred and seventy microlitres of buffer (100 mM NaCl, 100 mM Na_2PO_4), 10 µl of 50 mM substrate (dissolved in acetonitrile) and 10 µl of 1 % Triton X 100 were added to a 1 ml cuvette, mixed thoroughly by inversion and the absorbance measured by monitoring the production of *p*-nitrophenoxide at 405 nm over a period of 1 min using a Cary 50 Bio UV-Visible Spectrophotometer and Cary Win VU Kinetic Application 3.0 software (Australia), with a water-heated cuvette block at 25 ºC. This mixture was used as the blank and a new cuvette was used for each background/blank measurement. After addition of enzyme, the change in absorbance units per minute was measured for each substrate. Activity was calculated using the following equation:

$$A = \varepsilon . C . l$$

Where:

ε is the extinction co-efficient of *p*-nitrophenol (*p*-Np) which was experimentally determined by Dr. C. Heath and X.P. Hu as 13 900 $Lmol^1cm^1$.

A is the absorbance of the enzyme reaction based on V_{max} and the volume of enzyme used in the 1 ml assay.

C is the substrate concentration of $Lmol^{1}$ in the enzyme reaction.

l is the path length of light through the cuvette and has a value of 1 cm.

One unit of enzyme activity was defined as the amount of activity required to release 1 µmol *p*-Np per minute under the above condition. For the enzyme assays, pellet extracts were assayed with *p*-NP esters C_2 (acetate), C_3 (propionate), C_8 (octanoate), C_{10} (decanoate), C_{12} (laurate) and C_{14} (myristate).

2.14.1 Substrate Specificity

Substrates specificity of the purified RHgene34 protein was performed using *p*-NP esterified with fatty acid of different carbon chain lengths. The release of *p*-NP was measured by determining the absorbance at 405 nm. The assay was incubated at 25 ºC.

2.14.2 Determining the Effect of pH on Enzyme Activity

To determine the optimum pH of the pure enzyme, the following buffers were used in the reaction: 50 mM Glycine-HCl (pH 1.0 - 3.0), 50 mM sodium acetate (pH 3.0 - 5.0), 50 mM MES (pH 5.0 - 7.0), 50 mM Tris-HCl (pH 7.0 - 9.0) and 50 mM CAPS (pH 9.0 - 11.0. C_{10} was used as the substrate.

2.14.3 Determining the Effect of Temperature on Enzyme Activity

The optimal temperature of the purified enzyme was determined for a temperature range of 15 to 65 °C in which C_{10} was used as the substrate. Thermostability of the enzyme was determined at 35, 45 and 55 ºC. Enzyme samples were incubated for 5, 15, 30, 60 min at each temperature and the residual activities were determined using the standard *p*-Np assay at 25 ºC.

2.14.4 Determining Catalytic Efficiency

Enzyme kinetics was determined using *p*-Np esters at various substrate concentrations ranging from 0 – 2.5 or 3.0 mM). The standard enzyme assay in 50 mM sodium phosphate buffer (pH 7.5) incubated at 25 °C for 1 min was performed. The data was analysed using the Michaelis-Menten plot (Menten and Michaelis, 1913) to determine K_m and V_{max} values. The Hill plot (Hill, 1910) was used to determine $K_{0.5}$ and Hill coefficient (*h*).

2.15 Fast Performance Liquid Chromatography (FPLC) Determination of Quaternary Structure

The His-tag purified protein was dialysed overnight at 4 ºC against 50 mM Tris-HCl (pH 7.5) and 1 ml was used for analysis. Albumin (66 kDa), Cytochrome C (29 kDa) and Carbonic anhydrase (12.4 kDa) were used as standards where 0.2 mg was dissolved in 1 ml of the dialysis buffer.

Size exclusion chromatography was performed on the ÄKTA FPLC purifier system (Amersham Pharmacia Biotech Inc., USA) Sephadex^{TM} 75 HR 10/300 resin column with volume of 24 ml. The separation of the proteins was achieved using 50 mM Tris-HCl (pH 7.5), 1 M NaCl as mobile phase at a flow rate of 0.4 ml/min; column pressure limit was at 1.8 mPa. Detection of protein peaks was performed using UV detector. The Unicorn ver. 4.0 software (Amersham Pharmacia Biotech Inc., USA) was used to analyse the data. Selected fractions (1 ml each) were assayed to verify activity of peak.

2.16 High Performance Liquid Chromatography (HPLC) Analysis

To evaluate the capacity of RHgene34 gene product to hydrolyze the ester linkages of methyl *p*-coumarate, methyl caffeate, methyl ferulate and methyl sinapate, the HPLC-based detection system was utilised. Reaction mixtures (1 ml) were prepared with each methyl substrate (50 mM) in phosphate buffer (100 mM sodium phosphate, 100 mM NaCl; pH 7.5), 1 % Triton X-100 and reactions were initiated by the addition of RHgene34 protein (325 mg). The reactions were incubated at 25°C

for 30 min and stopped by incubation for 5 min at 70°C. The reactions were analysed on the Ultimate 3 000 HPLC. Methyl substrates were run as controls.

To detect the acid products present after lipolytic hydrolysis, the sample was injected into a reverse phase (C18) column and eluted using 30 % methanol, 1 % acetic acid as a mobile phase, at 48 °C, flow rate of 0.8 ml/min and a pressure of 91.7 bar limited by 120 bar. Samples were injected into a 20 µl super loop. The detection of eluted acids was performed with an ultraviolet detector at 220 nm. Results were analysed with Chromeleon (c) Dionex ver. 6.80 software.

CHAPTER 3

Screening and Sequence Analysis

Results and Discussion

3.1 Metagenomic Fosmid Library Verification

A high molecular weight DNA library was constructed using the fosmid CopyControl pCC1FOS vector in the EPI-300 *Escherichia coli* strain (Hu, X.P., 2010). The metagenomic library was functionally screened for hydrolysing activity on tributyrin indicator plates. The plates were randomly monitored during a two day incubation period at 37 ºC for the presence of zones of clearing around the colonies (Figure 3.1). Two clones, Try 11 and Try 12, formed halos on the indicator plates.

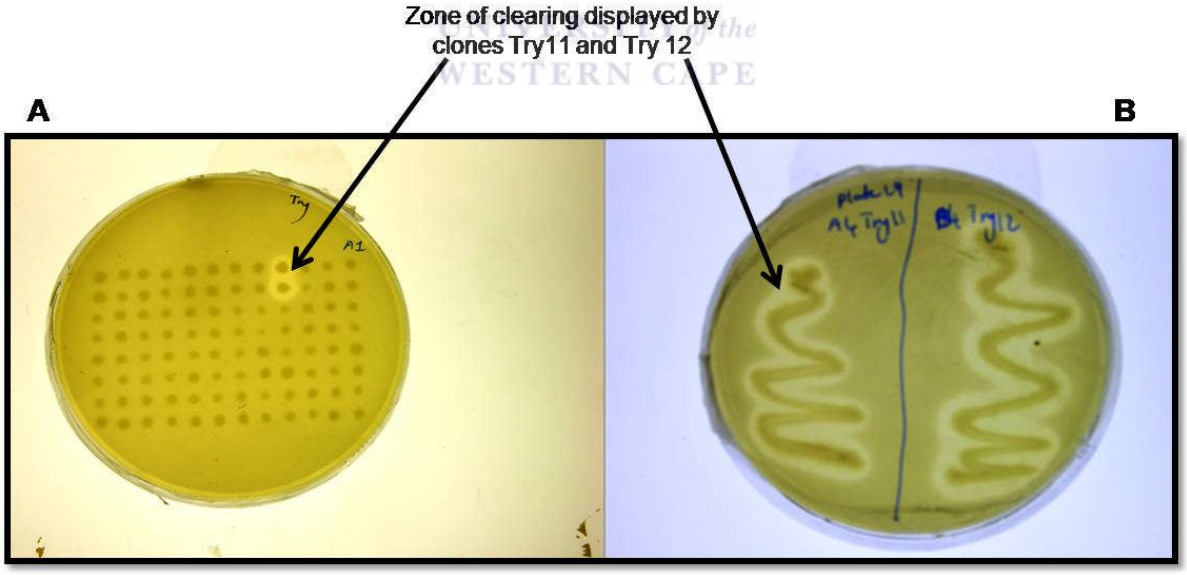


Figure 3.1: Screening a metagenomic library for tributyrin degrading enzymes (A). The two clones identified were Try 11 and Try 12, shown here depicting zones of clearing, after 2 days of 37 ºC incubation (B).

To distinguish between general lipolytic and ferulic acid esterase activity, the clones were streaked on LB plates containing ethyl 4-hydroxy-3-methoxycinnamate (ethyl ferulate), a substrate specifically degraded by ferulic acid esterase. Both clones (Try 11 and Try 12) had the ability to degrade ethyl ferulate (Figure 3.2, results for Try 12 are not shown).

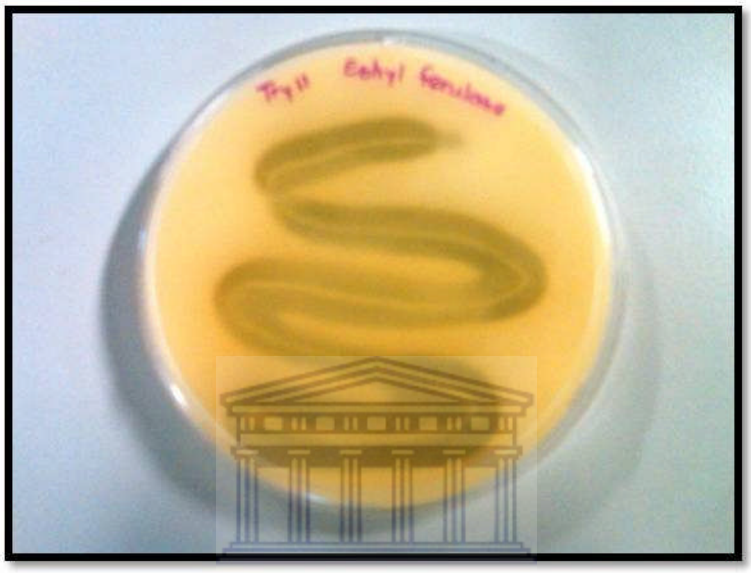


Figure 3.2: Try 11 clone exhibiting ferulic acid esterase activity on ethyl ferulate agar after 2 days incubation at 37 ºC.

Fosmid extractions were performed on the clones. Restriction endonuclease digestions were used to assess the insert size of the clones, using *Eco*RI and *Hind*III (Figure 3.3).

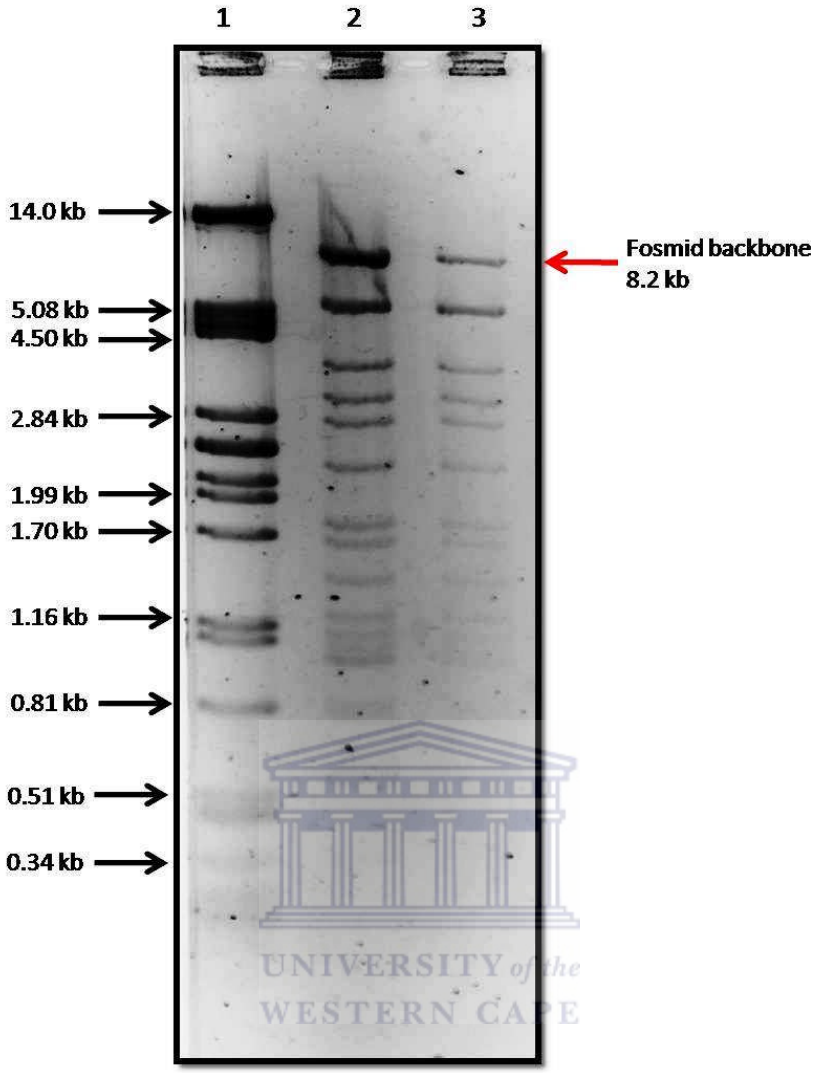


Figure 3.3: Restriction enzyme digestion profiles of clones Try 11 and Try 12, using *Eco*RI and *Hind*III. Lanes 2 and 3: Try 11 and Try 12, respectively. The 8.2 kb fosmid vector backbone is indicated by a red arrow. Lambda DNA digested with the *Pst*I restriction enzyme was used as a molecular weight marker (Lane 1).

The insert size of the halo-forming recombinant fosmid clones was 30.4 kb for both Try 11 and Try 12. The identical restriction profile indicated that these two clones contained exactly the same insert and were duplicate clones. For the purpose of this study, the lipolytic clone Try 11 was chosen for further analysis.

Clone verification by restriction enzyme digestion is a relatively crude method that is dependent on adequate separation of DNA fragments through low percentage agarose gels. Efficient estimates of insert size were obtained by this method and in

the absence of pulse-field apparatus, gel electrophoresis carried out at low voltage for longer time periods proved to be successful for separation of high molecular weight DNA fragments.

3.2 End-Sequencing

End-sequencing of Try 11 showed that the insert cloned into the vector was prokaryotic in origin (Table 3.1). The bacterial classes represented in these sequences are the γ-Proteobacteria (*Burkholderia cenocepacia* and *Tolumonas auensis*). *Burkholderia cenocepacia* are rod-shaped, free-living, motile Gram negative bacteria which are mostly found in soil, water and diseased plants (Miller *et al*., 2002). *Tolumonas auensis* inhabits anoxic sediments of freshwater lakes and has optimum temperature and pH for growth at 22 ºC and pH 7.2, respectively (Fischer-Romero *et al*., 1996). Both organisms are associated with soil and water but neither suggest a thermophilic nature. Because the hot spring is an open environment, it is conceivable that it may contain mesophilic contaminants. In addition, one must bear in mind that since this is only the end-sequence of the insert, it is possible that the other genes cloned into fosmid (Try 11) might be of a thermophilic nature.

Table 3.1: Nucleotide end-sequences of fosmid clone Try 11. The nucleotide identity of the closest match is indicated.

Clone Name	Nucleotide end-sequence	Identity (%)	e-Value	ID of nearest match (Accession number)
Try 11 PCCFOS1 Rev	ATGTCACCTGGATCGGCTCCAACTACTTCGTCAACACCGGCGGCTACTACGACACCTACA GGGCCAGTGCGCCTCGCGATGGTTGGGCCTACGACAGCAACCGCGATGCGGGCCTGACGC AAGTGCCCAGTGGCGCCGGCTATCCGACCTGCCGCCAGTGGTGGAGCGATGGCGGCAATG GCCTGCGTGCGCGGCTACTCGCCGAAGTCGACCCCACCCTGATGAATCGGCTGGCCGGCT GGGCCGGGTTCCTGAGTCGCGCCCAGGTTGATGACTCGGTGATCCGTGCGATCGCCTCGC CGCGCCAGCAGAAGCTCAACCAGGGCGCGGTCTACTCCGACTACGGCGGTCAGATCGACA AGACGCTGCCGAACATCGTCACCCGCGCTGCCGGCGACGTGGGTCTGGCGGTGGGATCGC TGGGCTTCTTCCCGGCCATGGATGTGGTGCGGCAGGCGCTCCCGATGGTGCTGTCACTGC TGAAGATGGCGCTGGTCATCTGCATCCCGCTGGTGCTTGTGGTCGGTACTTACGACCTGA AGACGGTGGTGACGGTCAGCGTGGTGCAGTTCTCGCTGTTCTTTGTGGACTTCTGGTTCC AGCTCGCCCGCTGGATCGACAGCACCATCCTCGACGCACTCTACGGCTGGGGGTGGGGCT GGAACCGGCCGCACTCGAACTTCGATCCGGTGATGGGGCTGAACAATGCCTTCGGCGACA TGTTGCTGAACTTCGTCATGGCGACGATGTTCTTGATTCTTCCTGCTTTCTGGGTTACTG CGCTTGCGTGGACA	651/781 (83 %)	0.00	*Burkholderia cenocepacia* J2315 chromosome 1, complete genome Hypothetical protein (AM747720)
Try 11 T7-promoter	ATCGGATTGTTGCTGGGCCAGGAGTGGAGCCGTGACGAGCATTGAGCCTAAGTCTCAACT TAGGAGCTCGATGATGGCACTCTCTGATCTGATCGTTCGCCAGGCCAAGACCACTGGCAA ACTCTACAACCTCCCCGATCTTGATGGCCTCGGCCTGGTGGTCTCACCGGTCGGCGGCAA GTCGTGGCACTTCCGTTACTACTGGCTCGGCAAGCAAAAGCGCATCTCTCTGGGTAATTA CCCAGAGATCGGCTTGCGCGATGCCCGCATCCTGCGAGACGAAGCCAGAGCCCTGCTGGC CAAGGGCATCAACCCTCACACCGATCGCAAGCAGAAGCGGCACGCTATCCAGCTTGCGGC CGAGCACACCTTCAAGGCGGTCTTCGATGCCTGGGTGGAGCATCGCCGCAAGGAGCTCAA GGAAGGTCGCCAAAGCACGCTCTCGCAGATCCTGCGGATATTTGACCGTGATGTGCTGCC CAGCCTGGGCAAGATGTCGATTTTCGACATCCGCCGCCCCCAACTCTTGGGCGTGCTGGC AAAGATCGAAGAACGAAAGGCCTTCACCACCGCGGAAAAGGTTCGCACCTGGTTCCGGCA GATGTTCCGCTACGCCCTGGTAATCGCCGAGGGGATGGAAGTCAATCCGGCTTCGGATCT TGATGTCGTCGCGGCCCCCAAGCCGCCGACTGCTCACAACCCCTATTTGCGCCTGCATGA ATTGCCCGACATGCTTCTCAAGCTGCGCGAATACCGTCGGCGGGGGCTGCAGACGCAGCT CGGCATCCGCCTGCTGTTCCTCACCGGCGTACGCACT	674/833 (80 %)	6e-175	*Tolumonas auensis* DSM 9187, complete genome Integrase family protein (CP001616)

3.3 Transposon Mutagenesis

In order to identify the gene conferring lipolytic activity, transposon mutagenesis was used. This is a powerful tool for creating knock-out mutants. The fosmid of lipolytic clone Try 11 was extracted and randomly mutated using the GPS® – Mutagenesis system. Unfortunately, knock-out mutants were not easily distinguished from non-mutated fosmids. Therefore, the fosmid was sequenced by pyrosequencing instead.

3.4 Sequence Analysis

Several bioinformatic tools were used to predict basic structural information based on raw sequence data. Using more than one bioinformatic tool is advantageous as it allows more accurate predictions and comparisons. The complete sequence of Try 11 fosmid insert was assembled from 6 contig sequences using CLC Genomics WorkBench 3. Alignment of the end sequences to the assembled contigs confirmed that the cloned insert was completely assembled (Table 3.1). The full sequence (34929 bp) was obtained, containing an average mol % GC content of 55 % and corresponded well to the size determined by restriction enzyme analysis (Figure 3.3). The ORFs within the insert were predicted using SoftBerry. A total of 40 ORFs, designated RHgene1 to RHgene40, were predicted on the Try 11 fosmid insert (Table 3.2, Figure 3.2). The majority of these ORFs displayed high sequence similarity (>90 %) to sequences deposited in GenBank. A previously uncharacterised α/β hydrolase fold protein (RHgene34) was identified. RHgene34 had the highest sequence similarity (41 %) to proteins of α/β hydrolase fold-3 of *Burkholderia* sp. H160 (Table 3.2), a thermophilic organism. Considering that the library was constructed from a Malawian hot spring soil sample (temperature 72-78°C), the occurrence of thermophilic species is expected and the similarity to proteins of their thermophilic homologues is not surprising. However, 38 % of proteins from the fosmid clone had high similarity to proteins from *Enterobacter,* a mesophilic organism. This may be due to the fact that Malawian hot spring is an open environment and remnant DNA may have been sampled, including DNA from dead mesophiles, or that currently, that is it the closest DNA match in the database.

Table 3.2: Predicted ORFs in fosmid Try 11.

Genes RHgene	O* 5' – 3'	Left end	Right end	Gene length (bp)	AA*	Closest protein match from GenBank (Identities)	Accession number	Organism
1	+	88	2019	883	643	Integrase family protein (499/643, 78 %)	YP_001414722.1	*Parvibaculum lavamentivorans* DS-1
2	-	2125	4359	984	744	Hypothetical protein Ajs_1568 (252/728, 35 %)	YP_985837.1	*Acidovorax* sp. JS42
3	-	4636	5082	148	148	Hypothetical protein (23/69, 33 %)	XP_001368753.1	*Monodelphis domestica*
4	-	5183	5695	142	170	Hypothetical protein Acid345_3844 (54/176, 31 %)	YP_592918.1	*Candidatus Koribacter versatilis* Ellin345
5	-	6194	9349	1573	1051	Peptidase families S8 and S53 domain protein (123/448, 27 %)	YP_003572398.1	*Salinibacter ruber* M8
6	-	9608	9700	78	30	Hypothetical protein AURANDRAFT_60578 (15/28 (54 %)	EGB12620.1	*Aureococcus anophagefferens*
7	+	9845	10021	102	58	Unknown		
8	-	9939	10166	121	75	Hypothetical protein mlr4028 (19/44, 43 %)	NP_104997.1	Mesorhizobium loti MAFF303099
9	-	10215	10442	110	75	Siderophore-interacting protein (25/64, 39 %)	ZP_07991351.1	*Corynebacterium variabile* DSM

								44702
10	-	10589	11374	438	261	Short-chain dehydrogenase/reductase SDR (197/260, 76 %)	YP_003258958.1	*Pectobacterium wasabiae* WPP163
11	-	11471	12184	182	237	LysR family transcriptional regulator (125/231, 54 %)	YP_584546.1	*Cupriavidus metallidurans* CH34
12	-	12611	14071	590	486	Putative outer membrane efflux lipoprotein (382/498, 77 %)	ZP_03825891.1	*Pectobacterium carotovorum* subsp. *brasiliensis* PBR1692
13	+	14160	14774	236	204	Regulatory protein, TetR (156/204, 76 %)	EGH45757.1	*Pseudomonas syringae pv. pisi str*. 1704B
14	+	15067	15963	499	298	Secretion protein HlyD (234/298, 79 %)	YP_348215.1	*Pseudomonas fluorescens* Pf0-1
15	+	15960	17066	527	368	Hypothetical protein SOD_c03250 (319/357, 89 %)	ZP_06190974.1	*Serratia odorifera* 4Rx13
16	-	17026	17463	351	145	PilT domain-containing protein (25/77, 32 %)	YP_920790.1	*Thermofilum pendens* Hrk 5
17	-	17648	17926	194	92	Hypothetical protein Bphyt_5853 (45/89, 51 %)	YP_001889560.1	*Burkholderia phytofirmans* PsJN
18	+	18089	18721	263	210	Transcriptional regulator protein (89/206, 43 %)	YP_002541214.1	*Agrobacterium radiobacter* K84
19	+	18721	18918	99	65	Unknown		
20	-	18851	19963	296	370	Oxidoreductase (154/366, 42 %)	YP_003766767.1	*Amycolatopsis mediterranei*

21	-	19967	20542	213	191	Acyl carrier protein phosphodiesterase (91/183, 50 %)	YP_001603996.1	*Gluconacetobacter diazotrophicus* PAl 5
22	+	20852	21091	121	79	Putative transcriptional regulator (53/79, 67 %)	YP_559233.1	*Burkholderia xenovorans* LB400
23	+	21236	21439	92	67	Unknown		
24	+	21445	21669	128	74	Unknown		
25	+	21700	22098	211	132	Transposase IS3/IS911 family protein (58/121, 48 %)	YP_001898744.1	*Ralstonia pickettii* 12J
26	+	22134	22421	169	95	Transposase (66/97, 68 %)	ZP_06686116.1	*Achromobacter piechaudii* ATCC 43553
27	+	22490	24070	993	526	Transposase IS66 (339/521 (65 %)	YP_002947819.1	*Variovorax paradoxus* S110
28	-	24087	25007	392	306	Transporter membrane protein (241/297, 81 %)	YP_004353562.1	*Pseudomonas brassicacearum* subsp.
29	+	25134	25448	100	104	Hypothetical protein MldDRAFT_3100 (21/59, 36 %)	ZP_01291063.1	*delta proteobacterium* MLMS-1
30	+	25445	26008	114	187	Transcriptional regulator, TetR family (80/181, 44 %)	ZP_07031205.1	*Acidobacterium* sp. MP5ACTX8
31	+	26141	26941	472	266	Short-chain dehydrogenase (169/263, 64 %)	YP_841302.1	*Ralstonia eutropha* H16

32	+	27021	27956	477	311	Short chain dehydrogenase (124/250, 50 %)	ZP_02355639.1	*Burkholderia oklahomensis* EO147
33	-	27913	28017	82	34	Unknown		
34	+	28011	28964	951	317	Alpha/beta hydrolase fold-3 domain protein (128/312, 41 %)	ZP_03268234.1	*Burkholderia* sp. H160
35	+	29102	30241	330	379	Major facilitator superfamily MFS_1 (206/372, 55 %)	ZP_03544743.1	*Comamonas testosteroni* KF-1
36	-	30338	32197	619	619	Relaxase (464/625, 74 %)	ZP_06845718.1	*Burkholderia* sp. Ch1-1
37	-	32394	32594	115	66	Unknown		
38	-	32646	33404	358	252	Hypothetical protein XCC3118 (192/252, 76 %)	NP_638465.1	*Xanthomonas campestris pv. campestris* str. ATCC 33913
39	+	33634	34014	142	126	Hypothetical protein Tmz1t_0940 (66/129, 51 %)	YP_002354602.1	*Thauera* sp. MZ1T
40	-	34026	34928	655	300	Hypothetical protein XCC3117 (242/300, 81 %)	NP_638464.1	*Xanthomonas campestris pv. campestris*

O* - Orientation AA* - Amino acid

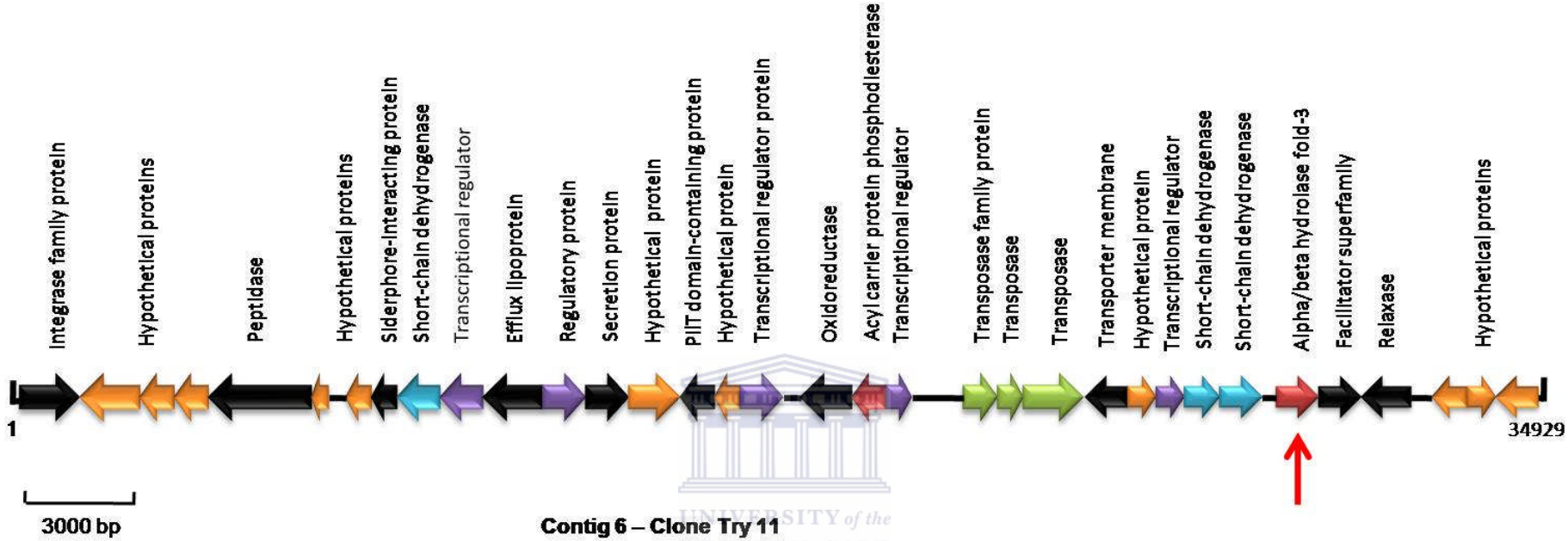


Figure 3.4: Arrangement of open reading frames identified in insert of Try 11. Arrows indicate the location and orientation of predicted open reading frames. Details of putative protein function and GenBank accession numbers are given in Table 4.2.

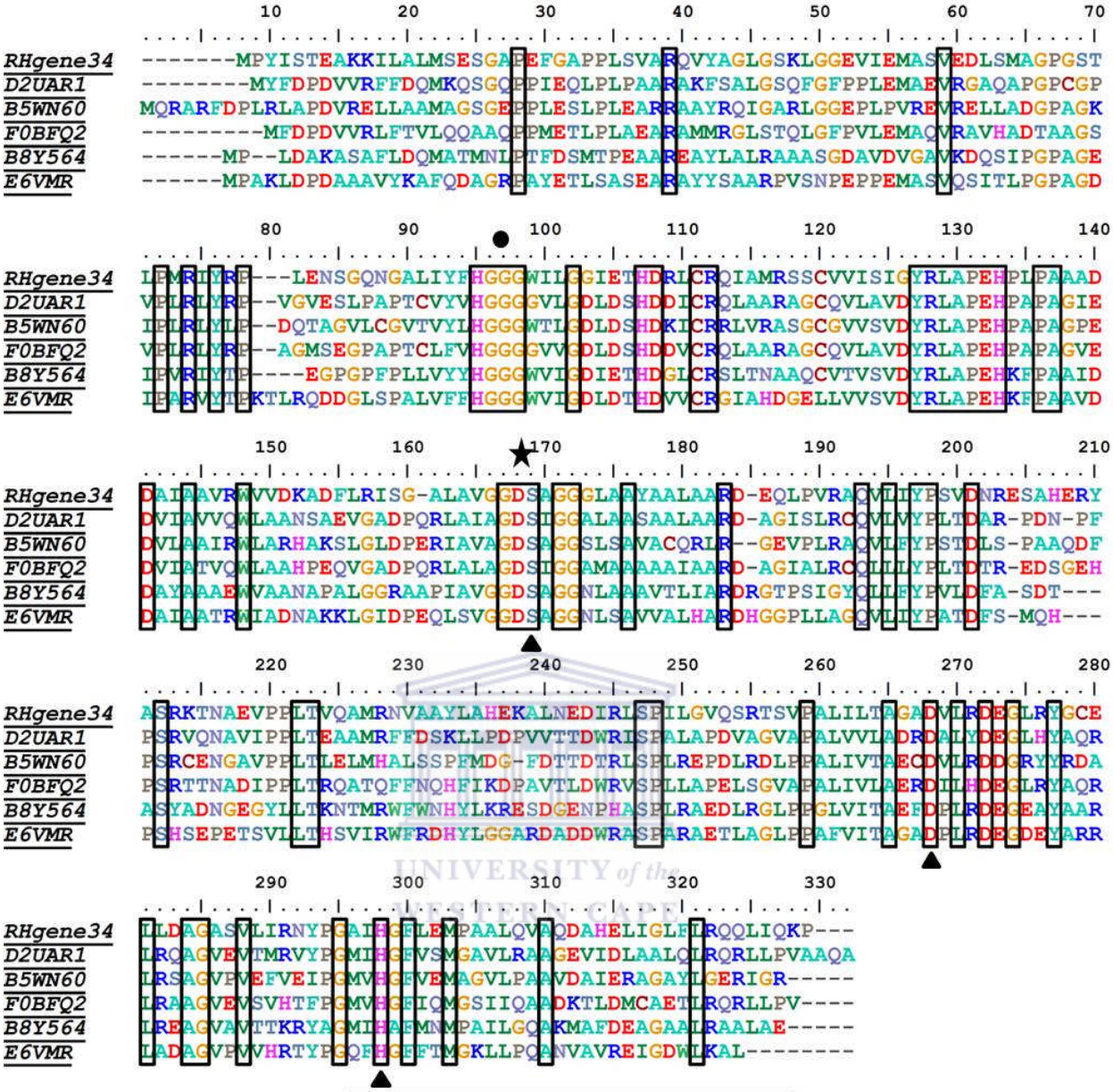
The identified lipolytic ORF RHgene34 encoded a protein of 317 amino acids and putative promoter regions (-10 and -35) and ribosomal binding site (RBS) are indicated (Figure 3.5).

```
            GCAGTGGACTTTAAACAATTTGATCCCCAGACGGCTATTGGTCCAATGATCGAGATCATAGAGAGCAAGGGGG
            GCAACTTTAGAAATGTGGTTCCGGAGTCCTTTGTCGCGGCTACCAAAAGCCTGCAAGCAAAAGTTTGGGACCT
            GAAGGTGTAGAACCTATTACTGCGCAGAAATCGAAATAACTGTCGTGAGAACTGTGAGGTAGT    -35
                                 -10                          RBS
1           ATGCCATATATTTCCACCGAAGCAAAAAAAAATACTCGCTCTCATGAGTGAAAGCGGCGCG
1            M  P  Y  I  S  T  E  A  K  K  I  L  A  L  M  S  E  S  G  A

61          CCAGAATTCGGAGCTCCGCCGCTATCGGTCGCGCGCCAGGTCTATGCAGGCCTTGGCTCA
21           P  E  F  G  A  P  P  L  S  V  A  R  Q  V  Y  A  G  L  G  S

121         AAGTTGGGCGGGGAGGTTATCGAGATGGCCTCTGTTGAGGATCTTTCGATGGCTGGGCCG
41           K  L  G  G  E  V  I  E  M  A  S  V  E  D  L  S  M  A  G  P

181         GGTAGTACTTTGCCGATGCGGATCTATCGCCCCCTTGAAAACTCTGGTCAAAATGGAGCT
61           G  S  T  L  P  M  R  I  Y  R  P  L  E  N  S  G  Q  N  G  A

241         TTAATTTATTTTCATGGTGGCGGGTGGATTCTGGGGGGCATAGAGACTCATGACAGACTC
81           L  I  Y  F  H  G  G  G  W  I  L  G  G  I  E  T  H  D  R  L

301         TGTCGACAAATCGCAATGAGATCCTCTTGCGTAGTTATCTCTATCGGTTACAGGCTTGCG
101          C  R  Q  I  A  M  R  S  S  C  V  V  I  S  I  G  Y  R  L  A

361         CCTGAGCACCCGCTGCCAGCGGCTGCAGACGATGCGATCGCTGCGGTGCGTTGGGTTGTG
121          P  E  H  P  L  P  A  A  A  D  D  A  I  A  A  V  R  W  V  V

421         GACAAAGCTGATTTCCTGCGAATCTCGGGCGCCCTAGCCGTTGGCGGTGACAGTGCAGGT
141          D  K  A  D  F  L  R  I  S  G  A  L  A  V  G  G  D  S  A  G

481         GGAGGGTTGGCTGCGTATGCTGCGCTCGCTGCTCGTGACGAACAATTGCCCGTCCGCGCT
```

```
161        G  G  L  A  A  Y  A  A  L  A  A  R  D  E  Q  L  P  V  R  A

541       CAGGTCCTAATCTATCCTAGCGTTGATAACCGGGAATCGGCACACGAGCGATATGCATCA
181        Q  V  L  I  Y  P  S  V  D  N  R  E  S  A  H  E  R  Y  A  S

601       CGAAAAACGAATGCAGAAGTGCCTCCTCTTACGGTGCAGGCAATGCGCAACGTTGCCGCA
201        R  K  T  N  A  E  V  P  P  L  T  V  Q  A  M  R  N  V  A  A

661       TATCTAGCGCATGAGAAAGCATTAAACGAAGATATTCGCCTGTCGCCTATTCTTGGTGTG
221        Y  L  A  H  E  K  A  L  N  E  D  I  R  L  S  P  I  L  G  V

721       CAAAGCCGGACATCGGTGCCTGCCTTAATACTGACCGCGGGAGCTGACGTGTTGAGGGAT
241        Q  S  R  T  S  V  P  A  L  I  L  T  A  G  A  D  V  L  R  D

781       GAAGGATTGCGCTATGGTTGCGAATTGCTCGATGCGGGTGCCTCGGTTTTAATTAGAAAC
261        E  G  L  R  Y  G  C  E  L  L  D  A  G  A  S  V  L  I  R  N

841       TATCCTGGTGCAATTCATGGTTTCCTCGAGATGCCAGCTGCCCTGCAAGTCGCACAAGAT
281        Y  P  G  A  I  H  G  F  L  E  M  P  A  A  L  Q  V  A  Q  D

901       GCGCACGAATTGATTGGGCTTTTTCTTCGGCAGCAATTAATTCAAAAGCCATAG
301        A  H  E  L  I  G  L  F  L  R  Q  Q  L  I  Q  K  P  *
```

Figure 3.5: The nucleotide sequence of the RHgene34 lipolytic gene from the metagenomic library. The predicted amino acid sequence of RHgene34 is given above the nucleotide sequence in the standard one-letter code. The putative promoter regions (-35 and -10 regions) and the ribosomal bind site (RBS) is highlighted in boldface and underlined. RHgene34 is 951 nucleotides in length with an ATG start codon. The asterisk denotes the stop codon (TAG) and the protein encoded a polypeptide with a molecular mass of 34 kDa.

The 951 bp sequence encoded a 317 amino acid polypeptide on the positive strand of the sequence. Multiple sequence alignment of RHgene34 protein to hits generated from UNIPROT BLAST identified the conserved motifs, including the putative pentapeptide active site GxSxG (Figure 3.6). Thus, the catalytic activity of RHgene34 likely involves the catalytic triad consisting of the catalytic nucleophile serine (Ser157), in the centre of GxSxG active site, the aspartate (Asp255) and the highly conserved histidine (His285). In addition, RHgene34 contains a strictly conserved family IV characteristic HGGG motif, which is located 95 amino acids upstream of the active site. It has been reported by Laurell *et al.* (2000), that hormone-sensitive lipase (HSL) family IV lipolytic enzymes have the HGGG motiff which is usually located 70 to 100 amino acids from the active site and is involved in hydrogen bonding interactions for stabillisation which may contribute to the formation of the oxyanion hole that is likely to act in the hydroysis process.



Accession number	% identity
D2UAR1	42%
B5WN60	41%
F0BFQ2	39%
B8Y564	39%
E6VMR1	40%

Figure 3.6: Multiple sequence alignment of the RHgene34 protein sequence with close relatives indentified from UNIPROT BLAST. Boxes indicate sequence similarity with a threshold value of 90 %. The putative catalytic triad residues composed of Ser157 (S), Asp255 (D) and His285 (H) are indicated with shaded triangles. The family IV most conserved motif, HGGG, is indicated by a shaded circle. The GxSxG lipolytic motif

(PROSITE Accession No. PS00120) is indicated with a shaded star. Accession numbers denote the following; D2UAR1: Putative esterase/lipase/thioesterase family protein [*Xanthomonas albilineans*], B5WN60: α/β Hydrolase fold-3 [*Burkholderia* sp. H160], F0BFQ2: Esterase/lipase [*Xanthomonas vesicatoria*], B8Y564: Lipolytic enzyme [Uncultured bacterium], E6VMR1: Putative lipase/esterase [*Rhodopseudomonas palustris* DX-1]. Percent identity of RHgene34 to the protein sequences is indicated in the Table. The sequences were aligned using ClustalW (Thompson *et al*., 1994).

The SignalP server was used to identify whether RHgene34 had an N-terminal signal sequence in the predicted gene. From the prediction, the RHgene34 gene product is a non-secretory protein (Figure 3.7) with a signal peptide probability of 0.001. This indicated the absence of an N-terminal leader peptide.

The rare codons predicted by the Rare Codon Caltor are based on the codon usage observed in the *E. coli* genome. Codon use differs among microorganisms and even among genes within a single genome (Grocock *et al*., 2002). The Rare Codon Caltor predicted a total of 35 rare codons (11 %) in the sequence, with the majority (8 %) being arginine (CGA, CGG, AGG and AGA) and glycine (GGA and GGG) (Table 3.3). Further analysis would be required in order to assess and compare the codon bias observed in the lipolytic gene RHgene34 to other microorganisms. The high frequency of rare codons in RHgene34 bears relevance to the expression work which will be discussed in chapter 4 section1.

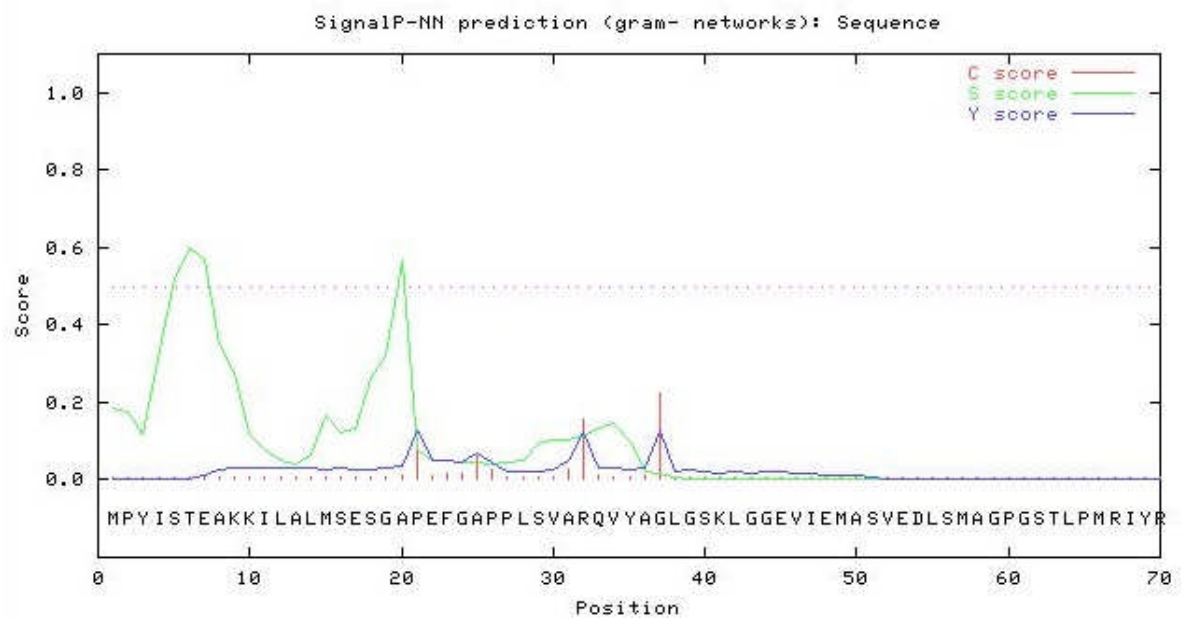


Figure 3.7: Prediction of N-terminal signal peptide cleavage site in polypeptide RHgene34.

Table 3.3: Rare codons and their frequency in the nucleotide sequence obtained for RHgene34 predicted by rare codon calculator for expression in *E. coli*.

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	4
	CGG	4
	AGG	2
	AGA	3
Glycine	GGA	5
	GGG	6
Isoleucine	AUA	3
Leucine	CUA	4
Proline	CCC	2
Threonine	ACG	2

3.5 Phylogenetic Analysis

To further clarify the phylogenetic relationship of RHgene34 with other lipolytic enzymes, a neighbour-joining phylogenetic tree was constructed using the amino acid sequence of 31 lipolytic enzymes, representing 9 different families (Figure 3.8) (Lee *et al*., (2006) and Arpigny and Jaeger (1999)). Phylogenetic analysis suggests that RHgene34 forms part of a distinct group of family IV hormone-sensitive lipases. This protein is closely related to a thermophilic carboxylesterase of the strain *Archaeoglobus fulgibus* (accession number AAB89533), carboxylesterase (accession number AY726780) from an uncultured archeon, esterase (accession number EU195805) from uncultured bacterium, and a putative esterase/lipase (accession number AM050333) from an unidentified microorganism, all of which cluster in family IV lipases. These results suggest that the RHgene34 is a new member of the family IV lipases.

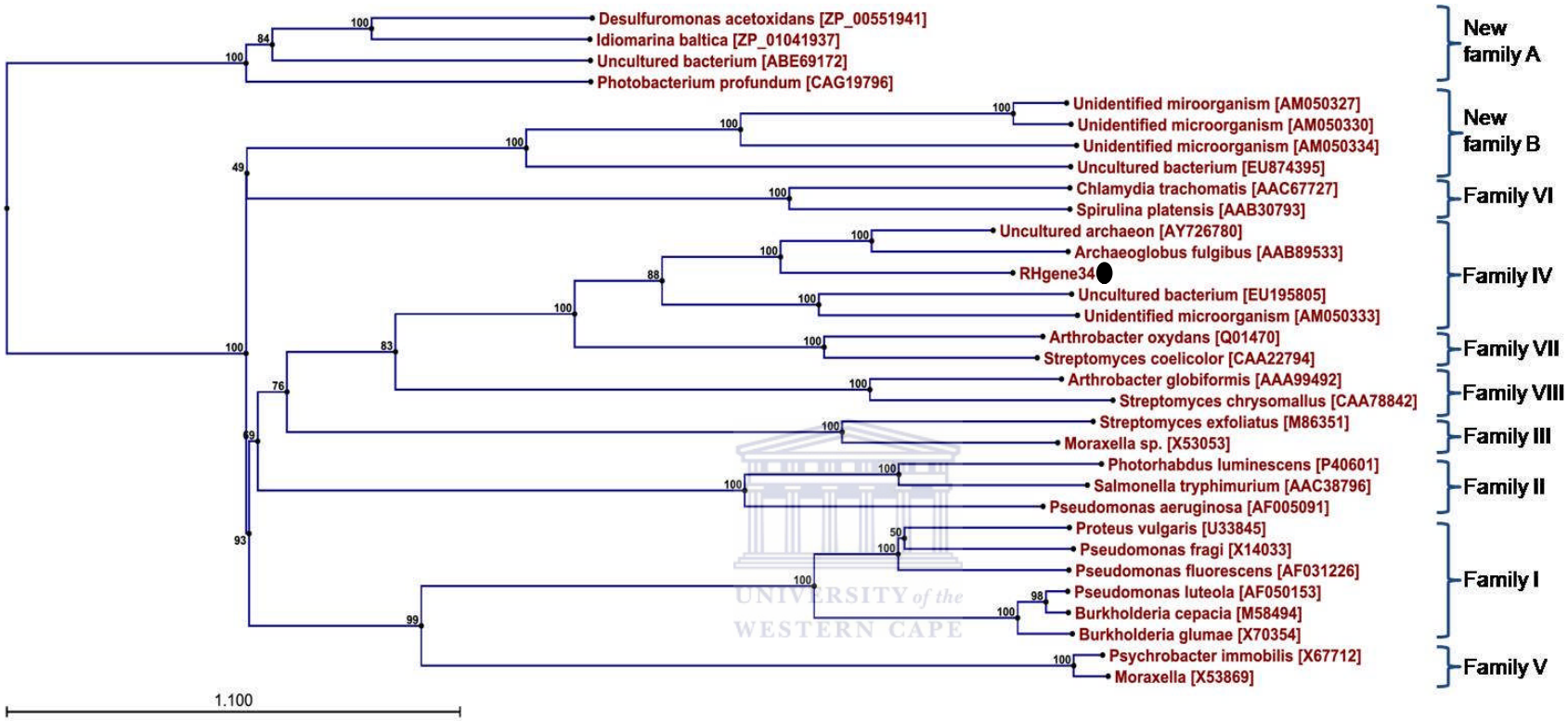
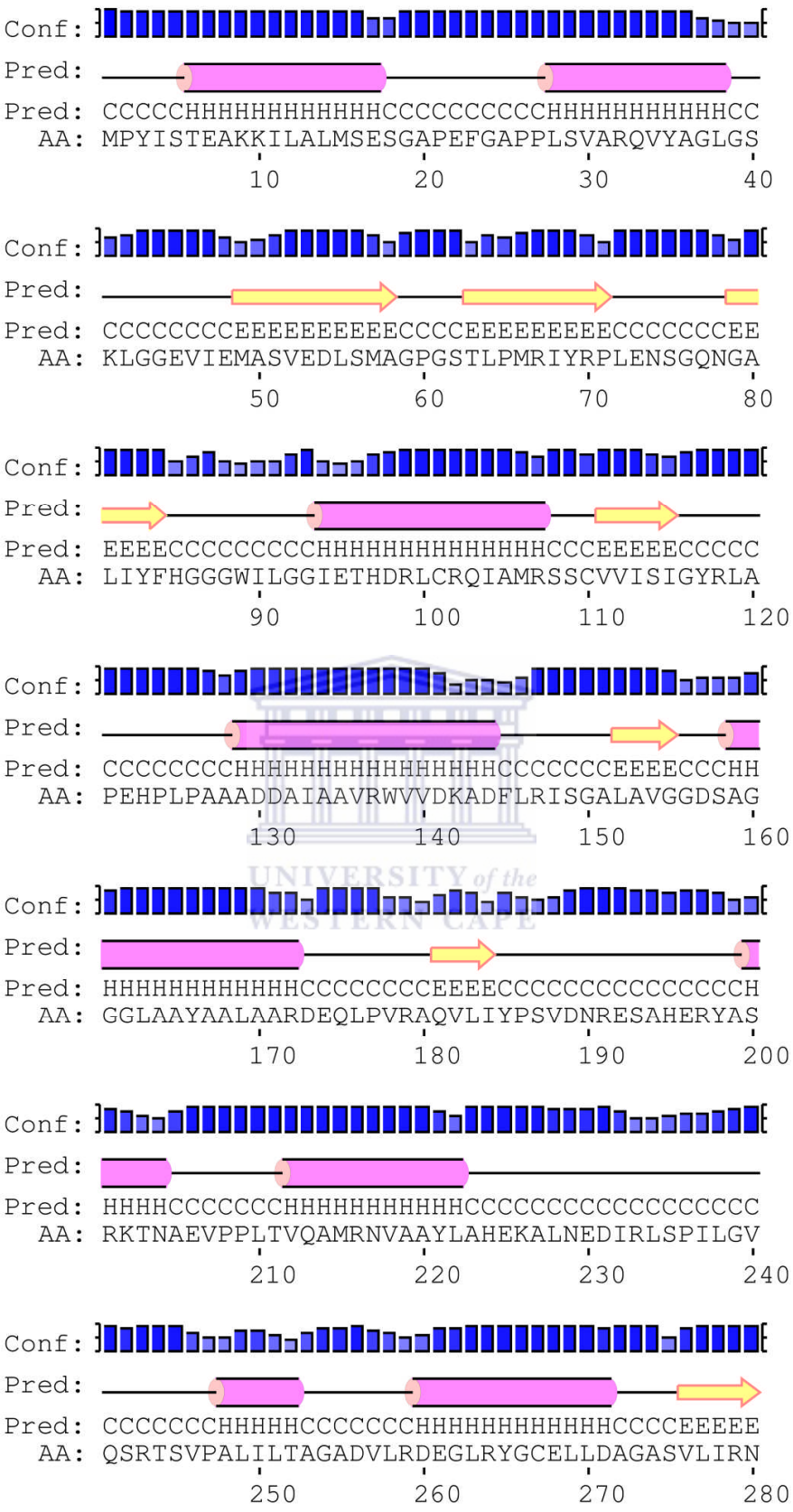


Figure 3.8: Phylogenetic analysis of RHgene34 and 31 selected lipolytic enzymes, representing 9 different families. Lipolytic families were defined by Lee *et al*., (2006) and Arpigny and Jaeger (1999). The phylogenetic tree was constructed using CLC Genomics Workbench 3.0 software with the neighbour-joining method. The numbers at nodes indicate the bootstrap percentage of 1000 resamples. RHgene34, denoted by shaded circle, is shown to belong to family IV lipolytic enzymes.

3.6 Homology Modelling

The ultimate goal of protein modelling is to predict a structure from its sequence with accuracy that is comparable to the best results achieved experimentally. Three dimensional protein structures give some information on molecular organization and function within a protein. When there is no experimentally-determined crystal structure, homology modelling can provide a rational opportunity to obtain a good 3D model to make some predictions on the enzyme's activity (Messaoudi *et al*., 2011; Krieger *et al*., 2003). Homology modelling involves taking a protein sequence with a known structure and mapping it against an unknown structure, in this case RHgene34. It is expected that the proteins would have similar structures as they are of similar origin and function. So, the known structure is used as a template to model the structure of the unknown (Krieger *et al*., 2003; Hilbert *et al*., 1993).

PSIPRED is used to predict secondary structure based on the neuronal networks of protein. Figure 3.9 shows the predicted secondary structure of RHgene34. It was found that the core of the enzyme had a definite order of 8 β-sheets connected by α-helices, which relates well to most lipolytic enzymes in literature (Carr and Ollis, 2009; Ollis *et al*., 1992). This structure corresponded with the secondary structure topology predicted by PDBsum, which is folded into a characteristic α/β hydrolase fold (Figure 3.10). This fold consists of β-sheets, with the second β-sheet antiparallel to the others, and strands β3 and β8 connected to helices packed on either side of the central parallel β-sheet (Jaeger *et al*., 1999).


Conf:
Pred:
Pred: CCCCCHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHCC
AA: MPYISTEAKKILALMSESGAPEFGAPPLSVARQVYAGLGS
10 20 30 40
Conf:
Pred:
Pred: CCCCCCCCEEEEEEEEEECCCCEEEEEEEECCCCCCCEE
AA: KLGGEVIEMASVEDLSMAGPGSTLPMRIYRPLENSGQNGA
50 60 70 80
Conf:
Pred:
Pred: EEEECCCCCCCCCHHHHHHHHHHHHHHCCCEEEEECCCCC
AA: LIYFHGGGWILGGIETHDRLCRQIAMRSSCVVISIGYRLA
90 100 110 120
Conf:
Pred:
Pred: CCCCCCCCHHHHHHHHHHHHHHHHHCCCCCCCEEEECCCHH
AA: PEHPLPAAADDAIAAVRWVVDKADFLRISGALAVGGDSAG
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHHHHHHCCCCCCCCCEEEECCCCCCCCCCCCCCCH
AA: GGLAAYAALAARDEQLPVRAQVLIYPSVDNRESAHERYAS
170 180 190 200
Conf:
Pred:
Pred: HHHHCCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCC
AA: RKTNAEVPPLTVQAMRNVAAYLAHEKALNEDIRLSPILGV
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCHHHHHCCCCCCCHHHHHHHHHHHCCCCEEEEE
AA: QSRTSVPALILTAGADVLRDEGLRYGCELLDAGASVLIRN
250 260 270 280

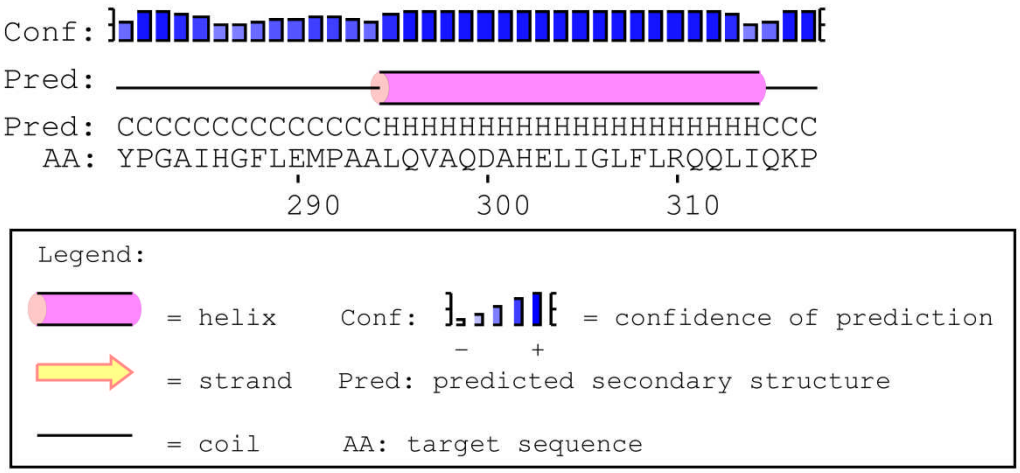


Figure 3.9: Secondary structure for the amino acid sequence obtained for RHgene34. Figure generated by PSIPRED (McGuffin *et al*., 2000; Jones, 1999).

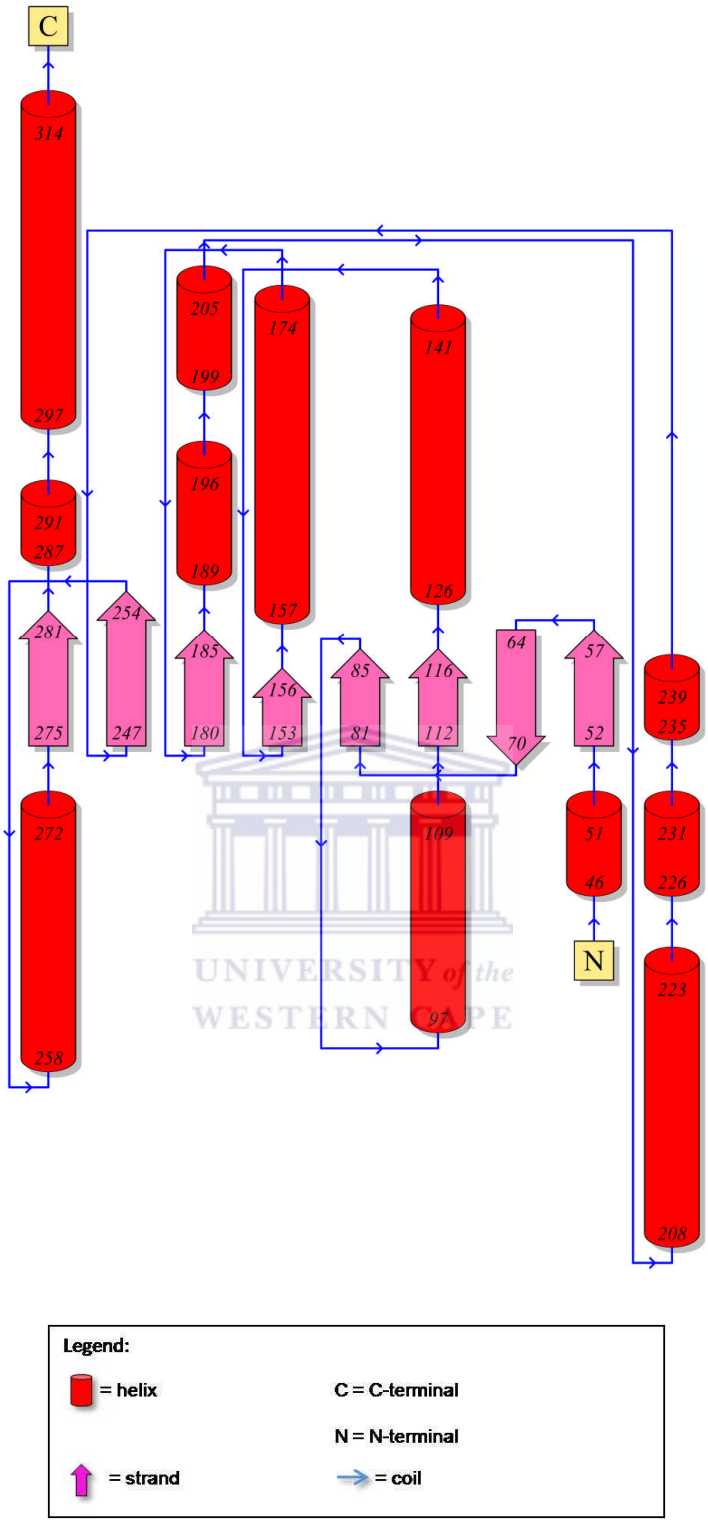
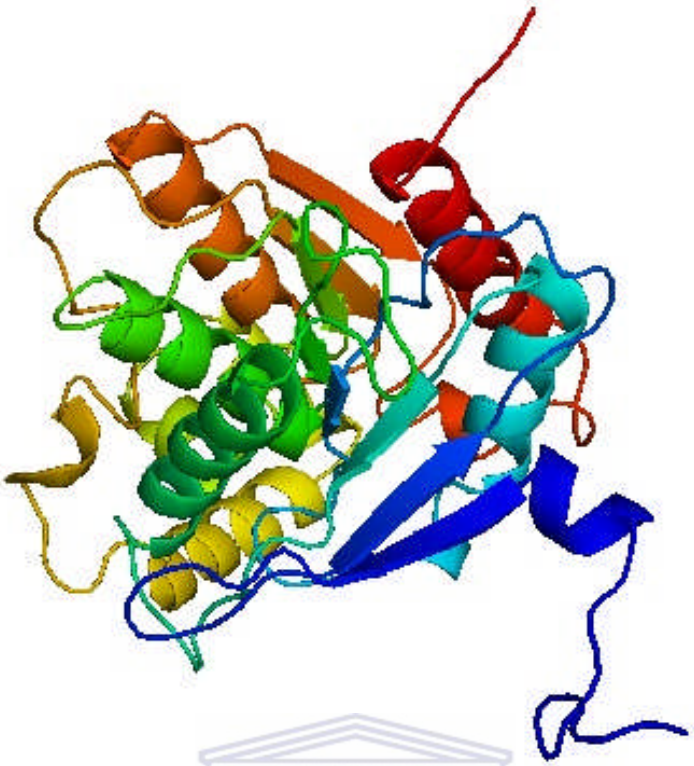
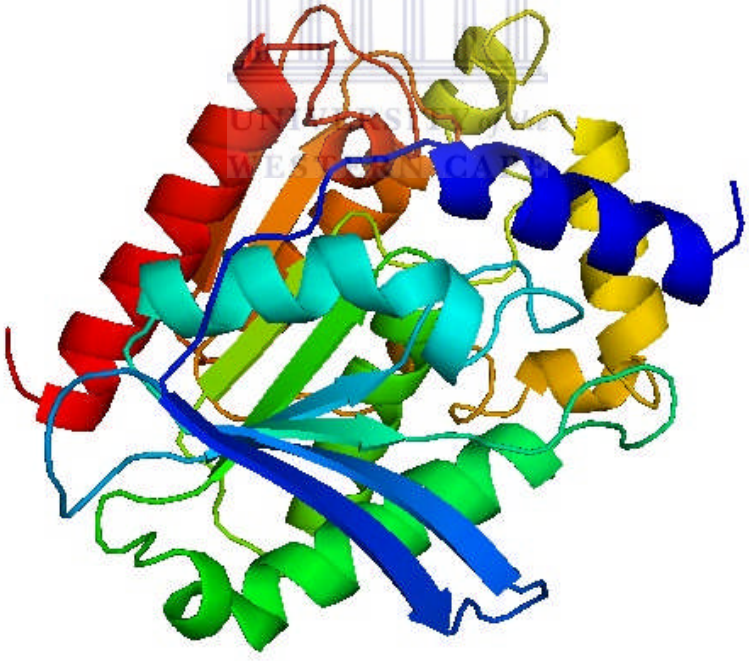


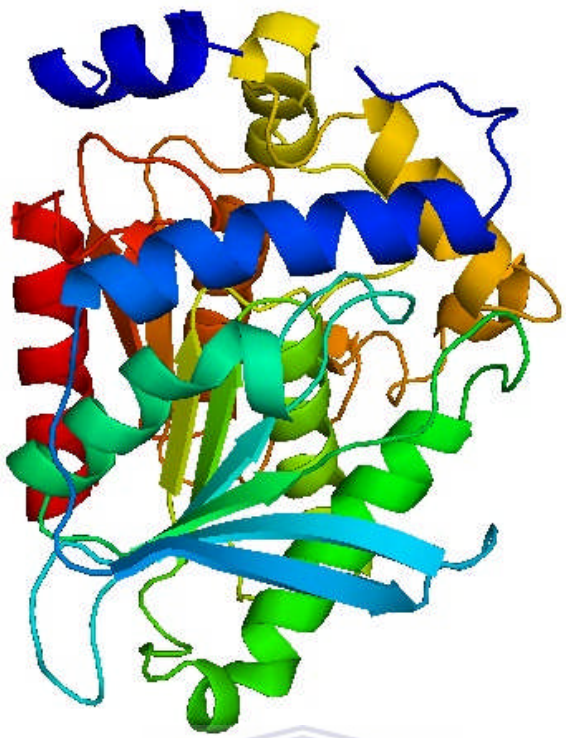
Figure 3.10: Secondary structure topology for the amino acid sequence obtained for RHgene34 using PDBsum (Morris *et al*., 1992). The elements of the secondary structure shown correspond to the canonical α/β hydrolase fold.



A



B



C

Figure 3.11: Homology model of the (A) RHgene34 protein built by the Swiss Model server using PDB structures 3ainD; residues 27 to 315 (B) and 2wirA; residues 3 to 313 (C) used as templates. The polypeptide spans across the colour spectrum from blue (N-terminal) to red (C-terminal).

The RHgene34 model (Figure 3.11) does not indicate the presence of a lid structure. Lid structures are usually found at the C-terminal of the polypeptide chain and consist of a single or double helix structure which covers the active site. However, RHgene34 does have an N-terminal loop structure which might be involved in catalysis or used for dimer formations. It is known that esterases (and a few lipases) do not display a lid structure, such as the lipase from *Candida antarctica* B (Uppenberg *et al*., 1994). RHgene34 could be one of those lipases without a lid structure.

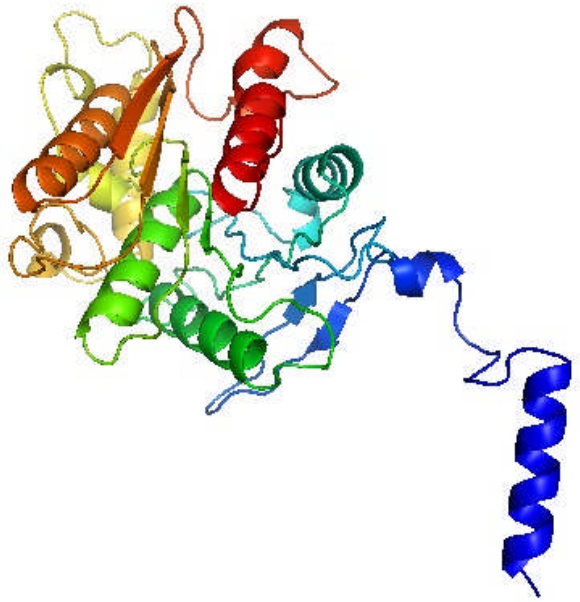


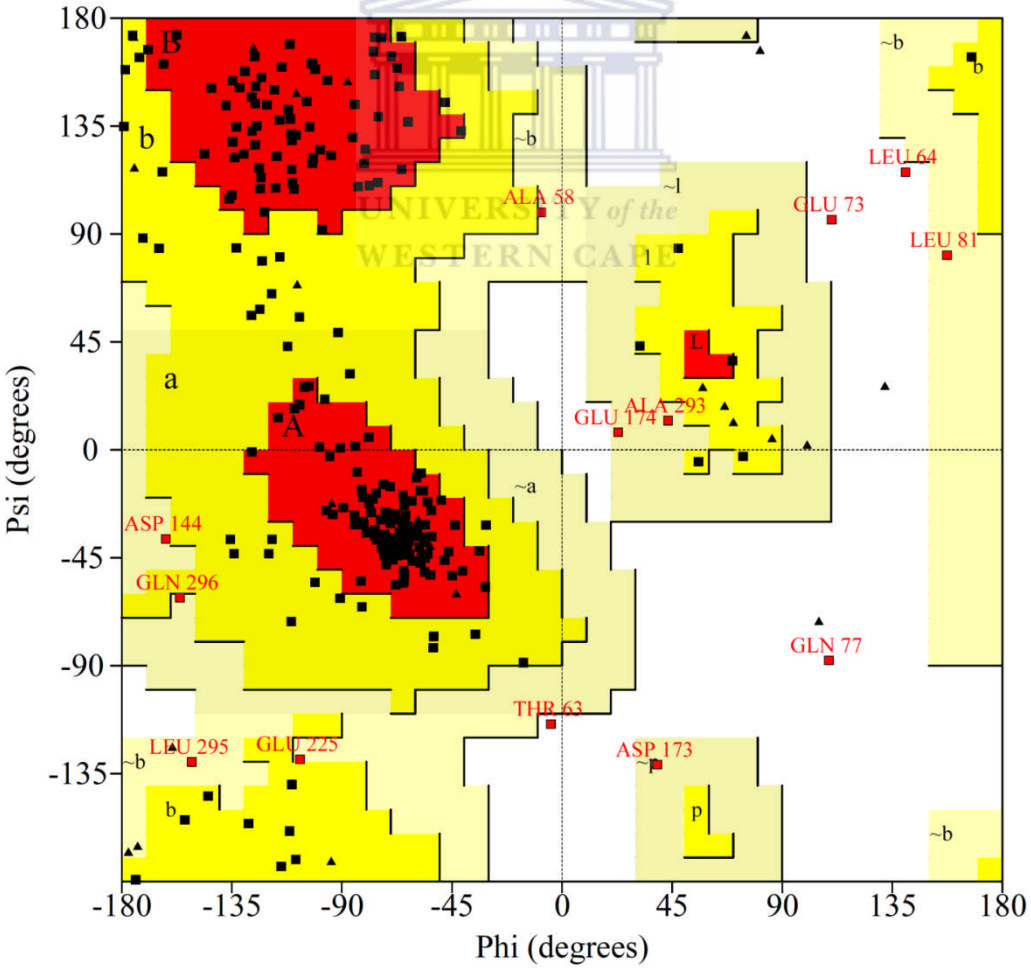
Figure 3.12: RHgene34 homology model built by the 3D JIGSAW model server. The polypeptide spans across the colour spectrum from blue (N-terminal) to red (C-terminal).

The 3D JIGSAW model server is not a reliable model builder as it builds a model according to physiochemical properties of the amino acids, without taking into account the already available models of close identity (Figure 3.12). However, it is still comparable to the Swiss Model because it takes into account other proteins of close identity and the characteristic α/β hydrolase fold is clearly visible. With the 3D JIGSAW the α/β hydrolase fold structure is partly constructed.

The growing knowledge of protein 3D-structures has prompted an attempt to classify proteins according to their fold. Hydrolases are found in the α/β fold group, also called the α/β hydrolase fold. All lipases, as well as the majority of the esterases, share this fold (Fojan, *et al*., 2000). Esterases and lipases generally display broad substrate specificity. Thus they cannot be classified solely by their function (Fojan, *et al*., 2000).

Most bioinformatic models built by homology modelling will have errors, for example side chains may be misplaced or even whole loops (Hooft *et al*., 1996); for this reason it is important to verify the quality of a model by using programs such as RAMPAGE. This program was used to visualize and assess the Ramachandran plots of the predicted protein structures of RHgene34. Ramachandran plots display

the *psi* and *phi* conformation angles for each residue in a protein in order to validate the predicted structure (Lovell *et al*., 2003). For a model to be accurate, 98 % of residues are expected in the favoured region, which are the areas in the plot that show the preferred regions of *psi*/*phi* angle pairs for residues in a protein. The allowed region should have 2 % of residues and a few will be in the disallowed region (Kleywegt and Jones, 1996). The model built by the Swiss model server contained 76.1 % residues in the favoured region, 18.5 % residues in the allowed region and 1.7 % residues in the disallowed region (Figure 3.13). This indicates that the Swiss model was not reliable as a high number of pairs were not in the favoured regions of the plot. This may be due to the fact that RHgene34 protein may be novel and was modelled against 3ainD and 2wirA, which are templates of low identity (35.9 % and 33.7 %, respectively).



Legend:

A - Core alpha
a - Allowed alpha
~a - Generous alpha
B - Core beta
b - Allowed beta
~b - Generous beta
L - Core left-handed alpha
l - Allowed left-handed alpha
~l - Generous left-handed alpha
p - Allowed epsilon
~p - Generous epsilon

Figure 3.13: Ramachandran plot for model of RHgene34 built by the Swiss model server, showing number of residues in favoured, allowed and disallowed region.

CHAPTER 4

Enzyme Characterisation

Results and Discussion

4.1 Cloning of the Lipolytic Gene RHgene34

A Malawian hot spring sediment metagenomic library was constructed where Try 11, a high molecular weight fosmid, was identified having lipolytic activity. RHgene34 was isolated from Try 11 and analysed using various bioinformatic tools. DNAMAN was used to predict restriction enzyme recognition sites occurring in the RHgene34 DNA sequence. It was established that *Nde*I and *Hind*III did not cut within the gene sequence and a primer pair [RhuF2 and LaniR2] (Table 2.2) was designed with these sites introduced at the ends of each primer. To obtain the full length gene (951 bp), gradient PCR was performed. Successful cloning was confirmed in three ways: (i) by restriction digestion using *Nde*I and *Hind*III (Figure 4.1) and (ii) by sequencing using the T7 promoter and terminator primers, and (iii) by PCR amplification (Figure 4.3). To investigate the property of RHgene34 protein, RHgene34 was expressed as an C-terminal His-tag fusion protein using pET-21a(+) expression system in *Escherichia coli* BL21 (DE3). As discussed in section 3.4, this was unsuccessful and was attributed to the utilisation of high percentage of rare codons. Instead, *E. coli* Rosetta was used which have high rare codon usage when compared to BL21 strain. Additionally, the clone designated RH-pET conferred a lipolytic phenotype on the host when transformed.

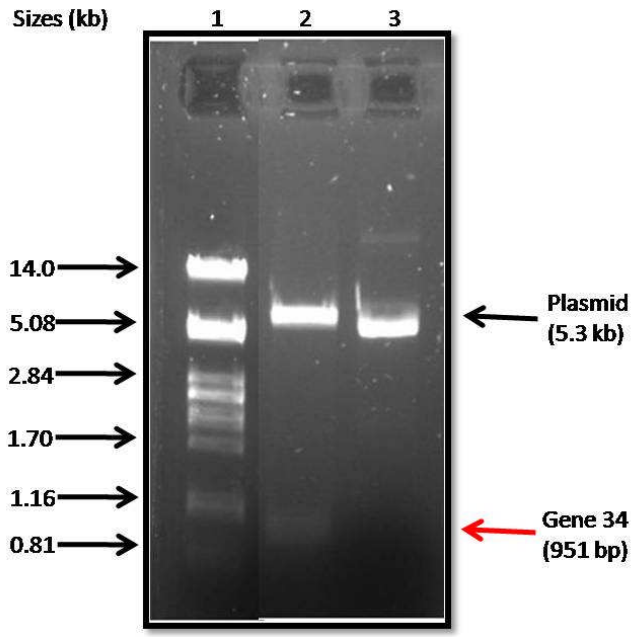


Figure 4.1: Restriction enzyme digestion of clone RH-pET. Lane 1: DNA molecular marker, lambda-*Pst*I digested DNA. Lane 2: *Nde*I and *Hind*III digested recombinant RH-pET plasmid construct. Lane 3: Uncut recombinant RH-pET plasmid.

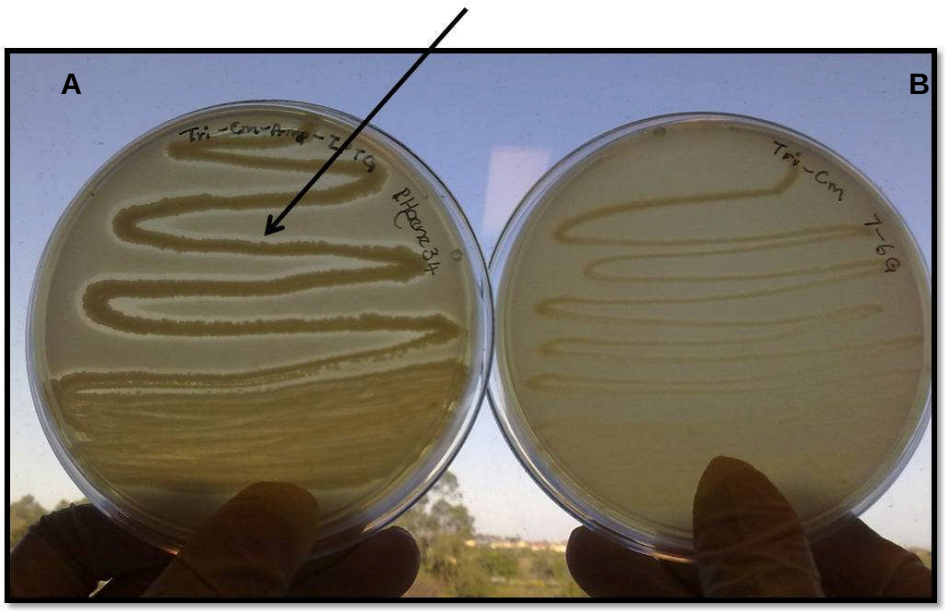


Figure 4.2: RH-pET clone harbouring RHgene34 gene with lipolytic activity (A) and control 7-6G clone with no lipolytic activity (B) on tributyrin agar indicator plates.

The tributyrin plate screen relies on the zone of clearing around the lipolytic bacterial colonies. However, tributyrin is not considered to be a true lipid as it can be cleaved by other esterases and disperses easily in water (Samad *et al*., 1989). For this reason, ethyl ferulate and olive oil-Rhodamine B agar plates were further employed to help distinguish RHgene34. In addition, to confirm that RHgene34 was responsible for the ferulic acid esterase activity conferred by Try 11 fosmid, RH-pET was tested on ethyl ferulate (Figure 4.4). Furthermore, RH-pET was screened on olive oil-Rhodamine B. The orange zone at the centre of the plate indicates the degradation of olive oil within the media (Figure 4.5). The results suggest that RHgene34 is not a true esterase as it showed fluorescence on olive oil-Rhodamine B plates. This enzyme could be a lipase as lipases can degrade tributyrin, ethyl ferulate and olive oil-Rhodamine B (Jaeger *et al*., 1999).

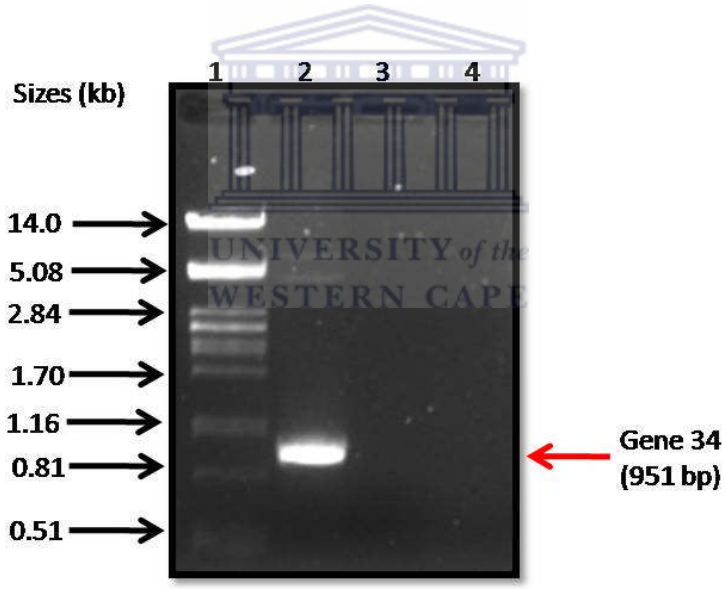


Figure 4.3: PCR amplification of RHgene34 using gene specific primers (Table 2.2) for confirmation of cloning into the pET vectors. Lane 1: DNA molecular marker Lambda *Pst*I digested DNA; Lane 2: PCR amplified gene of RHgene34; Lane 3 and 4: negative control.

Figure 4.4: RH-pET clone displaying FA activity on ethyl ferulate plates (B). The control (*E. coli* Rosetta pET21a) with no lipolytic activity (A).

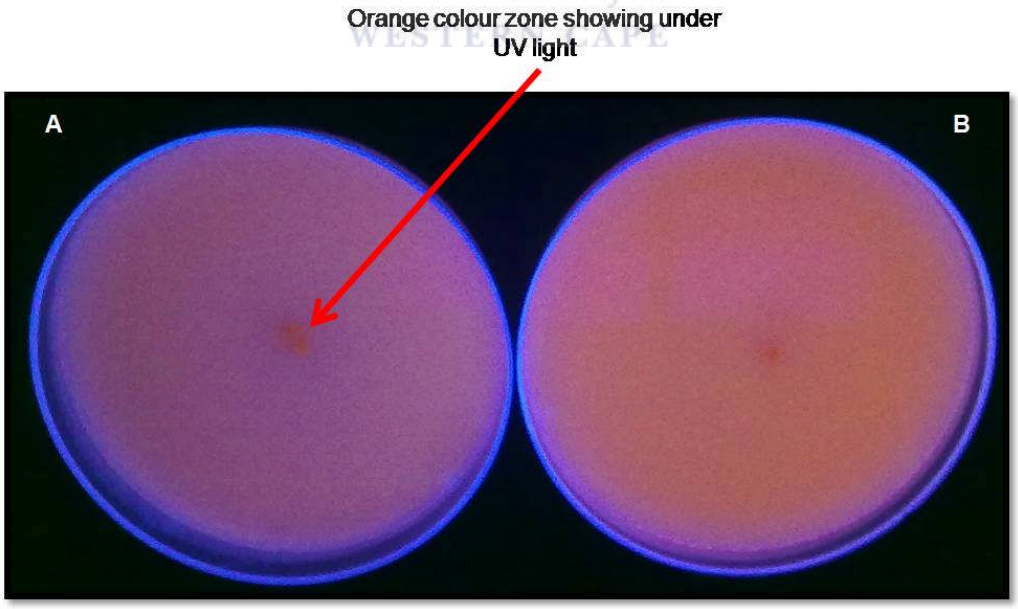


Figure 4.5: RH-pET clone displaying lipase activity (A) and control 7-6G clone (B) on olive oil-Rhodamine B lipase agar plates viewed under UV light.

4.2 Expression of RHgene34

Cell extracts of *E. coli* Rosetta(DE3)pLysS RH-pET were analysed on polyacrylamide gels and Coomassie brilliant blue staining in order to determine the expression of RHgene34 gene product and to monitor the degree of purity (Figure 4.6). A protein band migrating at ~34 kDa was present in cells which had been induced with IPTG (Figure 4.6: lanes 4 and 6). The calculated molecular mass of the gene product of RHgene34 is 34 kDa (including the His-Tag). Based on the information of SignalP (chapter 3, section 4), it was assumed that the expression of the RHgene34 gene in *E. coli* would not require co-expression with any flanking genes. This was proved to be the case as RHgene34-encoding gene was successfully expressed in *E. coli* Rosetta without the requirement of flanking sequences or genes. Expression was scaled up to larger volumes. Analysis by SDS-PAGE of the cytoplasmic fraction of RH-pET Rosetta, bound and eluted from His-bind resin, showed no additional protein band (Figure 4.7).

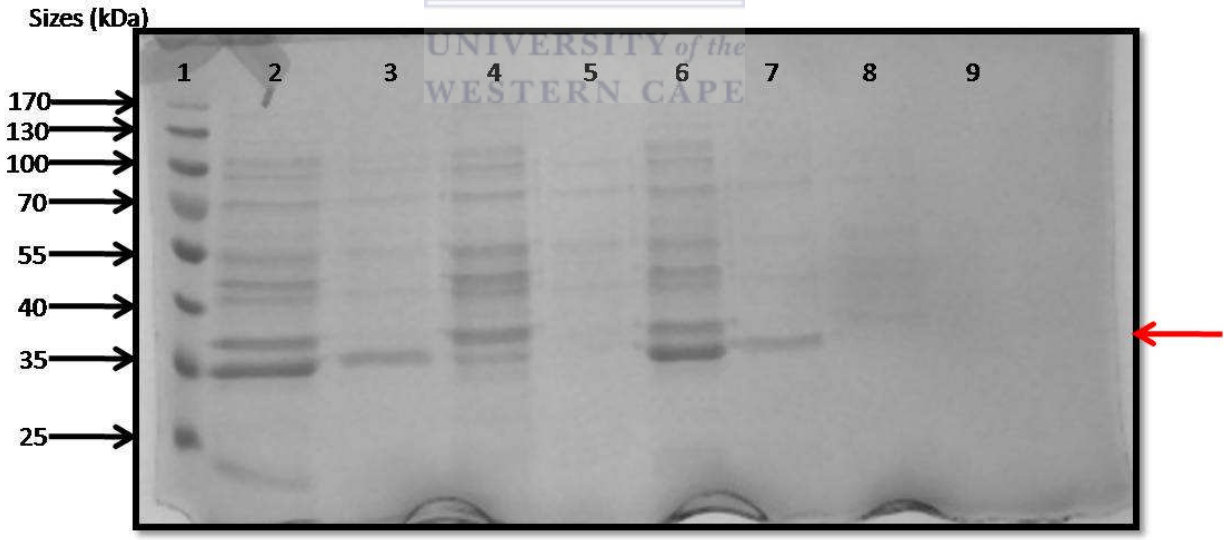


Figure 4.6: SDS-PAGE analysis of cell extracts of RH-pET in *E. coli* Rosetta(DE3)pLysS. The protein band corresponding to a size of 34kDa is indicated. Lane 1: protein molecular weight marker (#SM0671 Fermentas); Lane 2 and 6: IPTG induced total protein extract; Lanes 3 and 7: soluble fraction after overnight induction at 25 ºC and 30 ºC with IPTG, respectively; lane 4 and 8: uninduced total protein extract soluble; lane 5 and 9: soluble fraction of RHgene34 after overnight induction at 25 ºC and 30 ºC with IPTG.

Cell extracts of *E. coli* Rosetta(DE3)pLysS RH-pET were prepared by enzymatic methods or sonication and subjected to His-Tag affinity chromatography. The eluted fraction showed a protein band of approximately 34 kDa (Figure 4.7, lane 6). This correlated well with the predicted full length of RHgene34. The purity of the purified protein was more than 99 % according to SDS-PAGE analysis.

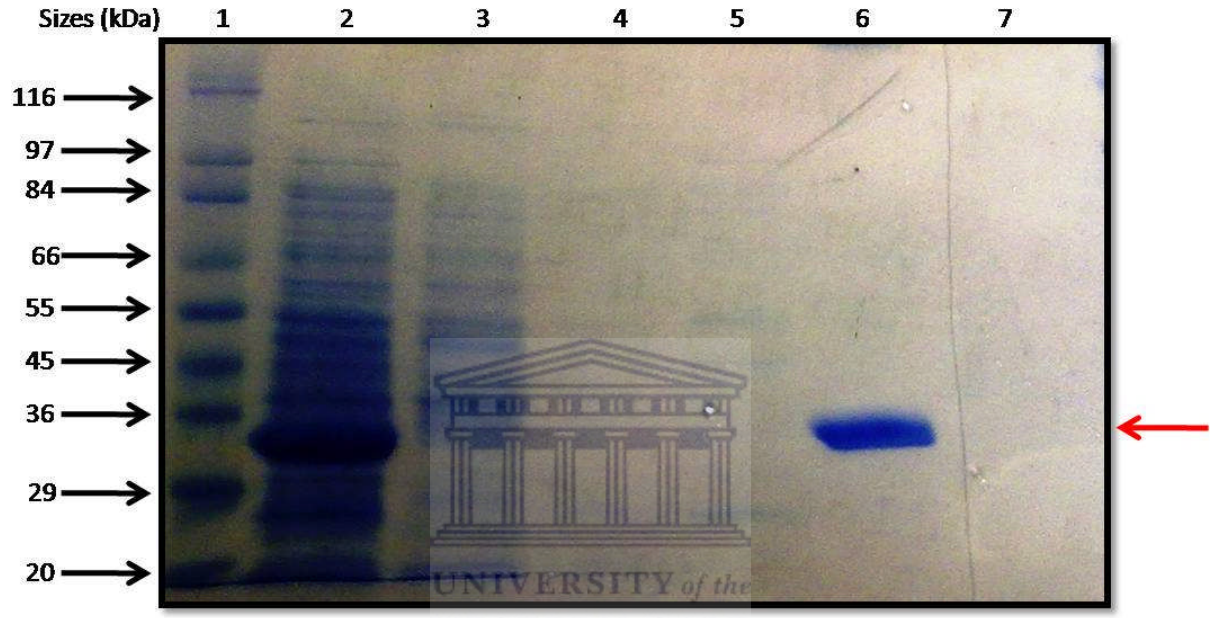


Figure 4.7: SDS-PAGE analysis of His-Tag purification of *E. coli* Rosetta(DE3)pLysS RH-pET. Lanes 1: protein molecular weight marker #S8445 (Sigma), lane 2: Total cell extract of RHgene34; Lane 3: Flow through elute; Lane 4: Elute from binding buffer; Lane 5: Elute from washing buffer; Lane 6: Eluted RHgene34 protein (MW 34 kDa). Lane 7: Elute from strip buffer.

To evaluate the quaternary structure for the RHgene34 protein, the apparent molecular mass was determined by size exclusion chromatography. The molecular weight (MW) of RHgene34 was calculated from the retention time of the peak absorbance by comparison with calibration standards having known molecular weights. The apparent molecular mass for the RHgene34 protein was estimated by extrapolation of the standard curve, using the straight line equation (y = -8.725 x + 194.7). Based on the calculation, the apparent molecular mass for RHgene34 protein is 66.08 kDa (RT: 15.39 min), as determined by SDS-PAGE the actual molecular

mass is 34 kDa. When the apparent mass is divided by actual mass we get a value of 1.8, this indicates that RHgene34 protein exists as a loose dimer, or a very tight monomer.

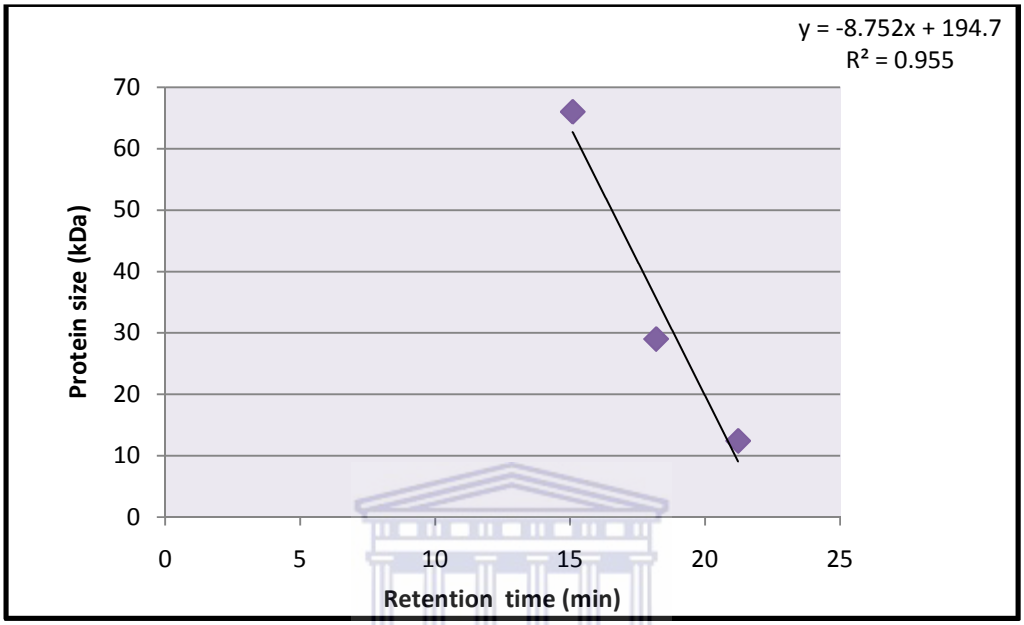


Figure 4.8: FPLC standard curve of the different retention times of albumin, cytochrome C and carbonic anhydrase. *E. coli* Rosetta(DE3)pLysS RH-pET was determined using the straight line equation determined by the line of best-fit.

4.3 Enzymatic Characterisation of the RHgene34 Gene Product

The substrate specificities of the RHgene34 gene product were examined with various *p*-Np ester substrates: C_2 (acetate), C_3 (propionate), C_8 (octanoate), C_{10} (decanoate), C_{12} (laurate) and C_{14} (myristate). RHgene34 was able to hydrolyse all esters with the highest activity towards long-chain fatty acids, and the C_{10} being the most favoured substrate (Figure 4.9). Activity on both short and long chain *p*-nitrophenyl ester substrates indicated a lipase-like activity for RHgene34. A similar substrate specificity profile was observed for a novel lipase isolated from mangrove sediment from the south Brazilian coast (Couto *et al*., 2010).

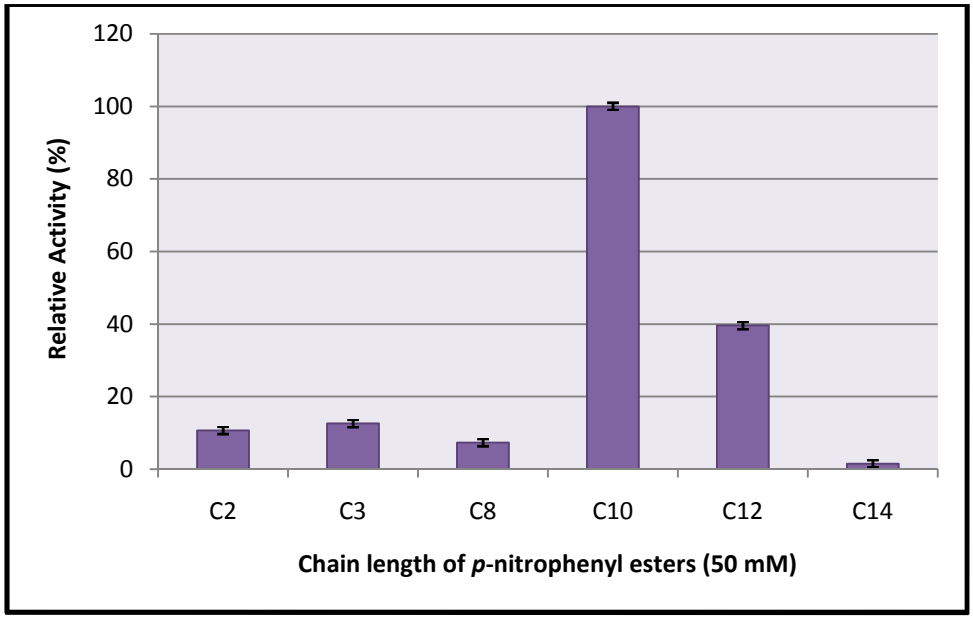


Figure 4.9: Substrate specificity of the purified RHgene34 enzyme toward *p*-nitrophenyl esters of varying chain lengths. Relative activity was shown as the percentage of the activity towards *p*-Np decanoate (C_{10}).

Qualitative analyses of the methyl ester substrates were performed. RHgene34 was found to be active on methyl *p*-coumarate, methyl caffeate and methyl ferulate however was less specific on methyl sinapate (Figure 4.10). This was indicative of a type B esterase (Wong, 2006).

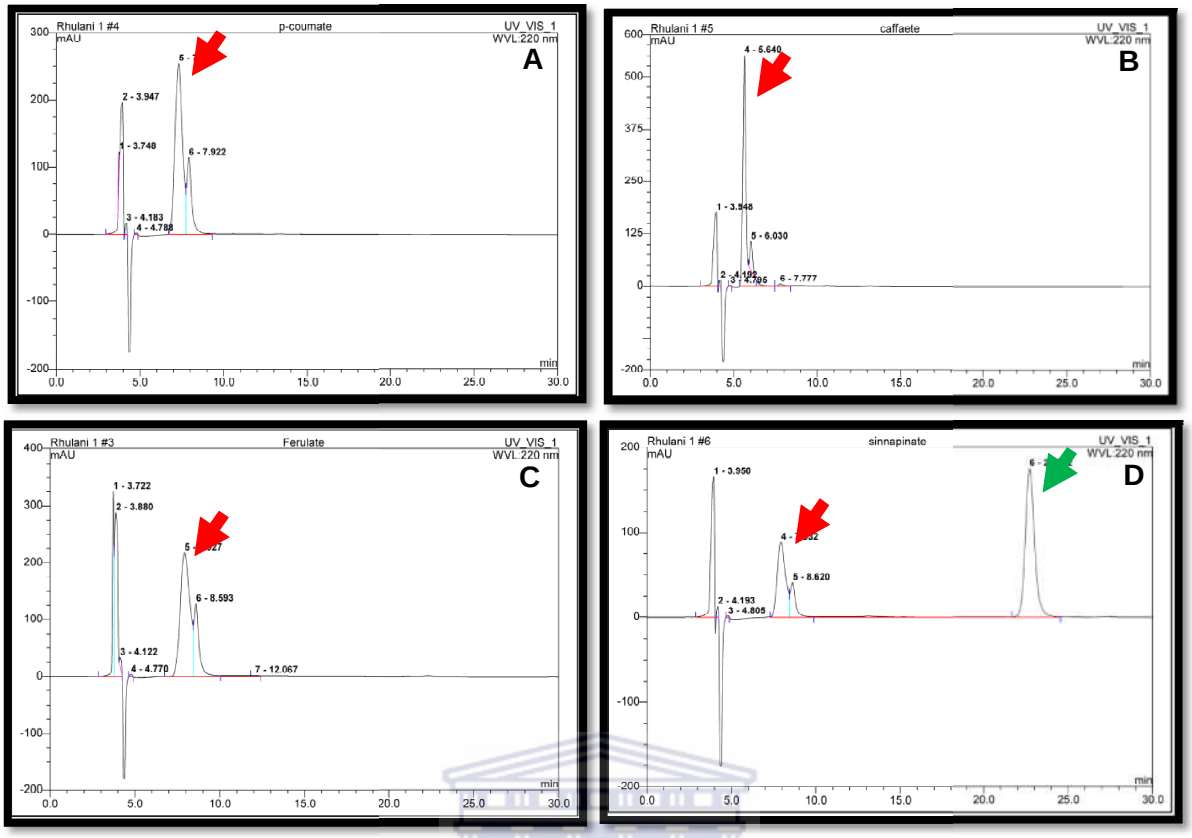


Figure 4.10: Reverse phase chromatography of reactions performed with purified RHgene34 on methyl *p*-coumate (A), methyl caffeate (B), methyl ferulate (C) and methyl sinapate (D). The red arrows indicate the product formed after hydrolysis and the green arrow, the substrate remaining.

Activity of the RHgene34 protein was tested over a pH range of 1.0 to 11.0, using C_{10} as a substrate (Figure 4.11). This enzyme showed activity in a broad pH range of 7.0 to 11.0. The RHgene34 gene product appears to be an alkaliphilic enzyme with an optimum activity at pH 9.0, since activity was observed from pH 7.0 and above and there was almost no hydrolysis of the substrate under extreme acidic conditions.

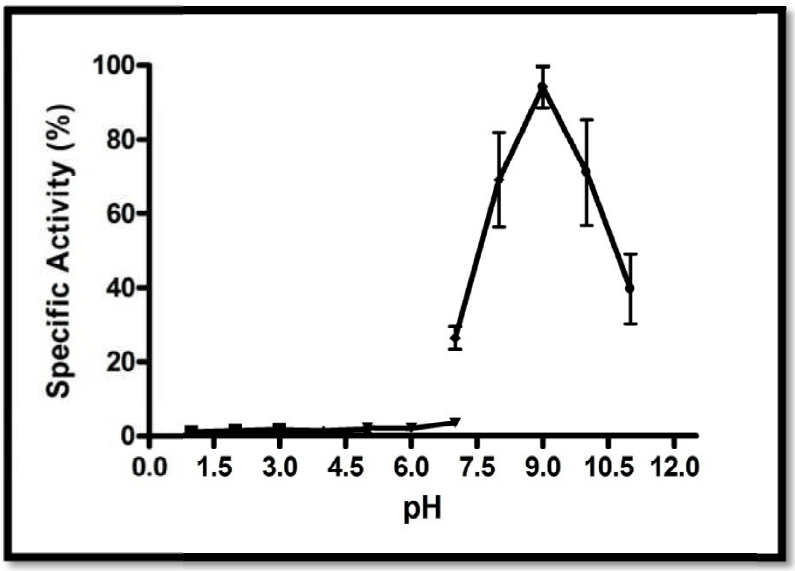


Figure 4.11: Effect of pH on purified RHgene34 protein activity using *p*-NP esterified with fatty acid of ten carbon chain length as the substrate. The activity was determined in various pH buffers at 25°C. Maximum activity at pH 9.0 was taken as 100 %.

Parameters such as temperature and pH of an enzyme are important to determine as they will determine the industrial viability of an enzyme. At temperatures higher than the optimal defined for a given enzyme, the protein will become denatured thereby rendering the fermentation process inefficient (Peteron *et al*., 2007). Fermentations conducted under thermophilic conditions (45 ºC to 80 ºC) are highly desirable as the risk of contamination by omnipresent mesophiles is reduced. Lipolytic activity of RHgene34 protein was determined from 15 ºC to 65 ºC with C_{10} as the substrate. The enzyme was active over a broad temperature spectrum (15 ºC to 55 ºC) and displayed optimum activity at 45°C (Figure 4.12), with complete deactivation at 65 ºC.

However, RHgene34 was not stable above 45 ºC, losing about 50 % of its activity after 30 min, suggesting that the enzyme is not an ideal candidate for industrial thermophilic fermentations. This may suggest that RHgene34 is either thermotolerant or a facultative thermozyme. A similar profile was published for an extracellular lipase from *Mucor hiemalis f. hiemalis*, which had an optimum temperature at 40 ºC and at 45 ºC had a residual activity of almost 80 % but at 55 ºC activity fell below 50 % and at 60 ºC was completely inactive (Hiol, *et al*., 1999).

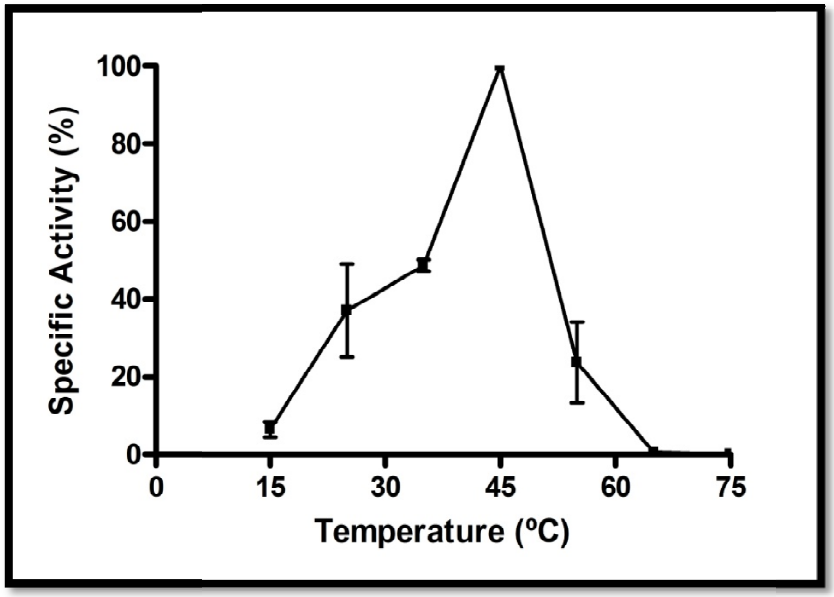


Figure 4.12: Effect of temperature on purified RHgene34 protein activity using *p*-NP esterified with fatty acid of ten carbon chain length as the substrate. The activity was determined at different temperatures at pH 7.5 in 100 mM sodium phosphate, 100 mM NaCl buffer. Maximum activity at 45 ºC was taken as 100 %.

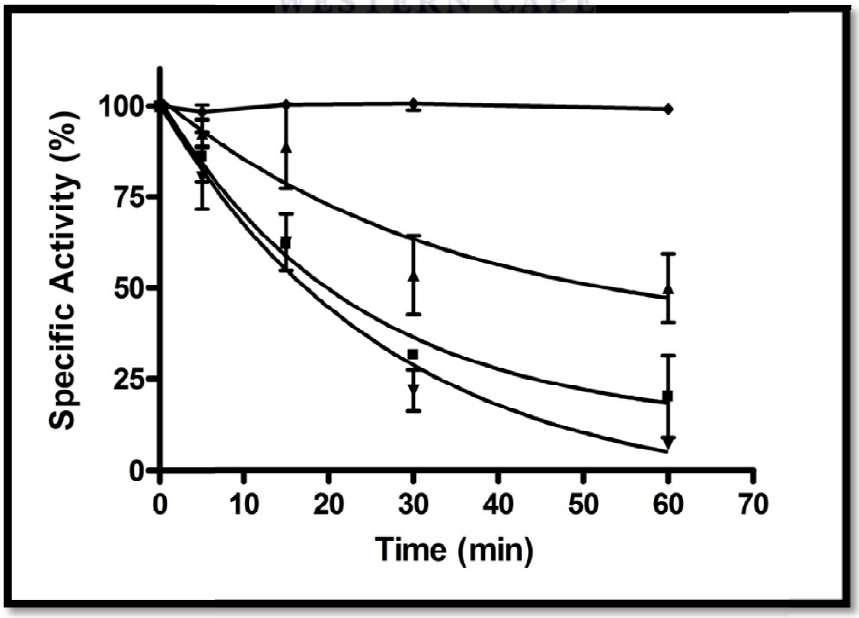


Figure 4.13: The thermal inactivation profile of RHgene34 at 25°C (♦), 35°C (▲), 45°C (■) and 55 ºC (▼).

Esterases and lipases differ both in chain-length substrate preferences and kinetics. While esterases catalyse small ester-containing molecules partially soluble in water, lipases have more affinity towards long-chain substrates insoluble in an aqueous environment. Therefore, esterases will display a typical Michaelis-Menten behaviour, while most lipases show an interfacial activation and therefore display a sigmoidal behaviour (Prim *et al*., 2006). The preliminary kinetic parameters of recombinant RHgene34 protein were determined. Values of k_{cat}, K_m or $K_{0.5}$ (apparent K_m) and V_{max} and Hill constant are shown in Table 4.1 and 4.2.

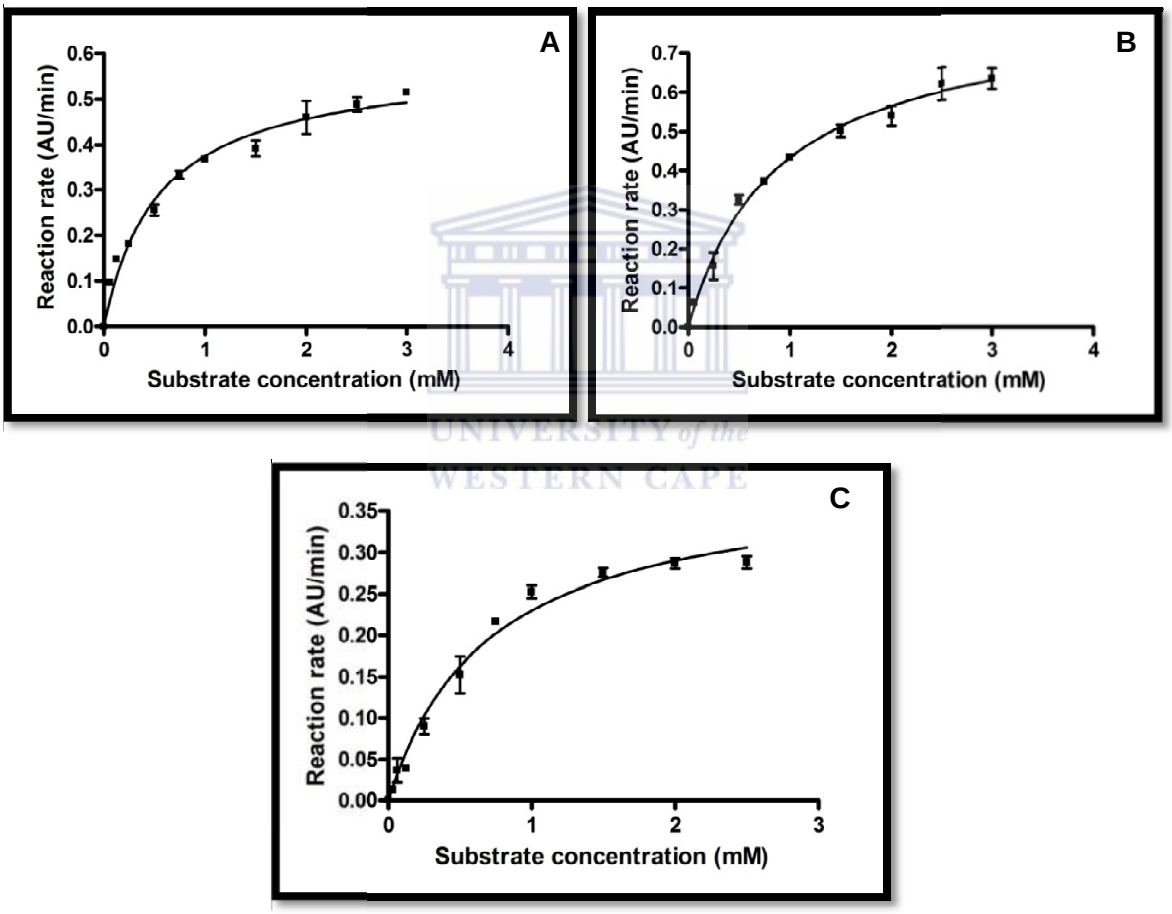


Figure 4.14: Kinetic parameters of purified RHgene34 gene product towards three *p*-Np ester substrates. **A**: Michaelis-Menten plot of RHgene34 towards C_2. **B**: Michaelis-Menten direct linear plot of RHgene34 towards C_3. **C**: Michaelis-Menten direct linear plot of RHgene34 towards C_8. Kinetic data were fitted using GraphPad Prism software (San Diego, USA).

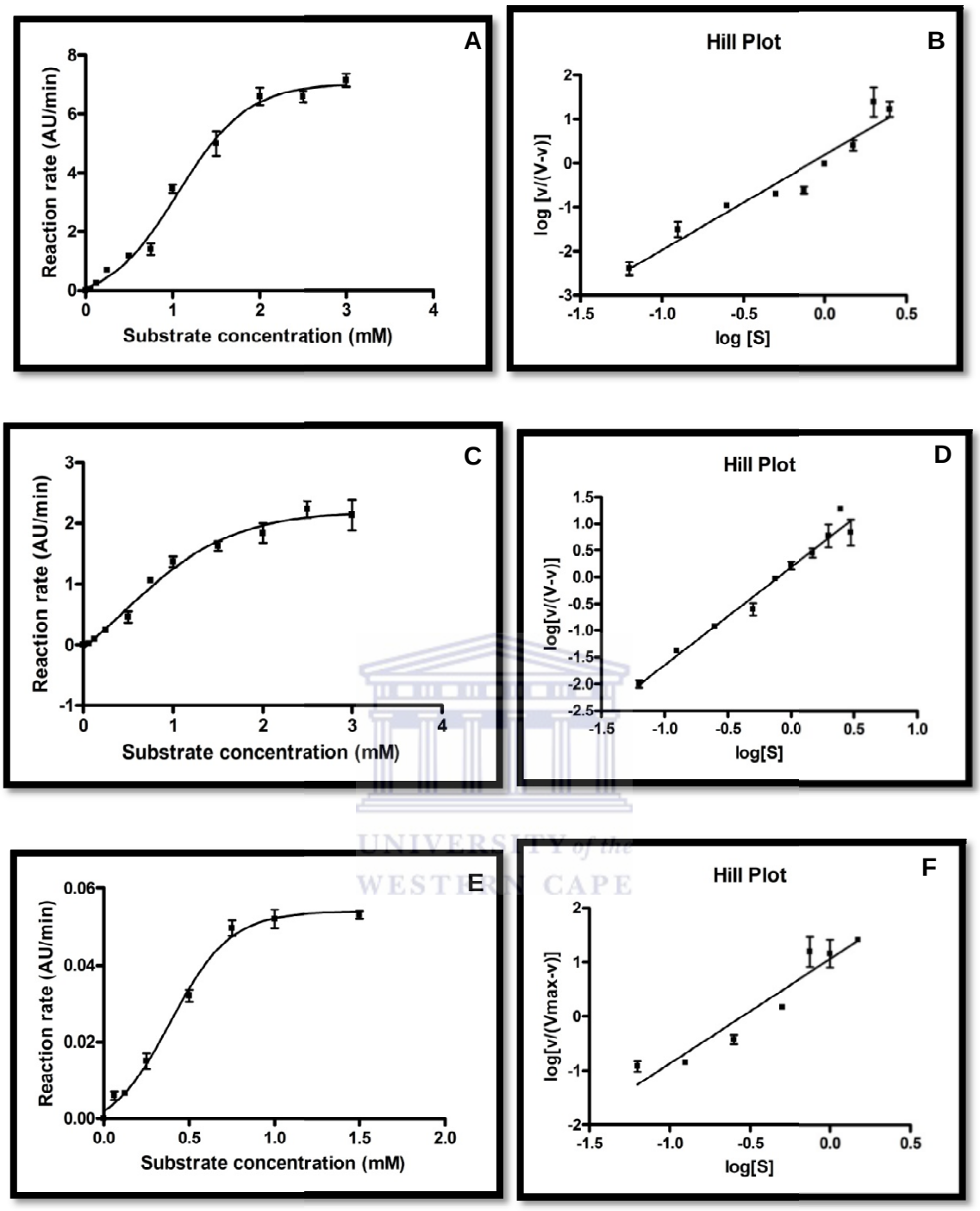
Table 4.1: Kinetic parameters of the RHgene34 enzyme using Michaelis-Menten *p*-Np esters.

Substrate	K_m (mM)	V_{max} (Umg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}/mM)
C_2	0.55	128.90	36.52	66.40
C_3	0.88	180.01	51.00	57.95
C_8	1.70	376.31	106.62	62.72

The kinetic constants for RHgene34 were calculated (Table 4.1) using the Michaelis-Menten nonlinear regression hyperbola plot. The K_m and V_{max} values for C_2 were approximately 3-fold lower than that of C_8. A high K_m value implies that the enzyme does not have a high affinity for the substrate and therefore a higher concentration of substrate is required to half-saturate the RHgene34. The catalytic turnover (k_{cat}) for C_2 was 3-fold lower than that for C_8 and, the overall efficiency (K_{cat} and K_{cat}/K_m) for C_8 is higher than C_2. This result indicates that although the enzyme is taking longer to recognise and bind to the C_8 substrate, once bound and activated the conversion to product is rapid.

The apparent K_m values for the longer chains (11.48 mM C_{10}, 2.75 mM C_{12} and 2.46 mM C_{14}) are shown in Table 4.2. Here, it is clear that RHgene34 has a lower substrate affinity for C_{10} as compared to C_{14}. However, the K_{cat} and $K_{cat}/K_{0.5}$ values show that the enzyme was more efficient in converting C_{10} to product. Similar to the short chain substrates, the enzyme takes longer to recognise and bind the C_{10} substrate, but as soon as it is bound, hydrolysis is rapid and could explain the sigmoidal curve. The variation of $k_{cat}/K_{m(0.5)}$ towards the short and long chain ester substrates suggests that different reaction mechanisms might be in use when substrates are bound with the enzyme (Ellenby *et al*., 1999).

The enzyme follows sigmoidal activity on long chain substrates, shown by a quick increase in velocity on the graph, reflecting how the binding of one subunit increases the chance that the other subunits will also bind to a substrate. The Hill coefficient, (*h*), is a quantitative measure of cooperativity in a binding process. This value is estimated from the Hill plot, where the value of 1 indicates independent binding and



Legend:

S – Substrate concentration (mM)

V – Final velocity (ms^{-1})

V – Initial velocity (ms^{-1})

Figure 4.15: Kinetic parameters of purified RHgene34 gene product towards three *p*-Np ester substrates. **A**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **B**:

logarithmic Hill plot towards C_{10}. **C**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **D**: logarithmic Hill plot towards C_{12}. **E**: Sigmoidal dose-dependent plot of RHgene34 with its corresponding **F**: logarithmic Hill plot towards and C_{14}. Kinetic data were fitted using GraphPad Prism software (San Diego, USA).

a value greater than 1 shows positive cooperativity where binding of one ligand increases the binding of subsequent ligands at other sites on a multimeric macromolecule (Weiss, 1997). The Hill coefficient of 2.16 on C_{10} suggests that the reaction of RHgene34 is positively cooperative (Hill, 1910), which correlates with the dimeric nature of this enzyme.

Table 4.2: Kinetic parameters of the RHgene34 enzyme using non Michaelis-Menten *p*-Np esters.

Substrate	$K_{0.5}$ (mM) [apparent K_m]	V_{max} (Umg^{-1})	k_{cat} (s^{-1})	$k_{cat}/K_{0.5}$ (s^{-1}/mM)	*h* (Hill coefficient)
C_{10}	11.48	1560.82	442.23	38.52	2.16
C_{12}	2.75	488.10	138.30	50.29	1.83
C_{14}	2.46	11.94	3.38	1.37	1.93

CHAPTER 5

General Discussion and Conclusion

The demand for bioethanol is continuously increasing due to the rapid depletion of fossil fuel reserves. Bioethanol is mainly produced from starchy (high in simple sugars) biomass sources like maize, sugarcane and wheat, known as first generation bioethanol. The source material which drives such production is making bioethanol more expensive than fossil fuel by diverting food crops, water and land resources for biofuel production, when food security is already a recognised problem (Srinivasan, 2009). This has forced the development of second generation biofuel from lignocellulosic biomass (Naik *et al*., 2010), to lower the production costs. Both depolymerising and side-chain cleaving enzymes, such as ferulic acid esterases (FAEs), are important in the complete degradation of lignocellulose polymers. This process generates a mixture of sugars when pretreated, alone or in combination, with enzymatic action followed by fermentation to bioethanol (Fazary and Ju, 2008). FAEs are inducible extracellular enzymes that hydrolyse the ester bonds between ferulic acid and the xylan backbone, resulting in increased accessibility for enzymatic attack on hemicellulose (Garcia *et al*., 1998; Mackensie *et al*., 1987). Due to this important role, the amount of research directed to FAEs has increased over the last decade. This is the result of recent discoveries in the isolation, purification and characterisation of fungal and bacterial FAEs (Fazary and Fu, 2007; Topakas *et al*, 2007). FAEs are classified into types according to their specificities to hydroxycinnamic acids (Rumbold *et al*., 2003). Differences in specificity are important in determining the optimal synergy between FAEs and other hemicellulases. When used in a mixed cocktail, the pH, temperature and stability profile of the enzymes would have to be compatible. Therefore, the discovery of novel FAEs with new properties is of importance (Shin and Chen, 2006).

Lipolytic enzymes are widely distributed among all the different bacterial taxa. Microbial lipases are one of the most versatile enzymes and offer a range of bioconversion properties such as esterification, interesterification, hydrolysis,

alcoholysis, acidolysis and aminolysis (Virk *et al*., 2011; Jeager *et al*., 1994). They also offer advantages which include stereospecificity, substrate specificity and the ability to hydrolyse substrates at the interface of water, as well as soluble and insoluble substrates. The ester products are used in the food industry to synthesise flavours and aroma constituents (Gandhi *et al*., 1995). Most industrial processes, such as crude oil refinery, where lipases play a role are required to be at temperatures over 45 ºC. Therefore the lipases need to be stable at temperatures around 50 ºC. Among other desirable characteristics, the most important for industrial lipases are alkali tolerance and thermostability (Haki and Rakshit, 2003; Sharma *et al*., 2001). Therefore, lipolytic enzymes are a highly sought after class of enzymes. The difference between these enzymes is that FAEs hydrolyse short ester-containing molecules that are partly soluble in water, while lipases preferentially hydrolyse water-insoluble long chain ester substrates (Virk *et al*., 2011).

Due to the increased demands for enzymes that can function efficiently under industrial conditions, considerable effort has been dedicated to the search for more robust varieties (Rozzell, 1999). Although many enzymes have been discovered, it is still not sufficient to meet the industry's demands (Madigan and Mars, 1997; Herbert and Sharp, 1992). Enzymes are the desired catalyst for industrial processes and, up to now, most industrial enzymes are of microbial origin. Microorganisms are a source of useful enzymes and are easy to manipulate to increase yields and quality (Price *et al*., 1995). Thermophiles are habitat-adapted and may therefore be a source of enzymes suited for harsh industrial processes. The valuable properties of such enzymes would be to operate at high temperatures, or solvents and detergents, giving these enzymes considerable potential for many biotechnological and industrial applications. Modern biotechnological processes generally operate at elevated temperatures because this reduces the risk of mesophilic contaminations. In addition, high temperatures increase enzyme reaction rates due to a decrease in viscosity and an increase in the substrate diffusion coefficient (Haki and Rakshit, 2003). This is why this study focused on screening thermophilic novel industrial enzymes.

Metagenomics is a rapidly growing field of research that has been successfully employed as a powerful tool for the discovery of enzymes with novel biocatalytic

activities from unculturable microbial communities (Kakirde *et al*., 2010; Kennedy *et al*., 2008). This technique does not require cultivation of microorganisms, which may be impossible to grow and, in addition, it provides genetic information on the overall microbial community from populations containing potentially uncultured members (Voget *et al*., 2003). The first step is to construct metagenomic libraries from environmental samples, then transform into *Escherichia coli*, and then isolate the gene conferring the desired activity. This method can be problematic as expression in *E. coli* is biased and inefficient. Additionally, it is difficult to develop assays using agar plates supplemented with substrates. The alternative is a PCR-based method, where novel genes are directly isolated from metagenomes. This particular method avoids the difficulties associated with protein expression in host cells but the main limitations of this approach is that homologous regions are required in order to probe the sequences (Uchiyama and Miyazaki, 2009; Rhee *et al*., 2005). When metagenomics is coupled with highly efficient and low cost high-throughput screening techniques, a large number of new biocatalysts and small molecules can be recorded (Simon and Daniel, 2009; Elend *et al*., 2006). A metagenomic strategy has previously been used to isolate lipolytic enzymes (Kakirde *et al*., 2010), such as FAEs and lipases.

Due to the potential that metagenomics offers for the isolation of a novel esterase, a metagenomic functional screening was employed and a large insert fosmid library prepared using sediment from a Malawian hot spring (72 - 78 ºC; pH 6.5) was screened in this study. Screening was performed on tributyrin-supplemented agar, a general esterase/lipase screen where positive clones are identified by a zone of clearing (Jarvis and Thiele, 1997). However, tributyrin is not a true lipid as it is soluble in water and cleaved by some esterases, and will therefore result in the isolation of a wide range of lipolytic activities (Litthauer *et al*., 2010). This screening led to the isolation of a fosmid clone, Try 11, which formed the basis of this study. This clone additionally gave positive activity on ethyl ferulate and olive-oil Rhodamine B, indicating that RHgene34 may be a lipase. To classify the enzyme, sequence comparison was used for preliminary analysis and functional characterisation to confirm activity.

In order to identify the gene sequence conferring the activity, the fosmid was fully sequenced. Annotation of the ORFs revealed that 38 % of proteins from the fosmid clone had high similarity to proteins from *Enterobacter,* a mesophilic organism. One of the ORFs (RHgene34) showed ≤41 % to known α/β hydrolase fold family protein in the NCBI database. This moderately low identity suggested that this may be a novel biocatalyst.

P-nitrophenyl esters of various carbon chain lengths (C_2 to C_{14}) were used to perform a preliminary biochemical characterisation of the enzyme activity. Maximum relative activity was obtained for the C_{10} chain substrate. The enzyme was found to have a pH optimum of 9, which could be advantageous in many biotechnological applications that require alkaline conditions (such as alkaline pulping) (Sigoillot *et al*., 2005; Record *et al*., 2003). The temperature optimum was 45 ºC; not surprising considering that it was isolated from a hot spring. However, considering the temperature of the hot spring, it was expected that this lipolytic enzyme would be more thermophilic, instead of having an optimum temperature of 45 ºC. This may be due to the isolation of remnant DNA from dead thermotolerant mesophiles during metagenomic extraction. It is important to consider that a hot spring is an open environment and therefore subject to contamination and metagenomics is unbiased to living or dead microorganisms. The enzyme was not very stable at 45 ºC, losing 50 % of its activity after 30 min, which suggests that the enzyme is not an ideal candidate for industrial thermophilic fermentations, but may be used for application in mesophilic fermentation processes employed in the pharmaceutical industry where enzymatic stability at 30 ºC to 45 ºC is ideal (Aravindan *et al*., 2007).

Considering highest relative activity was obtained on C_{10}, it suggested that this enzyme would be classified as a lipase. However, based on the activity on several substrates it is suggested that this enzyme has a broad substrate range. While *p*-Np esters aid in basic preliminary characterisation, they are not suitable for specific lipase assays because some esterases can also cleave them, but they do give an idea of the carbon chain length an enzyme can catalyse (Stuer *et al*., 1986). In addition, *p*-Np esters are easier to use and detect but because they have no substitutions or side chains, methyl substrates were used to give an indication of substrate specificity. RHgene34 displayed activity on methyl esters, namely methyl

ferulate, methyl *p*-coumarate, methyl caffeate but not as efficient on methyl sinapate (Wong, 2006). Based on this, we can classify RHgene34 as a typical B type esterase. This was not surprising as lipases are perfectly capable of catalysing a broad range of substrates including esterase substrates (Jaeger *et al*., 1999) such as methyl hydroxycinnamate substrates (Wong, 2006).

Most lipases are known not to follow the Michaelis-Menten model, because they have more than one homogeneous phase. Therefore, the Hill model is proposed to describe the kinetics of lipases. Based on the Hill plot analysis, it is proposed that RHgene34 functions as a dimer, because it describes the cooperativity between subunits of an enzyme. This was further supported by size exclusion, which indicated that it folds as a loose dimer. The RHgene34 was found to be positively cooperative, meaning that the binding of a substrate to one subunit increases the affinity at other subunits (Weiss, 1997). This is an interesting result since all lipases are reportedly monomeric (Aravindan *et al*., 2007). We may have isolated one of the first bacterial lipases that exist as a dimer. The dimerisation, however, appears to be loose based on the homology analysis, and this may explain the lack of thermostability. It has been found that thermostable proteins form tight dimers, deletion or shortening of loops and fewer or smaller cavities (Kumar and Nussinov, 2001), therefore, the tighter the protein the more thermostable it is.

Homology modelling of RHgene34 clearly identified the characteristic α/β hydrolase fold. The fold consists of 8 β-sheets connected by α-helices, which contain the residues of the catalytic triad (Jaeger *et al*., 1999). The RHgene34 protein does not appear to have the common lid structure that most lipases possess and the structures are not tightly packed. It should be noted that some lipases, such as that of *Candida antarctica* B (Fojan *et al*., 2000) and *Streptomyces exfoliatus* (Nardini and Dijkstra, 1999), have no lid structure but are still classified as lipases. The mechanism of action however, is not clear. These lipases are of fungal origin, while RHgene34 is of microbial origin, making it a very interesting lipase as there are very few known microbial lipases without a lid, with a broad substrate range and folding as a dimer.

The lid structure is very important in lipases; it is responsible for the unique interfacial activation that allows lipases to catalyse a broad range of substrates. When long chain fatty acids form an emulsion, there is a strong increase in enzyme activity. When the first 3D structures of lipases were first determined, interfacial activation was a clear explanation for this behaviour. The active site of lipases was found to be covered with a C-terminal lid structure which consequently makes the active site inaccessible to substrates i.e., making the enzyme inactive (Wong, 2006; El-Kouhen *et al*., 2005). However, a conformational change takes place when a lipase is bound to a lipid interface; causing the lid to open and exposing the active site. The lid reveals the hydrophobic residues to the lipid interface increasing the bond between the enzyme and the lipid surface. This phenomenon explains interfacial activation with the lid causing inactivation if no lipid interface is present, and has been used to distinguish between true lipases and esterases (Wong, 2006; Manco *et al*., 1998).

True lipases are usually defined as enzymes that show interfacial activation in the presence of long chain fatty acids. According to Jaeger *et al*., (1994), if an enzyme hydrolysing these substrates does not show interfacial activation, it should be classified as an esterase. Based on the predictive model, RHgene34 does not possess a lid. It does however, have an N-terminal loop-like structure which is not associated with the active site. The homology model was not a good representation of the enzyme, as shown by the Ramachandran plot; therefore it might be that the flap may be acting as a lid, covering the hydrophobic residues inside the active site. Additional, RHgene34 hydrolyses both short and long chain substrates demonstrating interfacial activation with long chain substrates. Addition research is warranted to fully characterise the true classification of this enzyme.

Therefore, in this study, metagenomics has proven very useful and successful in identifying a novel lipolytic enzyme, not only at the sequence level, but also at the structural level. Esterases and lipases have a wide range of biotechnological applications, from the food industry to the production of perfumes and fuels and the development of pharmaceuticals (Elend *et al*., 2006). Due to the ability of RHgene34 to hydrolyse a wide range of substrates, its alkaline pH and thermotolerance, it is an enzyme of value for industrial application, particularly in mesophilic fermentations.

Future work includes triacylgylcerol studies for accurate lipase classification, protein crystallisation and hydrolysis trials for bioethanol production.

CHAPTER 6

References

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

Antoni, D., Zverlov, V.V. and Schwars, W.H. (2007). Biofuels from microbes. *Applied Microbial Biotechnology* **26**: 47-55.

Aravindan, R., Anbumathi, P. and Viruthagiri, T. (2007). Lipase applications in food industry. *Indian Journal of Biotechnology* **6**: 141-158.

Archadi, M. and Sellstedt, A. (2008). Production of energy from biomass, *in Introduction to chemicals from biomass*. Clark, J. and Deswarte, F. (Eds). Pg. 152. John Wiley and Sons, Ltd, United Kingdom.

Arpigny, J.L. and Jaeger, K.E. (1999). Bacterial lipolytic enzymes: classification and properties. *Biochemical Journal* **343**: 177-183.

Aurilia, V., Parracino, A. and D'Auria, S. (2008). Microbial carbohydrate esterases in cold adapted environments. *Gene* **410**: 234-240.

Balat, M., Balat, M., Kirtay, E. and Balat H. (2009). Main routes for the thermo-conversion of biomass into fuels and chemicals. Part 1: Pyrolysis systems. *Energy Conversion and Management* **50**: 3147-3157.

Balat, M., Balat, H. and Oz, C. (2008). Progress in bioethanol processing. *Progress in energy and Combustion Science* **34**: 551-573.

Becker, J. and Boles, E. (2003). A modified *Saccharomyces cerevisiae* strain that consumes L-arabinose and produces ethanol. *Applied and Environmental Microbiology* **69**(7): 4144-4150.

Blum, D., Kataeva, I.A., Li, X-L. and Ljungdahl, L.G. (2000). Feruloyl esterase activity of the *Clostridium thermocellum* cellulosome can be attributed to previously unknown domains of XynY and XynZ. *Journal of Bacteriology* **182**(5): 1346-1351.

Blumer-Schuette, S.E., Kataeva, L., Westpheling, J., Adams, M.W. and Kelly, R.M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Current Opinion in Biotechnology* **19**: 210-217.

Bornscheuer, U.T. (2002). Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiology Reviews* **26**: 73-81.

Bradford, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254.

Brown, C.R. (2003). Biorenewable resources. Engineering New Products from agriculture. Iowa State Press. Blackwell Publishing. Pp. 59-73.

Bunzel, M., Ralph, J., Funk, C. and Steinhart, H. (2005). Structural elucidation of new ferulic acid-containing phenolic dimers and trimers isolated from maize bran. *Tetrahedron Letters* **46**: 5845-5850.

Carr, P.D. and Ollis, D.L. (2009). Alpha/beta hydrolase fold: an update. *Protein and Peptide Letters* **16**(10): 1137-1148.

Chang, V.S. and Holtzapple, M.T. (2000). Fundamental factors affecting biomass enzymatic reactivity. *Applied Biochemistry and Biotechnology* **84-86**: 5-37.

Chang, V.S., Nagwani, M., Kim C-H. and Holtzapple, M.T. (2001). Oxidative lime pretreatment of high-lignin biomass. *Applied Biochemistry and Biotechnology* **94**: 1-28.

Christov, L.P. and Prior, B.A. (1993). Esterases of xylan-degrading microorganisms: Production, properties and significance. *Enzyme Microbiological Technology* **15**: 460-475.

Cherubini, F. (2010). The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management* **51**: 1412-1421.

Cosgrove, D.J. (2001). Wall structure and wall loosening. A look backwards and forwards. *Plant Physiology* **125**: 131-134.

Couto, G.H., Glogauer, A., Faoro, H., Chubatsu, L.S., Souza, E.M. and Pedrosa, F.O. (2010). Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genetics and Molecular Research* **9**(1): 514-523.

Crepin, V.F., Faulds, C.B. and Connerton, I.F. (2003). A non-modular type B feruloyl esterase from *Neurospora crassa* exhibits concentration-dependent substrate inhibition. *Biochemical Journal* **370**: 417-427.

(a)Crepin, V.F., Faulds, C.B. and Connerton, I.F. (2004). Functional classification of the microbial feruloyl esterases. *Applied Microbiology and Biotechnology* **63**: 647-652.

(b)Crepin, V.F., Faulds, C.D. and Connerton, I.F. (2004). Identification of a type-D feruloyl esterase from Neurospora crassa. *Applied Microbiology and Biotechnology* **63**: 567-570.

Demirbas, A. (2007). Harnessing energy from plant biomass. *Current Opinion in Chemical Biology* **11**: 677-684.

Demirbas, A. (2009). Biofuels securing the planet's future energy needs. *Energy Conversion and Management* **50**: 2239-2249.

Demirijian, D.C., Morís-Varas, F. and Cassidy, C.S. (2001). Enzymes from extremophiles. *Current Opinion in Chemical Biology* **5**: 144-151.

de Souza Silva, C.M.M., de Melo, I.S. and de Oliveira, P.B. (2005). Ligninolytic enzyme production by *Ganoderma spp*. *Enzyme and Microbial Technology* **37**(3): 324-329.

de Vrije, T., Bakker, R.R., Budde, M.A., Lai, M.H., Mars, A.E. and Classen, P.A. (2009). Efficient hydrogen production from the llignocellulosic energy crop *Miscanthus* by the extreme thermophilic bacteria *Caldicellulosiruptor saccharalyticus* and *Thermotaga neapolitana*. *Biotechnology for Biofuels* **2**: 12.

Elend, C., Scheisser, C., Leggewie, C., Babiak, P., Carballeira, J.D., Steele, H.L., Reymond, J.L., Jaeger, K.E. and Wtreit, W.R. (2006). Isolation and biochemical characterisation of novel metagenome-derived esterases. *Applied and Environmental Microbiology* **72**(5): 3637-3645.

Ellenby, B., Sjoblom, B. and Lindskog, S. (1999). Changing the efficiency and specificity of the esterase activity of human carbonic anhydrase II by site-specific mutagenesis. *European Journal of Biochemistry* **262**: 516-521.

Emaunuelsson, O. Brunak, S., von Heijne, G and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953-971.

Etta, P.D., Preston, J.L., Bassham, S., Cresko, W.A. and Johnson, E.A. (2011). Local *De Novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS One* **6**(4): e18561.

Fan, L.T., Lee, Y.H. and Beardmore, D.H. (1980). Mechanism of the enzymatic hydrolysis of cellulose: Effects of major structural features of cellulose on enzymatic hydrolysis. *Biotechnology and Bioengineering* **22**: 177-199.

Faulds, C.B. (2010). What do feruloyl esterases do for us? *Phytochemistry Reviews* **9**: 121-132.

Faulds, C.B. and Williamson, G. (1991). The Purification and characterisation of 4-hydroxy-3-methoxycinnamic (ferulic) acid esterase from *Streptomyces olivochromogenes. Journal of General Microbiology* **137**: 2339-2345.

Fazary, A.E. and Ju, Y. (2007). Feruloyl esterases as biotechnological tools: current and future prospects. *Acta Biochimica et Biophysica Sinica* **19**: 811-878.

Fazary, A.E. and Ju, Y.H. (2008). The large-scale use of feruloyl esterases in industry. *Biotechnology and Molecular Biology Reviews* **3**(5): 95-110.

Fillingham, I.J., Kroon, P.A., Williamson, G., Gilbert, H.J. and Hazelwood, G.P. (1999). A modular cinnamoyl ester hydrolase from the anaerobic fungus *Piromyces equi* acts synergistically with xylanase and is part of a multiprotein cellulose-binding cellulase-hemicellulase complex. *Biochemical Journal* **343**: 215-224.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H-R., Ceric, G., Forslund, K., Eddy, S.R., Sannhammer, E.L.L. and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Research* **36**(Database Issue): D281-D288.

Fischer-Romero, C., Tindall, B.J. and Jüttner, B.J. (1996). *Tolumonas auensis* gen. nov., sp. nov., a toluene-producing bacterium from anoxic sediments of a freshwater lake. *International Journal of Systematic Bacteriology* **46**(1): 183-188.

Fojan, P., Jonson, P.H., Petersen, M.T.N. and Petersen, S.B. (2000). What distinguishes an esterase from a lipase: a novel structural approach. *Biochimie* **82**: 1033-1041.

Galbe, M., Zacchi, G. (2002). A review of the ethanol production from softwood. *Applied Microbial Biotechnology* **59**: 618-628.

Gandhi, N., Sawant, S. and Joshi, J. (1995). Studies on the lipozyme-catalysed synthesis of butyl laurate. *Biotechnology and Bioengineering Journal* **46**: 1-12.

Garcia, B.L., Ball, A.S., Rodriquez, J., Pérez-Leblic, M.I., Arias, M.E. and Copa-Patiño, J.L. (1998). Induction of ferulic acid esterase and xylanase activities in *Streptomyces avermitilis* UAH30. *FEMS Microbiology Letters* **158**: 95-99.

Garcia, B.L., Ball, A.S., Rodriguez, J., Perez-Leblic, M.I., Anas, M.E. and Copa-Patino, J.C. (1998). Production and characterisation of ferulic acid esterase activity in crude extracts by *Streptomyces avermitilis* CECT-3339. *Applied Microbiology and Biotechnology* **50**: 213-218.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASY server: (*In*) Walker, J.M. (*ed*): The proteomics protocols handbook. Humana Press. Pg 571-607.

Gírio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S. and Bogel-Łukasik, R. (2010). Hemicelluloses for fuel ethanol: A review. *Bioresource Technology* **101**: 4775-4800.

González-García, S., Gasol, C.M., Cagarrell, X., Rieradevall, J., Moreira, M.T. and Feijoo, G. (2010). Environmental profile of ethanol from poplar biomass as transport fuel in Southern Europe. *Renewable Energy* **35**: 1014-1023.

Goyal, A., Ghosh, B. and Eveleigh, D. (1991). Characterisation of fungal cellulases. *Bioresource Technology* **36**: 37-50.

Gray, K.A., Zhao, L. and Emptage, M. (2006). Bioethanol. *Current Opinion in Chemistry and Biology* **10**: 141-146.

Gromiha, M. M., Ahmed, S., et al. (2005). TEMBETA-NET: discrimination and prediction of membrane spanning beta-strands in outer membrane proteins. *Nucleic Acids Research* **33**: 165-167.

Hahn-Hagerdal, B., Galbe, M., Gorwa-Grauslund, M.F., Liden, G. And Zacchi, G. (2006). Bio-ethanol – the fuel of tomorrow from the residues of today. *Trends in Biotechnology* **24**: 549-556.

Haki, G.D. and Rakshit, S.K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresource Technology* **89**: 17-34.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*: 95-98.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **8**: 669-685.

Herbert, R. and Sharp, R. (1992). Molecular biology and biotechnology of Extremophiles, Chapman and Hall, NY.

Hilbert, M., Böhm,G. and Jaenicke, R. (1993). Structural relationships of homologous poroteins as a fundamental principle in homology modelling. *Proteins* **17**: 138-151.

Hill, A.V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology* **40**: 4-7.

Himmel, M.E. and Bayer, E.A. (2009). Lignocellulose conversion to biofuels: current challenges, global perspectives. *Current Opinion in Biotechnology* **20**: 316-317.

Hiol, A., Jonzo, M.D., Druet, D. and Comeau, L. (1999). Production, purification and characterisation of an extracellular lipase from *Mucor hiemalis f. hiemalis*. *Enzyme and Microbial Technology* **25**: 80-87.

Holtzapple, M. (2003). Cellulose, Hemicellulose and Lignin, in Encyclopedia of food science, food, technology and nutrition, 2nd Ed. Academic Press, London. Pp. 998-1007, 3060-3072, 3535-3542.

Hooft, R.W.W., Sander, C. and Vriend, G. (1996). Verification of protein structures: side-chain planarity. *Journal of Applied Crystallography* **29**: 714-716.

Hu, X.P. (2010). Identification and characterisation of novel cellulolytic genes using metagenomics. Department of Biotechnology, University of the Western Cape. Pp. 1-89.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., de Coastro, E., Lachaize, C., Langendijk-Genevaux, P.S. and Sigrist, C.J. (2007). The 20 years of PROSITE. *Nucleic Acids Research*: 245-249.

Jaeger, K.E., Dijkstra, B.W. and Reetz, M.T. (1999). Bacterial biocatalysts:molecular biology, three-dimensional structures and biotechnological applications of lipases. *Annual Reviews of Microbiology* **53**: 315-351.

Jaeger, K-E., Ransac, S., Dijkstra, B.W., Colson, C., van Heuvel, M. and Misset, O. (1994). Bacterial lipases. *FEMS Microbiology Reviews* **15**: 29-63.

Jarvis, G.N. and Thiele, J.H. (1997). Qualitative Rhodamine B assay which uses tallow as a substrate for lipolytic obligately anaerobic bacteria. *Journal of Microbiological Methods* **29**: 41-47.

Jegannathan, H.R., Chan, E-S. and Ravindra, P. (2009). Harnessing biofuels: A global Renaissance in energy production? *Renewable and Sustainable Energy Reviews* **13**: 2163-2168.

Jiyama, K., Lam, T.B-T. and Stone, B.A. (1994). Covalent cross-links in the cell wall. *Plant Physiology* **104**: 315-320.

Kakirde, K.S., Parsley, L.C. Liles, M.R. (2010). Size does matter: application-driven approaches for soil metagenome. *Soil Biology and Biochemistry* **42**: 1911-1923.

Kantarelis, E. and Zabaniotou, A. (2009). Valorisation of cotton stalks by fast pyrolysis and fixed bed air gastification for syngas production as precursor of second generation biofuels and sustainable agriculture. *Bioresource Technology* **100**: 942-947.

El-Kouhen, K., Blangy, S., Ortiz, E., Gardies, A.M., Ferté, N. and Arondel, V. (2005). Identification and characterisation of a triacylgylcerol lipase in Aradidopsis homologous to mammalian acid lipases. *FEBS Letters* **579**: 6067-6073.

Kennedy, J., Marchesi, J.R. and Dobson, A.D. (2008). Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microbial Cell Factories* **7**: 27.

Kleywegt, G.J. and Jones, T.A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure* **4**(12): 1395-1400.

Koseki, T., Furuse, S., Iwano, K. and Matsuzawa, H. (1998). Purification and characterisation of a feruloyl esterase from *Aspergillus awamori*. *Bioscience Biotechnology and Biochemistry* **62**(10): 2032-2034.

Koseki, T., Fushinobu, S., Ardiansyah, Shirakawa, H. and Komai, M. (2009). Occurrence, properties and applications of feruloyl esterases. *Applied Microbiology Biotechnology* **84**: 803-810.

Koskinen, P.E.P., Lay, C-H., Beck, S.R., Tolvanen, K.E.S., Kaksonen, A.H., Ã-rlygsson, J.H., Lin, C-Y. and Puhakka, J.A. (2007). Bioprospecting thermophilic microorganisms from Icelandic hot springs for hydrogen and ethanol production. *Energy and Fuels* **22**: 134-140.

Kotik, M. (2009). Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications. *Journal of Biotechnology* **144**: 75-82.

Krause, D.O., Denman, S.E., Mackie, R.I., Morrison, M., Rae, A.L., Attwood, G.T. and McSweeney, C.S. (2003). Opportunities to improve fibre degradation in the rumen: microbiology, ecology and genomics. *FEMS Microbiology Reviews* **797**: 1-31.

Krieger, E., Nabuurs, S.B. and Vriend, G. (2003). Homology modelling. Eds: Bourne, P.E. and Weissig, H. Structural bioinformatics. Wiley-Liss, Inc. Pp. 507-521.

Kumar, S. and Nussinov, R. (2001). How do thermophilic proteins deal with heat? A review. *Cellular and Molecular Life Sciences* **58**: 1216-1233.

Larkin, M.A., Blacksheilds, G., Brown, N.P., Chenna, R., McGettingan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007). ClustalW and ClustalX version 2.0. *Bioinformatics* **23**(21): 2947-2948.

Laurell, H., Contreras, J.A., Castan, I., Langin, D. and Holm, C. (2000). Analysis of psychrotolerant property of hormone-sensitive lipase through site-directed mutagenesis. *Protein Engineering* **13**(10): 711-717.

Lee, M.H., Lee, C.H., Oh, T.K., Song, J.K. (2006). Isolation and characterisation of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipase. *Applied Environmental Microbiology* **72**: 7406-7409.

Levisson, M., van der Oost, J. and Kengen, S.W.M. (2009). Carboxylic ester hydrolases from hyperthermophiles. *Extremophiles* **13**: 567-581.

Litthauer, D., Abbai, N.S., Piater, L.A. and van Heerden, E. (2010). Pitfalls using tributyrin agar screening to detect lipolytic activity in metagenomic studies. *African Journal of Biotechnology* **9**(27): 4282-4285.

Lovell, S.C., Davis, I.W., Arendall, W.B.III, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* **50**(3): 437-450.

Lynd, L.R. (1996). Overview and evaluation of fuel ethanol from cellulosic biomass: Technology, economics, the environment, and policy. *Annual Reviews on Energy and the Environment* **21**: 403-465.

Mabee, W.E. and Saddler, J.N. (2010). Bioethanol from lignocellulosics: Status and perpectives in Canada. *Bioresource Technology* **101**: 4806-4813.

MacKenzie, C.R. and Bilous, D. (1988). Ferulic acid esterase activity from *Schizophyllum commune. Applied Environmental Microbiology* **54**: 1170-1173.

Mackenzie, C.R., Bilous, D., Scheider, H. and Johnson, K.G. (1987). Induction of cellulolytic and xylanolytic enzyme systems in *Streptomyces spp*. *Applied and Environmental Microbiology* **53**: 2835-2839.

Madigan, M.T. and Mars, B.L. (1997). Extremophiles. *Scientific American* **April**: 66-71.

Malherbe, S., Cloete, T.E. (2002). Lignocellulose biodegradation: Fundamentals and applications. *Reviews in Environmental Science and Biotechnology* **1**: 105-114.

Manco, G., Adinolfi, E., Pisani, F.M., Ottolina, G., Carrea, G. and Rossi, M. (1998). Overexpression and properties of a new thermophilic and thermostable esterase from *Bacillus acidocaldarius* with sequence similarity to hormone-sensitive lipase subfamily. *Biochemical Journal* **332**: 203-212.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4): 404-405.

McMillan, J.D. (1996). Bioethanol production: Status and prospects. *Renewable Energy* **10**: 295-302.

Menten, L. and Michaelis, M.I. (1913). Die kinetic der invertinwirkung. *Biochemische Zeitschrift* **49**: 333-369.

Messaoudi, A., Belguith, H. and Hamida, J.B. (2011). Three-dimensional structure of *Arabidopsis thaliana* lipase predicted by homology modelling method. *Evolutionary Bioinformatics* **7**: 99-105.

Miller, S.C., LiPuma, J.J. and Parke, J.L. (2002). Culture based and non-growth dependent detection of the *Burkholderia cepacia* complex in soil environment. *Applied Environmental Microbiology* **68**(8): 3750-3758.

Morris, A.L., McArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**: 345-364.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M. and Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology* **96**: 673-686.

Moukouli, M., Topakas, E. and Christakopoulos, P. (2008). Cloning, characterisation and functional expression of an alkalitolerant type C feruloyl esterase from *Fusarium oxysporum*. *Applied Microbiology and Biotechnology* **79**: 245-254.

Naik, S.N., Goud, V.V., Rout, P.K. and Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* **14**: 578-597.

Nardini, M. and Dijkstra, B.W. (1999). α/β Hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology* **9**: 732-737.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G. and Goldman, A. (1992). The alpha/beta hydrolase fold. *Protein Engineering* **5**(3): 197-211.

Osmont, A., Catoire, L., Bocanegra, P.E., Gökalp, I., Thollas, B. and Kozinski, J.A. (2010). Second generation biofuel: Thermochemistry of glucose and fructose. *Combustion and Flame* **157**: 1230-1234.

Ostrander, E.A., Jong, P.M., Rine, J. and Duyk, G. (1992). Construction of small-insert genomic libraries highly enriched with microsatellite repeat sequences. *Proceedings of the National Academy of Science of United States of America* **89**: 3419-3423.

Palomo, J.M., Segura, R.L., Fernández-Lafuente, R. (2004). Purification, immobilisation and stabilisation of a lipase from *Bacillus thermocatenulatus* by interfacial adsorption on hydrophobic supports. *Biotechnology Progress* **20**: 630-635.

Panagiotou, G., Olavarria, R. and Olsson, L. (2007). *Penicillium brasilianum* as an enzyme factory; the essential role of feruloyl esterases for the hydrolysis of the plant cell wall. *Journal of Biotechnology* **130**: 219-228.

Polizeli, M.L.T.M., Rizzatti, A.C.S., Monti, R., Terenzi, H.F., Jorge, J.A. and Amorim, D.S. (2005). Xylanases from fungi: properties and industrial application. *Applied Microbiology Biotechnology* **67**: 577-591.

Prates, J.A.M., Tarbouriech, N., Charnock, S.J., Fontes, C.M.G.A., Ferreira, L.M.A. and Davids, G.J. (2001). The Structure of the feruloyl esterase module of xylanase 10B from *Clostridium thermocellum* provides insights into substrate recognition. *Structure* **9**: 1183-1190.

Price, C.P., Campbell, R.S. and Hammond, P.M. (1995). Novel enzymes as reagents. *Clinica Chimica Acta* **237**: 3-16.

Prim, N., Bofill, C., Pastor, F.I.J. and Diaz, P. (2006). Esterase EstA6 from *Pseudomonas* sp. CR-611 is a novel member in the utmost conserved cluster of family VI bacterial lipolytic enzymes. *Biochimie* **88**: 859-867.

Ranjitha, P., Karthy, E.S. and Mohankumar, A. (2009). Purification and partial characterisation of esterase from marine *Vibrio fischeri*. *Modern Applied Science* **3**(6): 73-82.

Record, E., Asther, M., Sigoillot, C., Pages, S., Punt, P.J., Delattre, M., Haon, M., van den Hondel, C.A., Sigoillot, J.C., Lesage-Meessen, L. and Asther, M. (2003). Over production of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Applied Microbiology and Biotechnology* **62**: 349-355.

Rhee, J.K., Ahm, D.G., Kim, Y.G. and Oh, J.W. (2005). New thermophilic and thermostable esterases with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Applied and Environmental Microbiology* **71**(2): 817-825.

Robinovich, M.L., Melnik, M.S. and Bolobova, A.V. (2002). Microbial cellulases: a review. *Applied Biochemistry and Microbiology* **38**(4): 305-321.

Rozzell, J.D. (1999). Commercial scale biocatalysis myths and realities. *Bioorganic and Medicinal Chemistry* **7**: 2253-2261.

Rumbold, K., Biely, P., Mastihubová, M., Gudelj, M., Gübitz, G., Robra, K-H. and Prior, B.A. (2003). Purification and properties of a feruloyl esterase involved in lignocellulose degradation by *Aureobasidium pullulans*. *Applied and Environmental Microbiology* **69** (9): 5622-5626.

Saha, B.C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology* **30**: 279-291.

Samad, M.Y.A., Razak, C.N.A., Salleh, A.B., Yunus, W.M.Z.W., Ampon,K. and Basri, M. (1989). A plate assay for primary screening of lipase activity. *Journal of Microbiological Methods* **9**: 51-56.

Sambrook, J., Fritsch, E.F., Cygler, M., Dijkstra, B., et al (1989). Molecular cloning: a laboratory manual. 2nd Ed. Cold Spring Harbour Laboratory Press, Cold Spring Habour.

Sanchez, O.J. and Cardona, C.A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology* **99**(13): 5270-5295.

Satpal, S.B. (2010). Thermophiles 2009. *Current Science* **98**(1): 19-20.

Schloss, P.D. and Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* **14**: 303-310.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* **31**(13): 3381-3385.

Shallom, D. and Shoham, Y. (2003). Microbial hemicellulases. *Current Opinion in Microbiology* **6**: 219-228.

Sharma, R., Christi, Y. and Banerjee, U.C. (2001). Production, purification, characterisation, and applications of lipases. *Biotechnology Advances* **19**: 627-662.

Sharma, R., Ranjan, R., Kapardar, R.K. and Grover, A. (2005). 'Unculturable' bacterial diversity: an untapped resource. *Current Science* **89**: 72-77.

Shin, H.D. and Chen, R.R. (2006). Production and characterisation of a type B feruloyl esterase from *Fusarium proliferatum* NRRL 26517. *Enzyme and Microbial Technology* **38**: 478-485.

Sigoillot, C., Camarero, S., Vidal, T., Record, E., Asther, M. Perez-Boada, M., Martinez, M.J., Sigoillot, J.C., Asther, M., Colom, J.F. and Martinez, A.T. (2005). Comparison of different fungal enzymes for bleaching high quality paper pulps. *Journal of Biotechnology* **115**: 333-343.

Simon, C. and Daniel, R. (2009). Achievements and new knowledge unravelled by metagenomics approaches. *Applied Microbiology and Biotechnology* **85**: 265-276.

Sommer, P., Georgieva, T. & Ahring, B. K. (2004). Potential for using thermophilic anaerobic bacteria for bioethanol production from hemicellulose. *Biochemical Society Transactions* **32:** 283-289.

Stuer, W., Jaeger, K.E. and Winkler, U.K. (1986). Purification of extracellular lipase from *Pseudomonas aeruginosa*. *Journal of Bacteriology* **168**: 1070-7074.

Vo, N.X.Q., Kang, H. and Park, J. (2007). Functional Metagenomics using stable isotope probing: a review. *Journal of Environmental Engineering* **12**(5): 231-237.

Shin, H-D. and Chen, R.R. (2006). Production and characterisation of a type B feruloyl esterase from *Fusarium proliferatum* NRRL 26517. *Enzyme and Microbial Technology* **38**: 478-485.

Shin, H-D. and Chen, R.R. (2007). A type B feruloyl esterase from *Aspergillus nidulans* with broad pH applicability. *Applied Microbiology and Biotechnology* **73**: 1323-1330.

Streit, W.R. and Schimtz, R.A. (2004). Metagenomics – the key to the uncultured microbes. *Current Opinion in Microbiology* **7**: 492-498.

Sun, Y. and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresources Technology* **83**: 1-11.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**(8): 1596-1599.

Tan, H.T., Lee, K.T. and Mohamed, A.R. (2010). Second-generation bio-ethanol (SGB) from Malaysian palm empty fruit bunch: Energy and energy analysis. *Bioresource Technology* **101**: 5719-5727.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **12** (14): 5627-5638.

Topakas, E. Stamatis, H. Biely, P. and Christakopoulos, P. (2004). Purification and characterisation of a type B feruloyl esterase (StFAE-A) from the thermophilic fungus *Sporotrichum thermophile. Applied Microbiology and Biotechnology* **63**: 686-690.

Topakas, E., Valiadi, C. and Christakopoulos, P. (2007). Microbial production, characterisation and application of feruloyl esterases. *Process Biochemistry* **42**: 497-509.

Tuffin, M., Anderson, D., Heath, C. and Cowan, D. (2009). Metagenomic gene discovery: how far have we moved into novel sequence space? *Biotechnology Journal* **4**: 1-13.

Uchiyama, T. and Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology* **20**: 616-622.

Uppenberg, J., Hansen, M.T., Patkar, S. and Jones, T.A. (1994). The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. *Structure* **15**: 293-308.

van den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Current Opinion in Microbiology* **6**: 213-218.

Vassilev, S.V., Baxter, D., Anderson, L.K. and Vassileva, C.G. (2010). An overview of the chemical composition of biomass. *Fuel* **89**: 913-933.

Vintila, T., Dragomirescu, M., Strava, S. and Croitoriu, V. (2009). Enzymatic hydrolysis of agricultural lignocellulosic biomass. *Zootehnie şi Biotechnologii* **42**(1): 125-129.

Virk, A.P., Sharma, P. and Capalash, N. (2011). A new esterase, belonging to hormone-sensitive lipase family, cloned from *Rheinheimera* sp. isolated from industrial effluent. *Journal of Microbiology and Biotechnology* **21**(7): 667-674.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E. and Streit, W.R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology* **69**(10): 6235-6242.

Wang, X., Geng, X., Egashira, Y. and Sanada, H. (2004). Purification and characterisation of a feruloyl esterase from the intestinal bacterium *Lactobacillus acidophilus*. *Applied and Environmental Microbiology* **70**(4): 2367-2372.

Weiss, J.N. (1997). The Hill equation revisited: uses and misuses. *The FASEB Journal* **11**: 835-841.

Wong, D.W.S. (2006). Feruloyl esterase: A key enzyme in biomass degradation. *Applied Biochemistry and Biotechnology* **133**: 87-113.

Wood, A.N.P., Fernandez-Lafuente, R. and Cowan, D.A. (1995). Purification and partial characterisation of a novel thermophilic carboxylesterase with high mesophilic specific activity. *Enzyme and Microbial Technology* **17**: 816-825.

Wood, T.M. (1991). Biosynthesis and biodegradation of cellulose. Marcel Dekker Inc., New York. Pp. 491-534.

Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. and Lee, Y.Y. (2005). Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology* **96**: 1959-1966.

Wyman, C.E. and Yang, B. (2009). Cellulosic biomass could help meet California's transport fuel needs. *California Agriculture* **63**(4): 185-190.

Yu, S., Zheng, B., Zhao, X. and Feng, Y. (2010). Gene cloning and characterisation of a novel thermophilic esterase from *Fervidobacterium nodosum* Rt17-B1. *Acta Biochimica et Biophysica Sinica*: 1-8.

Zaldivar, J., Nielsen, J. and Olssen, L (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering. *Applied Microbiology and Biotechnology* **56**: 17-34.

Zecca, A. and Chiari, L. (2010). Fossil-fuel constraints on global warming. *Energy Policy* **38**: 1-3.

Zhu, L., O'Dweyer, J.P., Chang, V.S., Granda, C.B. and Holtzapple, M.T. (2008). Structural features affecting biomass enzymatic digestibility. *Bioresource Technology* **99**(9): 3817-3828.