GENE DISCOVERY IN ANTARCTIC DRY VALLEY SOILS

by

Dominique Elizabeth Anderson

A thesis submitted in fulfillment of the requirements for the degree of

Magister Scientiae (M.Sc.)

Department of Biotechnology,

University of the Western Cape

Bellville

Supervisor: Professor D. A. Cowan

November 2008.

Table of contents

Acknowledgements

I would like to sincerely thank the following;

- The National Research Foundation for funding.
- Professor D. A. Cowan and Dr. C. Heath for supervision, support and encouragement.
- Dr. M. Mutondo, Dr. B. Kirby, Dr. M. Taylor and Dr. M. Tuffin who read this manuscript and gave valuable remarks and suggestions.
- IMBM and all its members for support and inspiration.
- Family and friends for always being there, through thick and thin, tears, lots of laughs and the occasional bottle of wine.
- Frank v. d. Berg for understanding, care and love.

Abstract

Gene discovery in Antarctic Dry Valley soils

Dominique E Anderson

MSc thesis, Department Biotechnology,
University of the Western Cape

The metagenomic approach to gene discovery circumvents conventional gene and gene product acquisition by exploiting the uncultured majority of microorganisms in the environment. It was demonstrated in this study that metagenomic methods are suitable for gene mining in extreme environments that harbor very high levels of unculturable microorganisms. DNA was extracted from Antarctic mineral soil samples taken from the Miers Valley, Antarctica. The metagenomic DNA was also used to construct a fosmid library comprising over 7900 clones with an average insert size of 29 kb. PCR amplification using bacterial and archaeal 16S rRNA gene specific primers and subsequent denaturing gradient gel electrophoresis (DGGE) of bacterial 16S rDNA amplicons showed that a small percentage of bacterial diversity (>1%) was captured in the metagenomic fosmid library. Activity-based screening for lipase and esterase genes using a tributyrin plate assay yielded twelve positive clones. LD1, a putative, novel cold-active GDSL lipase/esterase was identified and sequenced. The C-terminal domain of the ORF was found to be an autotransporter similar to those associated with type V secretion systems in Gram negative bacteria. Sub-cloning of the gene resulted in lipolytic activity in *E. coli.* Preliminary enzyme assays have determined that LD1 hydrolyses *p*-nitrophenyl esters with chain lengths shorter than C_{10}, an indication that the enzyme is an esterase. Complete purification and characterisation of this enzyme is subject to further study.

Keywords: Metagenomic DNA, fosmid library, functional screening, lipase, esterase, PCR, DGGE, diversity.

DECLARATION

I declare that *Gene discovery in Antarctic Dry Valley soils* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Dominique Elizabeth Anderson 14 November 2008

Abbreviations

CAPS	3-Cyclohexylamino-1-propanesulfonic acid
BrdU	5-Bromo-2-deoxyuridine
AU	Absorbance units
ATP	Adenosine triphosphate
APS	Ammonium persulphate
ARDRA	Amplified ribosomal DNA restriction analysis
AFPs	Antifreeze proteins
BAC	Bacterial artificial chromosome
BOD	Biological oxygen demand
BSA	Bovine serum albumen
BR	Broad range
C-terminus	Carboxy-terminus
$\times g$	Centrifugal force
CTAB	Cetyl trimethyl ammonium bromide
CAPs	Cold accumulatory proteins
CSPs	Cold shock proteins
cfu	Colony forming units
dH_2O	Demineralised water
DGGE	Denaturing gradient gel electrophoresis
dATP	Deoxy-adenine 5'-triphosphate
dCTP	Deoxy-cytosine 5'-triphosphate
dGTP	Deoxy-guanine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides
dTTP	Deoxy-thymine 5'-triphosphate
Des	Desaturases
et al	*et alia* (and others)
EtOH	Ethanol
EtBr	Ethidium bromide
EDTA	Ethylenediamine tetra-acetic acid
FACS	Fluorescence-activated cell sorting
GSP	General secretory pathway
GFP	Green fluorescence protein
HMW	High molecular weight
IPTG	Isopropyl-b-D-thiogalactopyranoside
kDa	Kilo Dalton
kV	Kilovolts
LB	Luri Bertani
LBA	Luri Bertani agar
µF	Micro Farad
Mbp	Million base pairs
MCS	Multiple cloning site
MVS	Miers Valley seal
TEMED	N,N,N',N'-Tetramethylethylenediamine
Ω	Ohm

ORF	Open reading frame
OTU	Operational taxonomic unit
OD	Optical density
PDB	Phage dilution buffer
*p*NP	*p*-nitrophenyl
RAPD	Random amplified polymorphic DNA
RISA	Ribosomal intergenic spacer analysis
RBS	Ribosome-binding site
sec	Second
SDS	Sodium dodecyl sulphate
SIGEX	Substrate induced gene expression
TGGE	Temperature gradient gel electrophoresis
T-RFLP	Terminal restriction fragment length polymorphism
TAE	Tris acetic acid EDTA
Tris	Tris-hydroxymethyl-aminomethane
UHQ	Ultra high quality
UTR	Untranslated region

List of Figures

List of Tables

Chapter 1: Literature review

1. 1 Antarctica

Antarctica broke away from the land mass Gondwana in the Cretaceous period approximately 120 million years ago (Peck *et al*., 2005). It separated from South America at the Eocene-Oligocene boundary 31 million years ago, at which time major cooling of the continent occurred. Geographic and environmental isolation coupled with a long existing extreme environment and strong selective pressures, is believed to have lead to the unique and highly adapted biota of Antarctic sea and land (Peck *et al*., 2005). Antarctica is the southernmost continent on earth, covering an area of 14, 2 million square kilometres. Most of this remote continent is covered by an expansive ice sheet and is the windiest and driest land mass known (Balks, 2001; Peck *et al*., 2005).

The continent harbours unique and diverse terrestrial and aquatic habitats, each with varying climatic conditions which impact directly on the diversity of life forms that exist (Wynn-Williams, 1996). Temperatures in winter range from -40 ^{o}C to -60 ^{o}C in the interior and -20 ^{o}C to -30^{o}C at the coastal regions with wind chill being a major contributor to the cold temperatures (Balks, 2001). The atmosphere contains very low levels of water vapour due to the cold temperatures and low precipitation on the continent, ultimately transforming it into a cold desert (Balks, 2001). The Antarctic continent is both physically and chemically demanding but is by no means devoid of life (Hogg *et al*., 2006)

The highest trophic level of non-migratory, endemic organisms in Antarctica includes the invertebrates such as nematodes and arthropods [mites and springtails] (Adams *et al*., 2006). Microbes are the dominant biomass of ecosystems in Antarctica (Vincent, 2000). Microbial evolution in the region has been influenced by the absence of gene flow from outside biota, as well as the extremes of environment. In Antarctic ecosystems, abiotic factors have a greater influence on the biota than biotic factors do (Hogg *et al*., 2006).

1.1.1 The Antarctic Dry Valley terrestrial environment

Ice free areas are scattered around the margins of the Antarctic continent and account for only 0.3% of the Antarctic land mass (Balks, 2001). The Dry Valleys consists of exposed soils (mineral soils being barren and with little water content, and moist soils with higher levels of moisture), glaciers, streams and lakes (both freshwater and saline) and permanently ice-covered lakes and are by no means homogenous (Cowan *et al*., 2004). In the summer months, air temperature averages 0 ^{o}C while the ground temperatures may increase to as much as 15 ^{o}C (Balks, 2001).

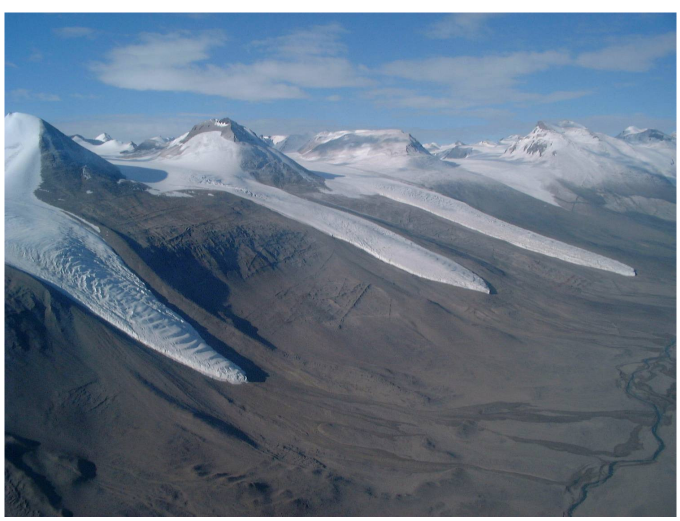


Figure 1 Aerial photograph of the Wright Valley, Antarctica.

The Dry Valley deserts lack a layer of topsoil and although processes such as leaching and weathering occur at extremely low rates, the soils are very susceptible to physical disturbances (Balks, 2001, Hogg *et al*., 2006). The upper level is referred to as the 'pavement' and is dominated by fine gravel, stones and boulders. Subsurface layers include an active layer that undergoes regular freeze-thaw cycles and permafrost, a mixture of permanently frozen ice and soil (Balks, 2001). Salts accumulate in the soils causing high salinity and the pH of soils in the Dry Valley area ranges from weakly acidic (pH 6-6.5) to alkaline (pH 7-9) in the coastal regions. This fluctuation in pH is mainly due to the low buffering capacity of the soils (Balks, 2001). Organic accumulation of carbon and nitrogen is low and a limiting factor in microbial populations that inhabit the Valley soils. Additionally, strong, low-humidity kabatic winds frequently sweep across the Valley floor (Balks, 2001). These factors, linked with solar radiation and low levels of precipitation and moisture strongly influence the physical, biological, ecological and chemical properties of the soil and thus the communities that inhabit it (Balks, 2001; Aislabie *et al*., 2006).

Figure 2 The 'pavement' structure of soils in the Dry Valleys, Antarctica.

1.1.2 Microbial diversity of the Dry Valleys

Due to the hyper-arid nature of the Dry Valleys it was, at one time, suggested that the soils were sterile and unable to support any life (Boyd, 1962; Cowan *et al*., 2002; Hogg *et al*., 2006). However, pioneering work in earlier decades proved that life did indeed occur in the cold desert soils (Ekelöf, 1908; Boyd, 1962). Due to the drawbacks of culture methods in these studies, a complete record of microbial community structure could not be provided. More recent studies on microbial diversity in the Dry Valleys have used culture independent approaches such as 16S rRNA gene amplification by PCR (Franzmann, 1995; Aislabie *et al*., 2006; Smith *et al*., 2006; Shravage *et al*., 2007).

Mosses, lichens and yeasts are often isolated from Dry Valley soils (Wynn-Williams, 1996; Adams *et al*., 2006) and mineral soils in the area contain both cosmopolitan and indigenous protozoa, fungi (eg. *Aspergillis*, *Penicillium*, *Alternaria*) and yeasts (eg. *Candida*) (Cowan *et al*., 2004; Adams *et al*., 2006). Stress resistant or dormant phases may offer protection against freezing and moisture absence to some groups such as lichens, algae, nematodes and bacteria. Protozoa in soils include flagellates, small amoeba and ciliates. Nematodes, rotifers and tardigrades have been isolated from Dry Valley soils but their distribution is irregular and generally limited to areas of higher soil moisture (Franzmann, 1995; Wynn-Williams, 1996; Adams *et al*., 2006;). Bacteria isolated and identified in Dry Valley soil belong to a wide range of genera, including *Arthrobacter*, *Bacillus*, *Bacteroidetes*, *Corynebacterium*, *Cytophaga*, *Flavobacterium*, *Micrococcus*, *Planococcus*, *Pseudomonas*, *Psychrobacter* and *Streptomyces* (Cowan *et al*., 2004; Adams *et al*., 2006; Smith *et al*., 2006; Shravage *et al*., 2007). Phylogenetic analysis of Dry Valley soils has shown that a large proportion of the bacterial diversity falls into the 'uncultured' class and may therefore represent a large pool of novel genera or species (Smith *et al*., 2006).

The biota of fresh water systems includes microalgal mats, diatoms and cyanobacteria. Cosmopolitan bacteriovorous flagellates, rhizopods, ciliates and heliozoans are also found along with grazing metazoa such as Rotifers and tardigrades (Wynn-Williams, 1996; Adams *et al*., 2006). Bacteria, yeasts, filamentous fungi and microalgae are the epiphytic communities associated with mosses found on the margins of these flowing water systems (Cowan *et al*., 2004; Adams *et al*., 2006). Lake communities are dominated by algal communities and microbial mats of filamentous cyanobacteria, bacteria, fungi and protozoa. Salinity, temperature and stratification are important factors contributing to the microbial ecology of Antarctic lakes (Cowan *et al*., 2004).

Lithic communities occur within the porous and translucent rocks of the Dry Valleys. These rocks offer a suitable habitat that buffers extreme temperature and humidity fluctuations (Cowan *et al*., 2004). Chasmoendolithic communities inhabit cracks in weathering rocks and consist of lichen and cyanobacterial associations. The cryptoendoliths, consisting largely of lichens and cyanobacteria, inhabit interstices of crystalline rock (Wynn-Williams, 1995; Cowan *et al*., 2002; Adams *et al*., 2006).

1.1.2.1 Seal Falls

The first sightings of mummified seal carcasses in the Dry Valleys of Antarctica dates back over a hundred years ago during Captain R. F. Scott's first expedition to the continent in 1901 (Barwick *et al*., 1967; Robson, Cowan, Cary, unpublished). Disoriented seals, the majority of which were identified as immature crabeater seals, wonder into the Dry Valley and die of starvation, dehydration and exhaustion (Barwick *et al*., 1967; Dort, 1982). The hyper-arid conditions that are prevalent in the Antarctic essentially mummify seal carcasses, which become eroded by high-speed wind-blown sands (Barwick *et al*., 1967). Remains of the carcasses range from complete, with little damage, to minimal scattered fragments. According to carbon dating, the ages of some of these carcasses range from 100 to 2000 years (Barwick *et al*., 1967). The rate of entry of seals into the Dry Valley was estimated at one every 4 to 8 years (Barwick *et al*., 1967). These carcasses influence the ecosystem dynamics in the Dry Valley soils by contributing carbon and nitrogen to an otherwise depleted pool. The carcasses not only contribute a substantial pool of organic nutrient to microbial communities directly beneath or in close proximity to the carcass, but may also protect soil microbes from desiccating winds and ultraviolet exposure (Hopkins *et al*., 2001; Robson, Cowan, Cary, unpublished).

The major bacterial classes found in soils beneath the seal carcasses include the Actinobacteria, Bacilli and the γ-Proteobacteria whereas the majority of microbes in the open Dry Valley soils fall into the 'uncultured' category, thereby indicating that bacterial diversity found beneath the seal carcasses differs substantially from those found in the open soils (Robson, Cowan, Cary, unpublished; Smith *et al*., 2006). The microbial communities may also exhibit valuable nutrient utilising capabilities including cold-adapted enzymes for utilisation of the seal derived substrates (Robson, Cowan, Cary, unpublished).

Figure 3 Mummified seal carcass in the Miers Valley, Eastern Antarctica. Damage due to wind blown sands occurs on the exposed surface of the carcass (photograph by D. A. Cowan).

1.2 Microbial biodiversity

Microorganisms are the most numerous organisms on earth and occupy every available niche on the planet. They are the driving force of ecosystem processes and perform many key functions including fundamental biogeochemical cycling of nutrients, energy flow, carbon sequestration, and the buffering and transformation of potentially harmful compounds and elements (Prosser, 2002; Ritz *et al.*, 2003). They are the foundation of food webs and are valuable environmental monitoring agents with respect to global ecosystem changes (Prosser, 2002; Ritz *et al.*, 2003; Singh *et al.*, 2006).

In soils, more than 10^9 bacteria per gram, with approximate biomass of 3000 kg per hectare, are supported (Ranjard *et al*., 2001). Soils therefore act as reservoirs of biodiversity in terms of physiology, metabolism and phylogeny (Hunter-Cevera, 1998). Population diversity and heterogeneity in soils is an integral part of ecosystem function. The importance of this biodiversity in soils is illustrated by the level of abundance and community structure (Hunter-Cevera, 1998; Coleman, 2005).The microbial biomass is implicated in soil structure and dynamics such as water retention, colour, texture and even the smell. Microorganisms are reservoirs of nitrogen, phosphorous, and sulphur (Ranjard *et al*., 2000; Targulian, 2004). Microorganisms respond quickly to changes in environmental conditions due to optimized biochemical and genetic regulation, ultimately resulting in metabolic and physiological alterations (Hunter-Cevera, 1998; Brookes, 2001; Ritz *et al*., 2003).

Microhabitats in soils are dynamic systems with chemical, physical and biological characteristics that differ in temporal and spatial dimensions (Nannipieri *et al*., 2003). Microbial activity in soil is a result of complex interactions in communities with different types, numbers and ratios of individual members (Hunter-Cevera, 1998; Brookes, 2001). These biotic interactions are susceptible to physical and chemical changes in habitats as well as the metabolic activities and physiology of the microorganisms themselves (Hunter-Cevera, 1998; Griffiths *et al*., 2000).

Microbial biodiversity encompasses three interrelated elements; genetic, phenotypic or taxonomic, and functional (Ritz *et al*., 2003). It can be defined as the hereditary on all levels from gene variability within species to local communities, including species richness and species abundance, and finally to the living ecosystems of the world (Hunter-Cevera, 1998; Torsvik *et al*., 1998; Wilson, 1997).

Biodiversity analysis in ecology is not only important for the conservation of microbial gene pools but also for linking diversity, ecosystem processes, physiology, function and decreases in environmental resilience, due to the loss of species with similar functional attributes (Griffiths *et al*., 2000; Prosser, 2002). Only 27 of the 53 bacterial phyla have to date been cultivated and described in pure culture (Coleman *et al*., 2005) and this is one of the main reasons why diversity and function of soil microbial communities is not well understood (Pace, 1997). Conflicting definitions of classification, taxonomic allocation and species concepts have also hampered and limited research output (Bohannan, 2003).

1.2.1 <u>Genetic analysis</u>

The application of molecular genetic techniques related to the DNA composition of the community are used to advance understanding of soil microbial community structure, determine phylogenetic relationships between organisms, track individual species and their dynamics and compare communities from different habitats (Ritz *et al*., 2003). Such methods are mostly sequence-based and involve the amplification and/or analysis of genes, thereby eliminating the limitations associated with culture dependant techniques (Prosser, 2002). Broad scale analysis of total community DNA extracted directly from the environment can be used to assess genetic diversity of the microbial population (Torsvik *et al*., 2002). An overview of methods used to assess microbial diversity is given in Figure 4.

1.2.1.1 Low resolution methods

Base composition and a shift in GC content is one of the properties of DNA that can be exploited to detect changes in microbial community structure. Similar % GC profiles may indicate related organisms but is by no means a confirmation of relationship. Differences in base composition do however provide evidence for a lack of relation (Øvreås, 2000; Torsvik *et al*., 2002; Nannipieri *et al*., 2003). Melting curves of DNA are used to determine the GC content based on the thermal denaturation of DNA. Melting curves of complex communities consist of many different melting points over a wide range of temperatures (Øvreås, 2000). Similarly, reassociation of single-stranded DNA can also be used to detect differences in diversity of the total community. High homology results in a faster reassociation rate (following second order kinetics) and is based on the variety of sequences present (Øvreås, 2000).

1.2.1.2 High resolution methods

Fingerprinting methods are based on the polymerase chain reaction (PCR) and provide a means to measure diversity and diversity changes within the whole community at a high resolution (Torsvik *et al*., 2002). Population DNA or RNA can be extracted from the sample and amplified using cycles of nucleic acid denaturation, primer binding and template elongation, using universal primers of the 16S rRNA genes of bacteria and archaea (Coleman *et al*., 2005). Amplification sequences of coding and/or non-coding regions give a community fingerprint which can be used to distinguish families, genera or even species, dependant on their sensitivity (Nannipieri, *et al* 2003).

a. Methods using the electrophoretic separation properties of amplified products

Denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) are techniques that utilize the electrophoretic properties of amplified products. Linear gradients of temperature or denaturing agents are created in vertical polyacrylamide gels and the differences in base composition of amplified products are reflected in generated profiles (Ranjard *et al*., 2000). During PCR, GC rich clamps (long terminal extensions of G and C residues) are included at the end of one primer to prevent complete denaturation of DNA into separate strands (Ranjard *et al*., 2000). Amplified DNA fragments of identical or near identical length will partially denature at different denaturant concentration or temperature profiles depending on the GC content of the sequences (Ranjard *et al*., 2000). Secondary analysis of individual bands may be performed after excision from the gel. DNA may be transferred to nylon membranes and probed with group- or species- specific oligonucleotides, or subjected to another PCR reaction and sequencing (Theron *et al*., 2000).

b. Methods that analyse length polymorphisms

Ribosomal intergenic spacer analysis (RISA) is an rRNA based method that involves analysis of length polymorphisms in the intergenic spacer region; the spacer between the small 16S and large 23S subunit of the rRNA genes (Øvreås, 2000; Ranjard *et al*., 2000). This region has a variable size of 50 bp to 1.5 kb, depending on the species of the organism (Ranjard *et al*., 2000). Primers target the conserved regions in the rrs (small ribosomal subunit) and rrl (large ribosomal subunit) genes and the PCR products have significant heterogeneity in length and nucleotide sequence when separated on polyacrylamide gels, thus creating specific community profiles (Øvreås, 2000). Taxonomic identification of populations in the community can then be obtained by excision and subsequent sequencing of particular bands (Ranjard *et al*., 2000).

c. <u>Methods using random primers</u>

The random amplified polymorphic DNA (RAPD) technique makes use of short, random sequences as primers in the PCR (Ranjard *et al*., 2000). These primers anneal at different sites on the DNA and products of various lengths are resolved on polyacrylamide gels. Primer design does not rely on previous knowledge of the genome, as is the case with other methods (Ranjard *et al*., 2000; Nannipieri *et al*., 2003).

d. <u>Methods using restriction digestion of amplified products</u>

Amplified ribosomal DNA restriction analysis (ARDRA) and terminal restriction fragment length polymorphism (T-RFLP) are techniques based on restriction enzyme digests of the PCR amplified products (Dahllöf, 2002). rRNA gene fragments are amplified using conserved primers and subjected to various restriction enzymes in order to detect small differences at the nucleotide level (Theron *et al*., 2000; Prosser, 2002). In T-RFLP, one primer is fluorescently labelled and fragments are resolved by polyacrylamide gel electrophoresis. T-RFLP gives only one band consisting of the fragment with the label, while ARDRA analysis produces multiple bands for a single species (Dahllöf, 2002). Each unique band in the resulting fingerprint is considered to be an operational taxonomic unit and the frequency of occurrence of each one can be calculated (Øvreås 2000; Ranjard *et al*., 2000).

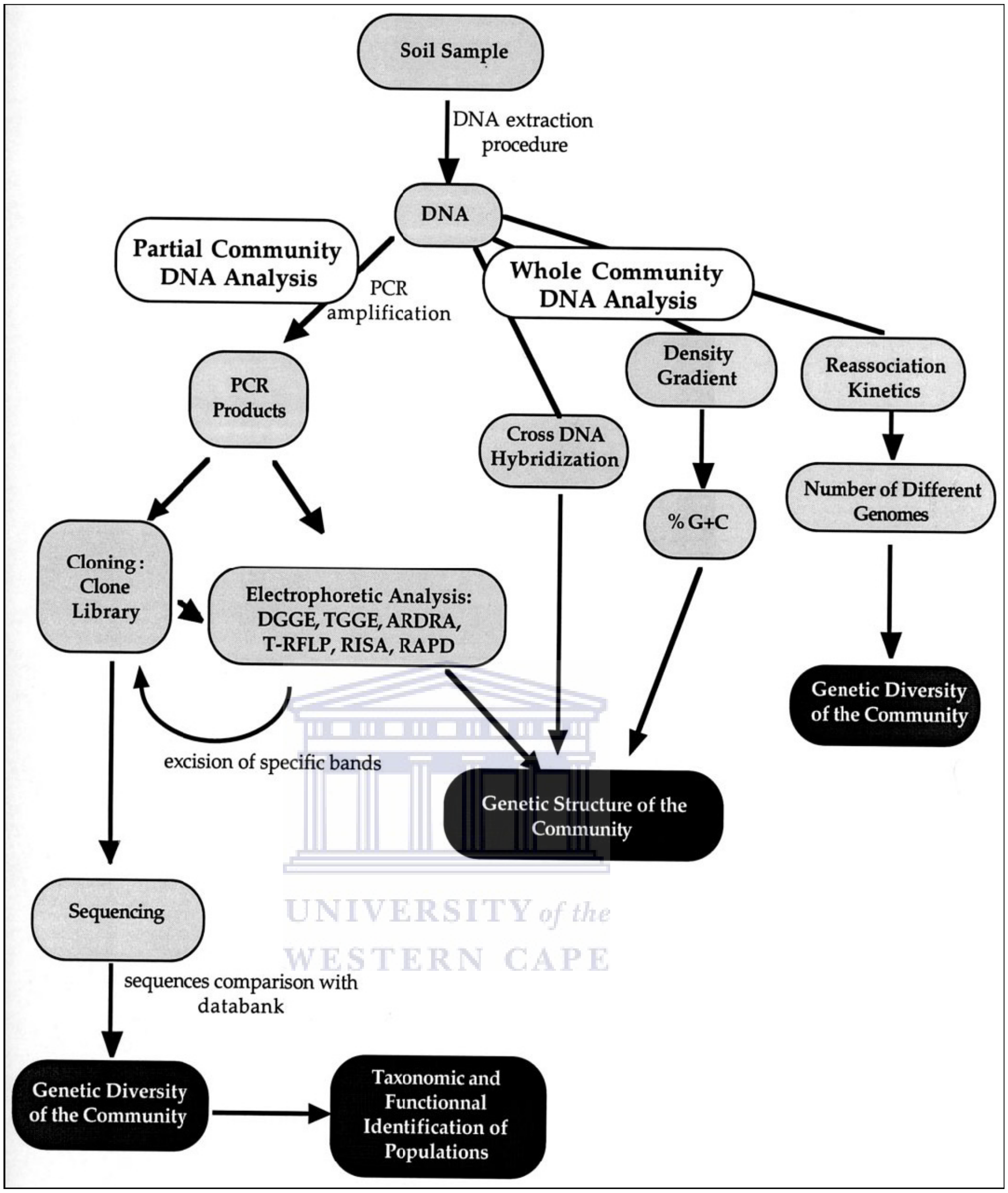


Figure 4 Summary of molecular methods used to assess genetic microbial biodiversity in soils. Taken from Ranjard *et al*., 2000.

1.3 Metagenomics and gene discovery

The biosphere is dominated by microorganisms which play central roles in geochemical and biological systems (Rodriguez-Valera, 2004; Xu, 2006). Prokaryotic life has been subjected to evolutionary pressure since its emergence 3.8 billion years ago, a factor that has contributed to the extreme heterogeneity of the microbial world (Xu, 2006). Molecular analysis has revealed that microbial diversity is more complicated than previously imagined (Béjà 2004). It is accepted that at least 99% of microbial species are presently still uncultured and the prokaryotic world is considered to be a dynamic pool rich in biodiversity with matching diversity in yet undiscovered compounds and metabolic pathways (Gillespie *et al*., 2002; Schloss *et al*., 2003; Lorenz *et al*., 2005; Schmeisser *et al*., 2007). Many microbes have remained recalcitrant to culturing due to strict physicochemical requirements (such as pH, temperature, nutrient requirements, salinity etc), and interdependence with other organisms (Lorenz *et al*., 2005). Since complex microbial interactions do not occur in pure culture, bacteria that inhabit complex consortia are difficult to reconstitute *in vitro* [even if multiple organisms are studied simultaneously in liquid enrichments where substrates and products diffuse freely among members] (Schloss *et al*., 2003; Cowan *et al*., 2004; Langer *et al*., 2006).

The term metagenomics was first coined in 1998 (Handelsman, 2004). Metagenomics focuses on the entire genetic complement of microbes in habitats or niches and is based on the extraction of total community DNA, and the genomic analysis of the data obtained (Cowan *et al*., 2004; Schmeisser *et al*., 2007). As metagenomic methods are DNA based, the need for culturing is removed and the bias associated with culture methods is reduced.

1.3.1 Metagenomic technology

Various methodologies are used in metagenomic studies and the choice of appropriate strategy is based on community complexity, research resources and goals, the amount of sample material available, microbial density and the nature of the substrate (Kowalchuk *et al*., 2007). An overview of metagenomic library construction is provided in Figure 5.

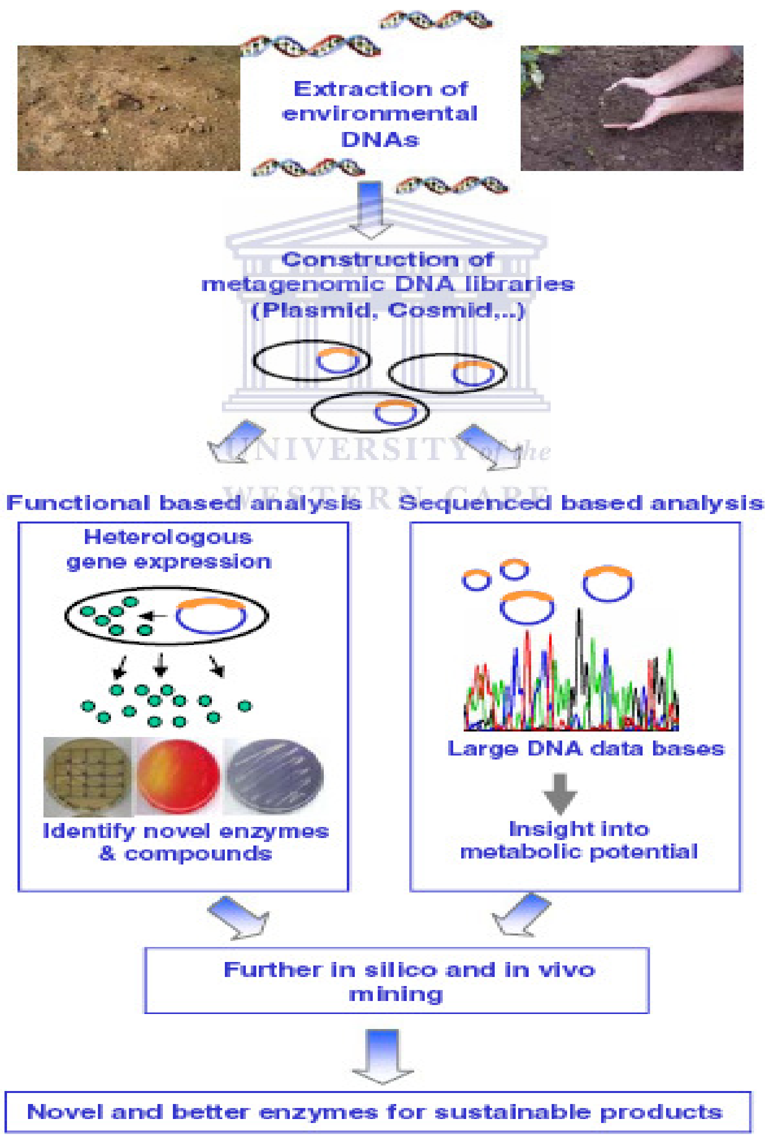


Figure 5 Overview of metagenomic library construction and the metagenomic approach to gene discovery. Taken from Schmeisser *et al*., 2007.

The construction of a metagenomic library starts with the extraction of total community DNA from a sample. The quality of a metagenomic library has a proportional relationship to the quality and clonability of the DNA obtained (Green, 2006). The purity and yield of DNA obtained from cell lysis affects downstream techniques such as PCR and cloning (Krsek, 1999). Polyphenolic compounds and other contaminants co-purified with DNA are not only difficult to remove but may negatively affect cloning efficiency and other downstream molecular techniques (such as PCR) due to interference of contaminating molecules with enzymatic steps (Daniel, 2004; Streit *et al*., 2004).

Three key issues need to be considered when extracting DNA from environmental samples, particularly when the DNA is used to construct large insert libraries; DNA must not be sheared as chimeric products are likely to be formed with smaller DNA fragments, it must be representative of a broad range of microbes so that the best genomic representation of a community or environment is obtained, and contaminating substances which may interfere with downstream processes must be absent (Schmeisser *et al*., 2007). The choice of a method for DNA extraction is based on the type of sample and the purpose of the metagenomic study. No single method of cell lysis and DNA extraction is applicable to all samples. Lysis may be mechanical, using methods such as bead beating or sonication; chemically, using various detergents and enzymes; or a combination of both treatments (Krsek, 1999).

There are two widely used methods of DNA extraction; the *ex situ* approach, whereby cells are extracted from the medium, concentrated and then lysed, is used for obtaining large DNA fragments (Krsek, 1999; Lorenz, 2002) and the *in situ* approach, where cell lysis occurs within the sample and this method is favoured smaller DNA fragments are required (Krsek, 1999; Lorenz, 2002).

In the total nucleic acid portion of the metagenome, target genes may represent a small fraction and pre-enrichment for genes of interest before lysis and DNA extraction can increase the screening hit-rate (Rappé, 2003; Cowan *et al*., 2005). Enrichment strategies generally involve the use of specific substrates in order to select for phenotypes of interest (Steele *et al*., 2005). Another method, which would increase the representation of certain genomes in the library, is via enrichment based on G+C content. Separation by BrdU enrichment is based on the selection of cells with the incorporated label in their DNA. Immunocapture is then used to recover the DNA. Using BrdU in conjunction with selective substrates increases the enrichment. Another method is the use of stable isotope labelling using compounds such as ^{13}C (Schloss *et al*., 2003). Enrichment can, however, lead to a decrease in diversity and changes in community structure leading to more biased libraries being created (Elend *et al*., 2006). Enrichment for selected common features allows researchers to obtain complete coverage of a subset of the environmental community, but genes and gene products that have optimal activity outside the range of enrichment conditions may be excluded (Schloss *et al*., 2003; Elend *et al*., 2006).

After fragmentation (either by mechanical or enzymatic methods) and purification of environmental DNA, the desired fractions are cloned into the appropriate vector system (Eyers *et al*., 2004).

Due to the large size of metagenomes and the need for efficient coverage, effective cloning strategies and high cloning efficiencies are required. Vector systems that are utilised in metagenomic studies include plasmids (Boubakri *et al*., 2006; Lämmle *et al*., 2006; Lee *et al*., 2004), fosmids (Hårdeman *et al*., 2006; Treusch *et al*., 2004), bacterial artificial chromosomes (Béjà *et al*., 2000; Rondon *et al*., 2000), and yeast artificial chromosomes (Béjà, 2004). Plasmid vector systems are generally employed for cloning small inserts and single genes or small operons are targeted (Rondon *et al*., 1999; Lorenz, 2002; Ward, 2006). Transcriptional promoters of the vector system are generally used when small inserts are cloned, and a start codon and upstream RBS (ribosome binding site) are usually supplied close to the MCS (multiple cloning site) (Gabor, 2004; Lorenz *et al*., 2005). Vectors carrying their own promoters and terminators are employed to ensure formation of full-length mRNA transcripts (Gabor, 2004). When these high copy number vectors are utilized in *E. coli*, cellular toxicity of expressed genes may be limiting (Ferrer *et al*., 2005).

Whether a microbial population is species rich or species poor, there is difficulty in obtaining DNA from all members of the community due to uneven species representation. Construction of large insert libraries can provide a more accurate representation of community members due to a greater genomic coverage of the community (Handelsman, 2005). These large insert libraries are used to target single genes and primary gene products, as well as secondary metabolites from the expression of complete operons (Rondon *et al*., 1999; Lorenz, 2002). Expression of genes encoded in large inserts depends on the presence of intrinsic promoter elements and transcriptional motifs of the original donor organism (Lorenz *et al*., 2005). The presence of a transcriptional promoter and a ribosome binding site (rbs) upstream of the start codon in the -20 to -1 region (9bp spacing in *E. coli*) is the minimal set of requirements for translation initiation.

Trans-acting elements, such as inducers, co-factors, chaperones and proper secretion machinery may also be required for formation of active protein and these must be provided by the host cell. For host transcription, translation and modification machinery to act upon foreign DNA, compatibility must exist with the transgenic genomic material (Gabor, 2004; Kowalchuk *et al*., 2007).

Bacterial artificial chromosomes (BACs) and fosmids (F1 origin based cosmid vectors) were introduced into the field of genomics in 1992 (Handelsman, 2005). These modified plasmids contain an origin of replication derived from the *E. coli* F factor and are the vector systems of choice for constructing large insert environmental libraries (Béjà et al., 2004). Replication of BAC and fosmid vectors is strictly controlled and maintained at very low copy number in heterologous host cells. BAC libraries require a greater amount of starting material, and minimized shearing of nucleic acids during extraction of high molecular weight DNA is essential (Béjà et al., 2004). Unlike the fosmid system, there is no selection against the cloning of small fragments and the inserts may range from 5 to 200 kb. Fosmid libraries can be constructed from a smaller pool of starting material and inserts are of uniform size up to 40 kb (Béjà, 2004).

Once vectors have been transformed into surrogate host cells, the resultant genomic libraries can be screened for novel genes and/or gene products (Eyers *et al*., 2004).

Direct cloning is, however, still hampered by inefficient heterologous expression of foreign genes in host strains (Gabor *et al*., 2007). The efficiency of gene expression in surrogate hosts is dependent on the presence of full-length genes, the recognition of expression signals and cis and trans acting sequences and post translational modification (Ferrer *et al*., 2005; Gabor *et al*., 2004). Recombinant proteins may not fold correctly due to the absence of appropriate chaperones in host strains. Furthermore, heterologous host strains may not synthesize essential co-factors, or the gene product may be toxic. Differences in codon usage may also contribute to low levels of protein expression and stability (Streit *et al*., 2004; Ferrer *et al*., 2005; Ferrer *et al*., 2007; Gabor *et al*., 2007; Lämmle *et al*., 2007).

The three most common strategies for the screening of metagenomic libraries include homology-based screening (requires sequence data in order to target genes), activity-based screening (clones that express desired functions are selected for), and substrate-induced screening (based on the principle that expression of catabolic genes is induced by substrate availability). Each strategy has advantages and disadvantages, but all have the potential for isolating genes and/ or gene products of interest (Lorenz, 2002; Daniel, 2005; Ferrer *et al*., 2005).

Screening by hybridisation or PCR can be used to detect genes homologous to those of a known sequence (Handelsman, 2005). This sequence-based screening method is also a powerful tool for the identification of gene function and putative roles of microorganisms in communities and has also driven demand for the development of new software and bioinformatics tools. However, product discovery is limited to previously described gene families and this method does not allow for the detection of truly novel genes and products (Schmeisser *et al*., 2007).

Biotechnological studies mostly make use of function-based screens as they have more potential in identification of entirely novel catalysts (Langer *et al*., 2006). In product driven studies, environments likely to contain the genes of interest should be selected. Activity based screening relies on gene expression in heterologous hosts as well as quick, easy and reliable protocols for the detection of function (Schmeisser *et al*., 2007).

Substrate induced gene expression (SIGEX) is a high throughput screening technique based on the knowledge that expression of catabolic genes is induced by substrate or metabolites of enzymatic breakdown of a substrate (Uchiyama, 2005). Operon-trap green fluorescence protein (gfp) expression vectors are used for cloning and fluorescence-activated cell sorting (FACS) is used to select for positive clones that express green fluorescence protein (GFP) when the target substrate is present (Handelsman, 2005; Uchiyama, 2005). Clones containing self-ligated plasmids are removed from the selection process by FACS after IPTG induction, thereby eliminating false positives and improving the possibility of selecting positive clones for further analysis (Uchiyama, 2005).

1.3.2 Gene discovery

There is a long tradition of human exploitation of microorganisms and their products e.g. baking, brewing, current food and feed processes, detergents etc (Lorenz *et al*., 2005). Biocatalysts obtained from natural resources have a number of advantages over their chemical counterparts such as, substrate specificity [which reduces the accumulation of large amounts of by-products], biodegradability, and ability to increase the sustainability of a process and deliver a cost advantage (Langer *et al*., 2006).

Past strategies for utilizing microbial diversity have relied on the ability to culture the organisms. However, the numbers of useful products being discovered is decreasing and commercial product development is impaired by a lack of suitable biocatalysts, not substrate (Lämmle *et al*., 2007; Langer *et al*., 2006; Schmeisser *et al*., 2007).

Established microbial collections from both common and extreme environments represent only a minor fraction of the microbial diversity and the challenge in current 'white' biotechnology is exploration beyond these culture collections (Lorenz, 2002). Novel cultivation techniques in which natural environments are simulated may offer temporary solutions but ultimately cultivation independent techniques are required to expand product discovery. Political pressure to increase sustainability, to reduce environmental impact and minimise resource consumption, is creating a demand for novel enzymes and biocatalysts (Lorenz *et al*., 2005).

The analysis of the genomes of uncultured microbes can not only provide a better understanding of global microbial ecology, but can also drive the supply of novel biocatalysts and biomolecules (Schmeisser *et al.*, 2007). The focus of many metagenomic research endeavours is the bioprospecting of novel products and metagenomics has become a powerful tool for the discovery of unknown and improved gene products that may be exploited for biotechnological and biomedical purposes (Cowan *et al*., 2004). Novel products and pathways already uncovered from a number of diverse environmental niches include enzymes; such as oxidoreductases (Knietsch *et al*., 2003), esterases (Elend *et al*., 2006) and lipases (Lee *et al*., 2004), hydratases (Liebeton, 2004; Ferrer, 2005) and alcohol dehydrogenases (Wexler, 2005); antibiotics, such as turbomycin (Gillespie *et al*., 2002), and even novel pathways for the degradation of xenobiotics (Boubakri *et al*., 2006; Eyers *et al*., 2004).

1.4 Cold adaptation

The ability of microorganisms to adapt to natural stress factors in environments has made them the Earth's most successful colonisers. Approximately 80% of our planets biosphere is below 5°C, with the polar regions representing at least 14% of permanently cold terrestrial and aquatic environments (Hébraud *et al*., 1999; Rodrigues *et al*., 2002). A wide range of adaptive strategies have been adopted by microbes in order to maintain vital cellular functions at cold temperatures. Physico-chemical properties vary in cold environments and the survival strategies used by organisms in each habitat will differ. Extremophiles generally encounter more than one stress factor in a cold habitat, such as desiccation, high or low pH, high osmotic pressure and low nutrient availability (Morgan-Kiss *et al*., 2006; Tehei *et al*., 2005).

1.4.1 <u>Low temperature adaptive strategies</u>

1.4.1.1 Membrane lipid composition

Microorganisms adjust the unsaturated fatty acid composition in their membranes according to changes in environmental temperatures. Alteration of lipid content is not a strategy used strictly by psychrotolerant bacteria and is employed by mesophilic and thermophilic organisms (Ray *et al*., 1998). However, the increased rate at which these changes occur is of great importance in habitats where thermal fluctuations occur (Hébraud *et al*., 1999). A decrease in temperature is generally accompanied by an increased ratio of polyunsaturated fatty acids, thereby reducing the phospholipids melting point and avoiding inflexibility of membrane structures (Nichols *et al*., 1993; Ulusu *et al*., 2001). The family of enzymes called desaturases (Des) is responsible for the introduction of double bonds into fatty acids. This occurs via an aerobic desaturation pathway which can occur independent of cell growth (Morgan-Kiss *et al*., 2006).

Other possible modulators of membrane fluidity have been proposed, particularly carotenoid pigment molecules that are associated with cell membranes. These pigments may buffer membrane fluidity and maintain homeoviscosity during temperature fluctuations (Chattopadhyay, 2006; Ray *et al*., 1998; Rodrigues *et al*., 2008).

1.4.1.2 The cold shock response

Cold shock response is induced when an organism is subjected to sub-optimal growth temperatures. The molecular mechanism involved in cold shock response has been extensively studied in the mesophilic bacterium, *E. coli*. In this organism, a rapid shift in growth temperature from 37°C to 10°C induces transient expression of a number of genes (Chattopadhyay, 2006; Horn *et al*., 2007; Ray *et al*., 1998). The gene products may either be directly or indirectly involved in protein transcription and translation (Horn *et al*., 2007). Expression of other genes is suppressed and the physiological growth declines until the organism is adapted to the new environment (Hébraud *et al*., 1999; Horn *et al*., 2007).

Cold shock proteins (CSPs) are a highly conserved family of single-stranded nucleotide binding proteins. In *E. coli*, 9 members of CSPs have been identified and CspA is the major protein involved in this response (Ray *et al*., 1998). Under cold stress the mRNA encoding CspA is stabilised and its expression is favoured. CspA can up-regulate its own transcription as well as that of other CSPs, by binding to the 5' UTR of Csp mRNA's (Horn *et al*., 2007). This enhances the half-life of RNA and reduces the degradation of mRNA by RNase by decreasing secondary structure formation in transcribed mRNAs. Additionally, expression is further enhanced by CspA due to its role in stabilising ribosome binding at the Shine-Dalgarno sequence (Horn *et al*., 2007). Genes homologous to CspA and other CSPs have been found in a number of cold-adapted microbes and a similar mechanism of cold shock response exists in these organisms (Horn *et al*., 2007; Ray *et al*., 1998). The main differences between mesophiles and psychrophiles lies in the fact that theses genes are constitutively expressed, synthesised at higher rates and are maintained in the cytosol after cold shock in cold adapted microbes (Hébraud *et al*., 1999; Horn *et al*., 2007; Ray *et al*., 1998; Ulusu *et al* 2001).

A second class of proteins produced by psychrophilic bacteria are the cold accumulatory proteins [CAPs] (Chattopadhyay, 2006). During prolonged growth at low temperatures CAPs are over-expressed in cold-tolerant bacteria. However, the molecular mechanism of this response in psychrophiles is yet to be elucidated (Chattopadhyay, 2006). Other proteins that are implicated in cold adaptations is summarised in Table 1.

Table 1 Proteins implicated in cold adaptation. Taken from Rodrigues *et al*., 2008.

Protein(s)	Function(s)
AceE	Decarboxylation of pyruvate (pyruvate dehydrogenase)
AceF	Dihydrolipoyltransacetylase (pyruvate dehydrogenase)
CspA	RNA chaperone
CspB	RNA/DNA chaperone (?)
CspE	Regulation of CspA
CspG	RNA/DNA chaperone (?)
CspI	Unknown
CsdA	RNA unwinding activity
DnaA	DNA binding and replication (initiation); transcriptional regulator
RbfA	30S ribosomal binding factor
InfA	Initiation factors; binding of charged tRNA-fmet to the 30S ribosomal subunit
InfB	
PNP	Degradation of RNA
Hsc66	Molecular chaperone
HscB	DnaJ homolog
HU-β	Nucleoid protein; DNA supercoiling
Trigger factor	Prolyl-isomerase activity and other functions
RecA	Recombination factor
GyrA	DNA topoisomerase
H-NS	Nucleoid-associated DNA-binding protein
NusA	Involved in termination and antitermination
OtsA	Trehalose phosphate synthase
OtsB	Trehalose phosphatase
Desaturases	Unsaturation of membrane lipids
Dihydrolipoamide acetyltransferase	Decarboxylation of pyruvate
Alpha-glutamyltranspeptidase	Glutathione metabolism

1.4.1.3 Enzyme adaptation

Considering that temperature is one of the most important environmental factors governing biochemical reactions, enzymes need to be suitably adapted in order to perform their catalytic activity (D'Amico *et al*., 2002). According to the Arrhenius equation, any decrease in temperature will cause an exponential decrease of reaction rates catalysed by enzymes (D'Amico *et al*., 2002). Increased flexibility or plasticity of enzymes has been proposed as the main structural feature of cold adaptation, by allowing better accessibility of substrates to the catalytic cavity at low temperatures (Gerday *et al*., 2000). Plasticity of enzymes does however result in increased thermal sensitivity and a decrease in the enzyme stability (D'Amico *et al*., 2002).

This flexibility of cold-adapted enzymes is achieved by a number of structural modifications such as decreased core hydrophobicity, decreased arginine: lysine ratios, increased clustering of glycine residues near functional domains, fewer salt bridges, decreased isoleucine content, decreased amounts of aromatic-aromatic interactions and modified α-helix dipole interactions (Cavicchioli *et al*., 2002; Nichols *et al*., 1999; Ray *et al*., 1998; Russell, 2000). Since active site residues are conserved in homologous enzymes, molecular and structural changes that are observed in cold-adapted biocatalysts must occur elsewhere in the protein (D'Amico *et al*., 2002).

The increased flexibility of enzymes is accompanied by a decrease in the activation enthalpy of a reaction and lowers the energetic cost of conformational changes required when enzyme and substrate interact (D'Amico *et al*., 2002; Gianese *et al*., 2001). Different strategies of structural adaptation may be adopted by different enzyme families and may be unique to each enzyme (Gerday *et al*., 2000; Gianese *et al*., 2001). For example, in a comparison of structures of 21 psychrophilic enzymes belonging to different families, significant substitution of proline residues was only observed for the α-amylase family (Gianese *et al*., 2001). Similarly, when an Antarctic lipase was structurally compared to its mesophillic counterpart, *Pseudomonas glumae*, by Arpigny and co-workers, it was observed that the absence of certain salt bridges in the Antarctic lipase was essential for cold adaptation and high flexibility of the active site (Arpigny *et al*., 1997). Considering that other ecological and physico-chemical parameters are involved with protein structure and modification, it is important that all the characteristics of an environment be taken into account when assessing adaptive strategies utilised by microorganisms (D'Amico *et al*., 2002).

1.4.1.4 Other strategies

Some features exhibited by microorganisms during cold stress include slower overall growth rates, reduction or inhibition of cell division and long life cycles (Peck *et al*., 2005; Ulusu *et al*., 2001). The formation of dormant cell types which continue to respire and utilise substrates, is also a possible survival strategy employed by bacteria under adverse conditions (Chattopadhyay, 2006).

Production of antifreeze proteins (AFPs) has been well documented in Antarctic fish species such as *Trematomus bernacchi* (De Vries *et al*., 1970). The role of AFPs in bacterial cold adaptation received attention after these molecules where detected in 11 Antarctic lake isolates (Gilbert *et al*., 2004). AFPs are believed to contribute to freeze tolerance and not necessarily freeze avoidance (Chattopadhyay, 2006). Uptake or production of cryoprotectants such as glycine betaine in bacteria is thought to prevent protein aggregation that is induced by cold stress (Chattopadhyay, 2006; Rodrigues *et al*., 2008).

Polyhydroxy-alkanoate and polyamide compounds serve as intracellular carbon and nitrogen reservoirs (Rodrigues *et al*., 2008). Prolonged extreme cold conditions may restrict the uptake of these molecules and the reserves can therefore insure a supply of carbon and nitrogen to bacterial cells.

1.5 Lipolytic enzymes

Lipids constitute a large and essential portion of biomolecules in living systems and participate in energy storage, cell signalling processes and as structural components in membranes (Gilham *et al*., 2005; Hasan *et al*., 2006). Lipolytic enzymes are ubiquitous in nature and include esterases (E.C 3.1.1.1) and lipases (E.C 3.1.1.3). These enzymes are responsible for the metabolism of lipids within cells as well as in the extracellular milieu (Hasan *et al*., 2006). Esterases preferentially hydrolyse short chain (<C10) ester-containing molecules that are partly soluble in water, while true lipases exhibit a broader substrate range with maximal activity towards water-insoluble fatty acyl molecules (>C10) (Arpigny *et al*., 1999; Jaeger *et al*., 1999; Fojan *et al*., 2000; Gilham *et al*., 2005). Due to the important role that these enzymes play as virulence factors, produced and secreted by pathogenic bacteria as well as their application in a variety of industrial and biotechnological processes (see section 1.5.2.1), they are receiving considerable attention (Rosenau, 2000).

Three dimensional structures of both enzymes show that they exhibit a definite order of α-helixes and β-sheets, known as the α/β hydrolase fold (Figure 6). This fold consists of 8 β-sheets, (of which the second is anti-parallel to the others) with β3 and β8 connected by α-helices packed on either side of the central parallel β-sheet (Jaeger *et al*., 1999).

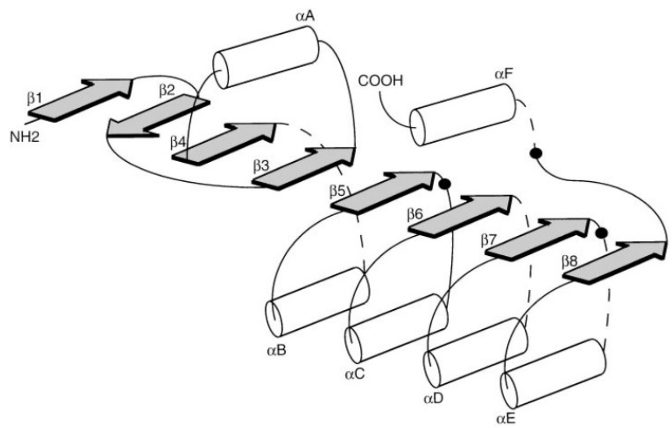


Figure 6 The canonical structure of the α/β hydrolase fold. α-Helices are shown as cylinders and β-sheets as shaded arrows. Solid circles indicate the topological position of active site residues (nucleophile after β5, Asp/Glu after β7 and His in the loop between β8 and αF). Taken from Jaeger *et al*., 1999.

For both lipases and esterases, the reaction mechanism of hydrolysis is essentially the same and consists of four steps [Figure7] (Jaeger *et al*., 1999; Bornscheuer, 2002).

1. Binding of the substrate to the active site serine results in formation of a transient tetrahedral intermediate, stabilised by interactions with the NH-groups of the catalytic His and Asp residues.
2. The histidine residue donates a proton and the alcohol component of the substrate is released.
3. Nucleophilic attack by water on the carbonyl C atom of the covalent intermediate occurs (deacylation step).
4. 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, liberating the free enzyme.

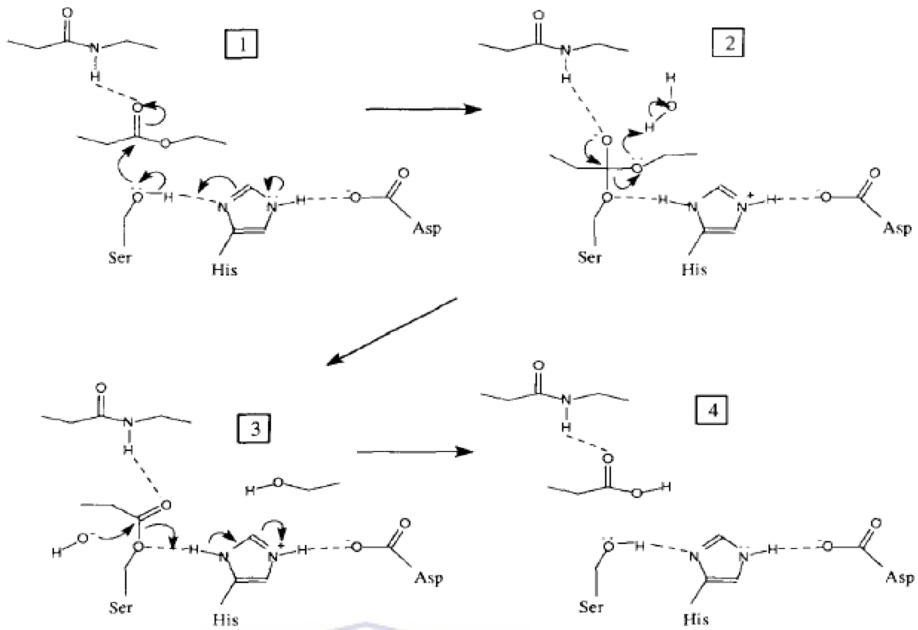


Figure 7 Mechanism of ester bond hydrolysis by lipolytic enzymes. Taken from Jaeger *et al*., 1994.

Lipases and esterase are generally distinguished from each other on the basis of substrate specificity and the phenomenon of interfacial inactivation (Jaeger *et al*., 1999; Verger, 1998). This is, however, a crude method of distinction, since lipases are capable of hydrolysing short-chain esters and not all of these enzymes exhibit interfacial inactivation (Jaeger *et al*., 1999; Verger, 1998). Esterases follow classical Michaelis-Menten kinetics, where activity is a function of substrate concentration and the maximal rate is achieved at substrate saturation (Jaeger *et al*., 1994; Bornscheuer, 2002). Lipases (with some exceptions) exhibit increased enzymatic activity on emulsions (insoluble substrates) when compared to monomeric (soluble) solutions of the same substrate (Jaeger *et al*., 1994; Verger, 1998).

A lid, consisting of a single or double helix, or a loop region, covers the active site in the absence of a lipid-water interface. Increased activity in the presence of hydrophobic substrates causes a conformational rearrangement of the active site, making catalytic residues more accessible (Jaeger *et al*., 1994; Verger, 1998).

Arpigny and Jaeger (1999) classified 53 bacterial lipolytic enzymes into 8 families based on amino acid sequence, important structural features and fundamental biological properties such as mechanism of secretion (Table 2). For the purpose of this review, only one family will be discussed due to its relevance to the current study. The catalytic triad of lipolytic enzymes is commonly composed of a Ser-Asp-His, with the active site serine found in a GXSXG consensus motif (referred to as the nucleophile elbow) located in the middle of the gene (Arpigny *et al*., 1999). It has been shown (Upton *et al*., 1995; Arpigny *et al*., 1999) that not all lipolytic enzymes contained this GXSXG consensus motif, but that family II GDSL esterases/lipases contained a GDS(L) motif located closer to the N-terminal of the protein (Upton *et al*., 1995; Jaeger *et al*., 1999). Furthermore, these enzymes have four strictly conserved residues, Ser-Gly-Asn-His, in four conserved blocks; I, II, III and IV, respectively. Each block plays an essential role in catalytic function of these enzymes (Akoh *et al*., 2004). Due to the absence of the nucleophile elbow and a different tertiary fold structure, these enzymes are not members of the α/β hydrolase-fold superfamily but rather belong to the SGNH hydrolase superfamily (Akoh *et al*., 2004, Jaeger *et al*., 1999; Verger 1998). These enzymes also share very little sequence homology to true lipases and some members exhibit protease, arylesterase and thioesterase activity (Arpigny *et al*., 1999; Bornscheuer, 2002; Akoh *et al*., 2004). Another characteristic feature of these enzymes is a covalently bound C-terminal autotransporter domain, which forms a β-barrel pore in the outer membrane of Gram negative bacteria (Arpigny *et al*., 1999; Jacob-Dubuisson, 2004).

Both gram negative and gram positive bacteria generally utilise the Sec-dependant pathway (also termed the general secretory pathway; GSP) for protein secretion across the inner membrane (Henderson *et al*., 2004; Jacob-Dubuisson *et al*., 2004). A terminal branch of the GSP is the autotransporter secretion pathway (also termed the type V secretion system) that enables gram negative bacteria to export proteins across the cellular envelope (Thanassi *et al*., 2000; Henderson *et al*., 2004; Jacob-Dubuisson *et al*., 2004). Proteins secreted via autotransporter systems typically contain N-terminal signal peptides allowing for their targeting and transport across the inner membrane via the GSP system (Henderson *et al*., 2004; Jacob-Dubuisson *et al*., 2004).

Internal passenger domains are covalently linked to their C-terminal translocator units, which consist of 250-300 amino acid residues (Henderson *et al*., 2004). The translocation units are predicted to contain a α-helical linker region, followed by 10 to 14 amphipathic β-strands that form an anti-parallel β-barrel structure in the outer membrane (Jacob-Dubuisson *et al*., 2004). Phenylalanine or tryptophan is the terminal amino acid residue in autotransporters and is preceded by alternating hydrophilic and hydrophobic residues (Henderson *et al*., 2004; Jacob-Dubuisson *et al*., 2004). Furthermore, mutagenesis studies by Lee and Byun (2003) showed that two strictly conserved residues [proline and glycine] are essential for proper folding of β-barrels and for active translocation of the N-terminal domains (Lee *et al*., 2003).

Once transported across the inner membrane, the proprotein exists as a periplasmic intermediate, where partial folding of the autotransporter may occur. This intermediate protein is also accessible to periplasmic enzymes (Henderson *et al*., 2004). Translocated passenger domains undergo alternative processing steps; they may be processed and released into the extra-cellular medium, or, once cleaved, remain in close association with the bacterial cell surface by non-covalent β-domain interactions (Jacob-Dubuisson *et al*., 2004). Passenger domains may not be cleaved at all, thereby remaining as intact proteins, membrane bound by the C-terminal and with N-terminal domains extending into the extra-cellular matrix (Henderson *et al*., 2004; Jacob-Dubuisson *et al*., 2004). A model of passenger domain secretion across the outer membrane is shown in Figure 8.

Yen *et al.,* (2002) classified autotransporters into 10 phylogenetic clusters. Cluster 10 represents those autotransporters with lipase/esterase passenger domains. Some examples from this cluster include; EstA (*Pseudomonas aeruginosa*), ApeE (*Salmonella enterica serovar typhimurium*), Lip1 (*Photorhabdus luminescens*) and McaP (*Moraxella catarrhalis*) (Yen *et al*., 2002; Henderson *et al*., 2004). Interestingly, all of these are also classified as members of the GDSL family of lipolytic enzymes, based on conserved motifs found in their sequences (Yen *et al*., 2002).

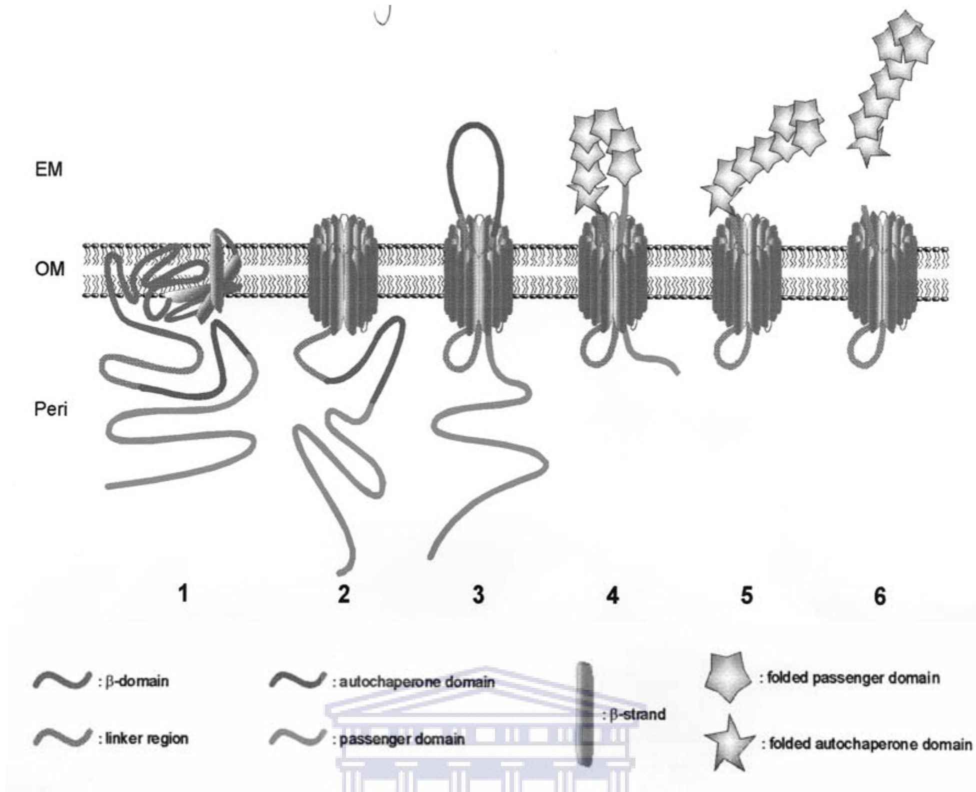


Figure 8 A model of the passenger domain secretion across the outer membrane (Peri: periplasm; OM: outer membrane; EM: extracellular milieu). Step 1 to 2: insertion and folding of the β-domain into the outer membrane. Steps 2 to 3: the linker region initiates secretion through the pore. Steps 3 to 6: folding and secretion of the passenger domain followed by its release into the extracellular milieu. Taken from Desvaux *et al*., 2004.

Table 2 Families of lipolytic enzymes. Taken from Arpigny *et al*., 1999.

Family	Subfamily	Enzyme-producing strain	Accession no.	Similarity (%) Family	Similarity (%) Subfamily	Properties
I	1	*Pseudomonas aeruginosa**	D50587	100		True lipases
		Pseudomonas fluorescens C9	AF031226	95		
		Vibrio cholerae	X16945	57		
		Acinetobacter calcoaceticus	X80800	43		
		Pseudomonas fragi	X14033	40		
		Pseudomonas wisconsinensis	U88907	39		
		Proteus vulgaris	U33845	38		
	2	*Burkholderia glumae**	X70354	35	100	
		*Chromobacterium viscosum**	Q05489	35	100	
		*Burkholderia cepacia**	M58494	33	78	
		Pseudomonas luteola	AF050153	33	77	
	3	*Pseudomonas fluorescens SIK W1*	D11455	14	100	
		Serratia marcescens	D13253	15	51	
	4	*Bacillus subtilis*	M74010	16	100	
		Bacillus pumilus	A34992	13	80	
	5	*Bacillus stearothermophilus*	U78785	15	100	
		Bacillus thermocatenulatus	X95309	14	94	
		Staphylococcus hyicus	X02844	15	29	Phospholipase
		Staphylococcus aureus	M12715	14	28	
		Staphylococcus epidermidis	AF090142	13	26	
	6	*Propionibacterium acnes*	X99255	14	100	
		Streptomyces cinnamoneus	U80063	14	50	
II (GDSL)		*Aeromonas hydrophila*	P10480	100		Secreted acyltransferase
		*Streptomyces scabies**	M57297	36		Secreted esterase
		Pseudomonas aeruginosa	AF005091	35		OM-bound esterase
		Salmonella typhimurium	AF047014	28		OM-bound esterase
		Photorhabdus luminescens	X66379	28		Secreted esterase
III		*Streptomyces exfoliatus**	M86351	100		Extracellular lipase
		Streptomyces albus	U03114	82		Extracellular lipase
		Moraxella sp.	X53053	33		Extracellular esterase 1
IV (HSL)		*Alicyclobacillus acidocaldarius*	X62835	100		Esterase
		Pseudomonas sp. B11-1	AF034088	54		Lipase
		Archaeoglobus fulgidus	AE000985	48		Carboxylesterase
		Alcaligenes eutrophus	L36817	40		Putative lipase
		Escherichia coli	AE000153	36		Carboxylesterase
		Moraxella sp.	X53868	25		Extracellular esterase 2
V		*Pseudomonas oleovorans*	M58445	100		PHA-depolymerase
		Haemophilus influenzae	U32704	41		Putative esterase
		Psychrobacter immobilis	X67712	34		Extracellular esterase
		Moraxella sp.	X53869	34		Extracellular esterase 3
		Sulfolobus acidocaldarius	AF071233	32		Esterase
		Acetobacter pasteurianus	AB013096	20		Esterase
VI		*Synechocystis sp.*	D90904	100		Carboxylesterases
		Spirulina platensis	S70419	50		
		*Pseudomonas fluorescens**	S79600	24		
		Rickettsia prowazekii	Y11778	20		
		Chlamydia trachomatis	AE001287	16		
VII		*Arthrobacter oxydans*	Q01470	100		Carbamate hydrolase
		Bacillus subtilis	P37967	48		*p-Nitrobenzyl esterase*
		Streptomyces coelicolor	CAA22794	45		Putative carboxylesterase
VIII		*Arthrobacter globiformis*	AAA99492	100		Stereoselective esterase
		Streptomyces chrysomallus	CAA78842	43		Cell-bound esterase
		Pseudomonas fluorescens SIK W1	AAC60471	40		Esterase III

*Lipolytic enzyme with known 3D structure.

1.5.1 Biotechnological application of lipolytic enzymes

The use of enzymes in industrial processes allows a high level of control of the products being manufactured (Hasan *et al*., 2006). Unwanted side reactions may be reduced due to specificity of the enzyme used. These biomolecules are biodegradable and contribute minimal biological oxygen demand (BOD) in waste streams. Enzymes of microbial origin are useful in many industrial processes due to their stability, higher yield and regular supply (Hasan *et al*., 2006; Joseph *et al*., 2008).

Lipases are the third largest group of commercial enzymes (based on total sales volume) and are exceeded only by proteases and carbohydratases (Hasan *et al*., 2006). Lipolytic enzymes are valuable as they show broad substrate range, are stable in organic solvents, may be purified in large quantities and catalyse both anabolic and catabolic reactions (Hasan *et al*., 2006).

1.5.1.1 Lipolysis

Lipolysis is the catabolism of fats or esters into constituent acid and alcohol/glycerol in the presence of water (Gandhi, 1997). The hydrolytic properties of lipolytic enzymes are employed in a number of industrial processes.

Microbial carboxyl esterases are employed in the hydrolysis of pectin or xylan in plant cell walls to liberate ferulic acid and also in catabolism of aryl-esters (Bornscheuer, 2002). In the leather industry, lipases are used for the removal of residual fats and protein debris associated with hair and hides (Hasan *et al*., 2006). Thin layers of fat must be removed in activated sludge and aerobic waste treatment to permit oxygen transport. Lipases are used to degrade the lipid-rich liquid that is skimmed from these systems (Gandhi, 1997).

Lipases are utilized extensively in the food industry. Conventional chemical processing of fats and oils requires harsh conditions of temperature and pressure that produce undesirable side reactions such as decolourisation, odour and oxidation of fatty acids (Jaeger *et al*., 1994; Gandhi, 1997; Jaeger *et al*., 2002). Acceleration of flavour development occurs when free fatty acids and soluble peptides and amino acids are formed in the maturation stages of a dairy product (Hasan *et al*., 2006). Lipases impart rich creamy flavours to coffee whiteners, caramels and toffees, and chocolate (Gandhi, 1997). Extended shelf-life of breads and improved crumb structure of baked goods is achieved when lipolytic enzymes are used in the bakery industry (Hasan *et al*., 2006). One of the most important applications of lipases in industry is the resolution of racemic mixtures and synthesis of chiral building blocks for pharmaceuticals (drug production) and agrochemicals (pesticides) (Jaeger *et al*., 1994; Gandhi, 1997; Hasan *et al*., 2006). Other applications that utilise the hydrolytic power of lipases and esterases include oil biodegration and biodiesel production (Hasan *et al*., 2006), digestive aids as well in the paper and pulp industry (Gandhi, 1997).

1.5.1.2 Ester synthesis

Lipolytic enzymes are capable of catalysing the reverse reaction and in the process, liberate water. In low water activity systems, the normal hydrolytic equilibrium can be reversed in favour of esterification reactions (Jaeger *et al*., 1994; Sharma *et al*., 2001). Acidolysis, interesterification and alcoholysis reactions give rise to acids, esters or alcohol instead of water

(Gandhi, 1997). This ability of lipases is important in oleochemical processes where less useful fats may be converted to more valuable ones (Hasan *et al*., 2006). Interesterification reactions refers to the simultaneous hydrolysis and esterification and has been applied for the conversion of palm oil into cocoa butter, a high value product used in food, confection, pharmaceuticals and the cosmetic industry (Gandhi, 1997; Sharma *et al*., 2001; Hasan *et al*., 2006).

1.5.2 <u>Cold-active lipolytic enzymes</u>

Permanently cold habitats exert high selective pressure on the resident population. Organisms colonising these environments have developed strategies of adaptation, allowing survival under extreme physico-chemical conditions (Ferrer *et al*., 2007; Gerday *et al*., 2000). For example, increasing membrane fluidity of cells by tailoring acyl chains in membranes, thereby increasing lipid saturation, allows for appropriate exchange of solutes between cells and the external medium (Gerday *et al*., 1997). Since low temperatures can slow down or even inhibit biochemical reactions, enzymes are key targets for cold adaptation (D'Amico *et al*., 2002; Gerday *et al*., 2000). Table 3 gives a summary of selected cold-active lipolytic enzymes have been discovered and characterised.

Lipolytic enzymes produced by cold-adapted microbes have evolved structural features conferring a high degree of flexibility around the active site (Joseph *et al*., 2007). As a result, low activation enthalpy and high specific activity at low temperatures is observed. This flexibility could be caused by structural changes such as a low number of arginine residues compared with lysine, low proline content particularly in loop regions, increased clustering of glycine residues, a small number of salt bridges and aromatic-aromatic interactions, decrease in the number of hydrogen bonds and weakening of hydrophobic clusters (Joseph *et al*., 2007; Joseph *et al*., 2008; Rodrigues *et al*., 2008; Russell, 2000).

1.5.2.1 Applications of cold-active lipolytic enzymes

High catalytic activity at low temperatures and thermolability of lipolytic enzymes are the key to their success in the detergent industry, the food industry, environmental bioremediation and the textile industry (Joseph *et al*., 2007). Unwanted side reactions are eliminated and energy consumption and environmental impact are greatly reduced (Joseph *et al*., 2008). Rapid inactivation of heat liable enzymes increases mechanical resistance of fabrics, while cold washing reduces the wear and tear on fabrics (Gandhi, 1997). In bioremediation schemes, seasonal fluctuations influence the effectivity of pollutant degradation. Application of both mesophilic and psychrophilic enzyme preparations may enhance the process due to activity over a varied range of temperatures (Gandhi, 1997; Joseph *et al*., 2007).

Table 3 A selection of cold-active bacterial lipolytic enzymes.

Organism	Source	Optimum temperature	Optimum pH	Mr (kDa)	Substrate specificity (*p*-nitrophenyl esters)	Conserved motif	Reference
Pseudomonas sp. Strain B1-11	Alaskan tundra soils	45	8	33	C_4	GXSXG	Choo *et al.*, 1998
Unknown, metagenomic fosmid library	Baltic sea sediment	35	8	35.4	C_8	GXSXG	Hårdeman *et al.*, 2007
Aeromonas sp. Strain LPB4	Sea sediment	35	Not specified	50	C_6	GXSXG	Lee *et al.*, 2003
Serria marcenscens	Raw milk	37	8	52	Not specified	GXSXG	Abdou, 2003
Pseudomonas sp. KB700A	Subterranean environment	35	8.5	49.9	C_{10}	GXSXG	Rashid *et al.*, 2001
Actinobacter sp. Strain 6	Siberian soil	20	Not specified	Not specified	C_8	Not specified	Suzuki *et al.*, 2001

1.6 Aims

The general aim of this study was to construct a fosmid library using Antarctic soil metagenomic DNA and screen this library for putative novel lipolytic genes. Additionally, the library will be screened for 16S rRNA signals (archaeal and bacterial) using sequence-based methods in order to access 16S rDNA capture in the fosmid library.

- Construct a fosmid library of metagenomic DNA extracted from soil samples collected from under seal carcasses in the Dry Valleys of Antarctica.

- Screen the library for lipolytic function using activity-based methods.

- If lipolytic activity was detected, obtain the gene sequence and bioinformatically analyse the gene

- Clone, express and partially characterise genes of interest

Chapter 2: Materials and methods

2.1 General microbiological techniques

2.1.1 Media

Luria-Bertani (LB) broth contained 1% [w/v] tryptone, 0.5% [w/v] yeast extract and 1% [w/v] NaCl. Luria-Bertani (LB) agar was prepared from LB broth with the addition of 1.3% [w/v] bacteriological agar. Tributyrin agar consisted of 1% [w/v] tryptone, 0.5% [w/v] yeast extract, 1% [w/v] NaCl, 1.3% [w/v] bacteriological agar, 1% [v/v] tributyrin, 1% [w/v] gum arabic. SOB broth consisted of 2% [w/v] tryptone, 0.5% [w/v] yeast extract, 0.05% [w/v] NaCl, 0.02% [w/v] KCl. All components were mixed together with distilled water and the pH was adjusted to 7.0 using 1 M NaOH. Media were autoclaved at 121°C for 20 minutes.

SOC was prepared from SOB with the addition of filter sterilised 2 M $MgCl_2$ to 0.5% [w/v] and 1 M glucose to 2% [w/v].

After the media was autoclaved and cooled to ~50°C, the appropriate filter sterilised antibiotic was aseptically added. Final concentrations of antibiotics were: (unless otherwise stated) chloramphenicol (cam), 12.5 µg/ml; carbenicillin (carb), 50 µg/ml; kanamycin (kan), 30 µg/ml and gentomycin (gen), 20 µg /ml.

2.1.2 Growth of *E. coli* strains

Bacterial strains were grown in broth or on solid media supplemented with the appropriate antibiotic. The native EPI-300 *E. coli* strain was grown on media with no antibiotic. Strains were inoculated using aseptic technique. Unless otherwise stated, cultures were incubated at 37°C. If strains were grown in broth, incubation was accompanied by agitation at 150 to 225 rpm.

<u>Table 4</u> Strains, plasmids and primers used in this study.

	Characteristics	Source
<u>Bacterial strains</u>	Genotype	
<u>*E. coli*</u>		
EPI-300	F- mcrA D(*mrr*-*hsd*RMS-*mcr*BC) f80d*lac*ZDM15 D*lac*X74 *rec*A1 *end*A1 *ara*D139 D(*ara*, *leu*)7697 *gal*U *gal*K l- *rps*L *nup*G *trf*A *ton*A *dhfr*	Epicentre Biotechnology (USA)
ArcticExpress (DE3)	*E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte [*cpn10 cpn60* Gentr]	Strategene
Rosetta(DE3)pLysS	F^- *ompT* *hsdS*$_B$(r_B^- m_B^-) *gal* *dcm* (DE3) pLysSRARE (Cam^R)	Novagen (USA)
BL21(DE3)pLysS	F–, *omp*T, *hsd*S_B (r_B–, m_B–), *dcm*, *gal*, λ(DE3), pLysS, Cm^r	Invitrogen (USA)
<u>Plasmids/vectors</u>		
pCCFos1	$Chloramphenicol^R$ 12.5 µg/ml	Novagen (USA)
pUC19	$Ampicillin^R$ 100 µg/ml	Novagen (USA)
pET28a	$Kanamycin^R$ 30µg/ml	Novagen (USA)
<u>Primers</u>		
Universal 16S rRNA genes		
<u>Bacteria</u>		
341 F-GC	5' CGCCCGCCGCGCGCGGCGGGCGGGG CGGGGGCACGGGGGGCCTACGGGAGGC AGCAG 3'	Muyzer *et al*., 1993
534 r	5' ATTACCGCGGCTGCTGG 3'	
<u>Archaea</u>		
Ua1204 R	5' TTMGGGGCATRCIKACCT 3'	Baker *et al*., 2003
A571Fb	5' GCYTAAAGSRICCGTAGC 3'	
AB927R	5' CCCGCCAATTCCTTTAAGTTTC 3'	Jurgens *et al*., 1997
A3FA	5' TCCGGTTGATCCYGCCGG 3'	Baker *et al*., 2003

Table 4 continued

Primer walking		
Transposon mutagenesis N primer	5´ACTTTATTGTCATAGTTTAGATCTATTT TG 3’	New England Biolabs
Transposon mutagenesis S primer	5´ATAATCCTTAAAAACTCCATTTCCACC CCT 3´	New England Biolabs
LD1-TM3 F	5’ CCTTAACTGGTAAATGTGG 3’	This study
LD1-TM3 R	5’ GCACCTAAGCGTTTAGATG 3’	This study
LD1-RUS1	5’ GCTTGAGCCAAACGACAGTGC 3’	This study
LD1-US2	5’ GCGCGCCATCTCTGGTAAC 3’	This study
LD1-RUS2	5’ CAGCGCAATAACCTCAGC 3’	This study
LD1-RUS3	5’ CGTCTACAACGACAGAACCATCA GC 3’	This study
Sub-cloning LD1		
LD1-R-X1	5’ ACCTCGAGTTACCAGTTAAGGCT TAC 3’	This study
LD1-F-N1	5’ GTGCATATGAAGAAGGTACTGG 3’	This study
LD1-F-Bh1	5’ GCGGATCCATGAAGAAGGTACT GG 3’	This study

2.2 General molecular biology techniques

2.2.1 DNA extraction

a. Zhou method

Extraction buffer (0.1% [w/v] CTAB; 100 mM Tris, pH 8; 100 mM NaH_2PO_4, pH 8; 100 mM EDTA; 1.5 M NaCl; 0.02% [v/v] Protease K) was prepared and an equal volume was added to pre-weighed soil samples. Tubes were incubated horizontally for 30 minutes at 37°C and 225 rpm. 20% [w/v] SDS (750 µl per 5 ml) was added to each tube followed by further incubation at 65°C for 2 hours with gentle inversion every 20 minutes, then centrifuged at 3000 × *g* for 10 minutes (Stach *et al*., 2001).

b. Crude DNA extraction

E. coli cells were streaked onto LBA plates and grown overnight. One colony was transferred to 5 ml LB broth and incubated overnight. Four milliliters of the culture was harvested by centrifugation at 10000 × *g* for 5 minutes. The pellet was re-suspended in 500 µl of TE buffer (pH 7.7) and centrifuged at 10000 × *g* for 2 minutes. The supernatant was discarded; the pellet was re-suspended in 500 µl of TE buffer (pH 7.7) and boiled at 100°C for 10 minutes. Once the sample was cooled the tubes were centrifuged at 10000 × *g* for 2 minutes and 400 µl of the supernatant was transferred to a sterile eppendorf tube and stored at 4°C.

2.2.2 Phenol: chloroform: isoamyl alcohol (25:24:1) and ethanol precipitation

After cell lysis, the supernatant was transferred to sterile eppendorf tubes. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added and tubes were centrifuged at 13000 $\times g$ for 10 minutes. The supernatant was transferred to sterile eppendorf tubes, an equal volume of chloroform was added and tubes were centrifuged at 13000 $\times g$ for 10 minutes. The supernatant was transferred to sterile eppendorf tubes, 0.6 × vol of isopropanol was added and tubes were left at room temperature for 1-2 hours to precipitate DNA. Tubes were centrifuged at 10000 $\times g$ for 10 minutes. The DNA pellet was washed in ice-cold 70% ethanol, centrifuged at 1000 $\times g$ for 10 minutes and the supernatant decanted. The wash step was repeated.

2.2.3 Agarose gel electrophoresis

0.7% or 1 % (w/v) agarose was dissolved in 0.5 × TAE buffer (0.2% [w/v] Tris base, 0.5% [v/v] glacial acetic acid, 1% [v/v] 5 M EDTA [pH 8]. Cast gels were electrophoresed at 100 V in 0.5 × TAE buffer. To allow visualisation of the DNA on a UV transilluminator, the gels were supplemented with 0.5 μg/ml ethidium bromide. Samples were mixed with standard loading dye (60% [v/v] glycerol, 0.25% [w/v] Orange G) and loaded into the wells of the cast gels. DNA was sized according to its migration in the gel as compared to that of DNA molecular markers used (Lambda DNA restricted with *Hind*III; Lambda DNA restricted with *Pst*I; Fosmid control DNA [Epicentre]).

2.2.4 <u>DNA quantification</u>

DNA concentration was measured by fluorometery using the Quanti-iT ™ ds DNA BR assay kit and the Qubit™ system (Invitrogen, Oregon, USA) according to the manufacturers specifications.

2.2.5 <u>DNA purification</u>

a. GELase (EPICENTRE®, Madison, Wisconsin)

Based on the assumption that 1 mg of solidified agarose will yield 1 µl of molten agarose upon heating, 3 µl of 50 × GELase buffer (2 M Bis-Tris [pH 6], 2 M NaCl) was added to a 150 mg LMP agarose gel slice. The agarose was melted at 70°C for 5 to 10 minutes, quickly transferred to 45°C and allowed to equilibrate for 5 minutes. One and a half units of GELase enzyme (1 U/ml) preparation was added to each tube (1 unit for every 100 µl molten agarose) and incubated overnight at 45°C. The enzyme was inactivated by incubation for 15 minutes at 70°C.

Five hundred microliter aliquots were transferred to 1.5 ml sterile eppendorf tubes, chilled on ice for 5 minutes and centrifuged at 10000 × *g* for 20 minutes to pellet any insoluble oligosaccharides. The upper 90%-95% of resulting supernatant, containing the DNA, was removed, transferred to clean tubes and precipitated for 2 hours at -20°C, using 0.1 vol 3 M NaOAc (pH 7) and 2.5 vol of ice-cold absolute ethanol.

b. GFX ™

The Illustra ™ GFX™ PCR DNA and gel band purification kit (GE Healthcare Limited, Buckinghamshire, UK) was used to purify DNA from solution or agarose according the manufacturers specifications.

2.2.6 Fosmid extraction

Selected clones were inoculated into 5 ml LB-cam and incubated overnight. One millilitre of the culture was inoculated into a tube containing 9 ml LB-cam and 10 µl induction solution (EPICENTRE®) and grown for 5 hours at 37°C with agitation. Tubes were centrifuged at 4000 × *g* for 30 minutes at 4°C. The supernatant was decanted and the tubes inverted on a paper towel to remove any excess medium. Cells were re-suspended in 1 ml of cooled GET buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl) and 24 µl of 10 mg/ml RNase A (Fermentas) was added. One millilitre of lysis solution (0.2 M NaOH, 1% SDS) was added and the tubes were gently inverted. One millilitre 3 M KOAc (pH 5.5) was added, the tubes were inverted and cells were incubated on ice for 5 minutes. The tubes were inverted again and incubated on ice for a further 10 minutes.

2.2.7 Restriction enzyme digestion

Restriction enzyme digestions were performed in sterile eppendorf tubes in small reaction volumes (10–50 µl). The reactions contained the appropriate volume of 10 × or 2 × buffer (supplied by the manufacturer for the specific enzyme) and 5-10 U of enzyme per µg of plasmid or genomic DNA. Reactions were incubated for either a 2 hour period, or overnight at 37°C. The digestion products were analysed by gel electrophoresis on 0.7% or 1% (w/v) agarose gels [section 2.2.3].

2.2.8 Preparation of electrocompetent *E. coli* cells

A single colony of EPI-300 *E. coli*, streaked from a glycerol stock onto LBA-cam $_{\text{and}}$ grown overnight, was used to inoculate 10 ml SOB. This starter culture was grown overnight at 37°C with shaking (225 rpm). One litre of SOB media was inoculated with the starter culture and grown at 37°C until an OD_{600} of 0.6-0.9 was reached. The cells were kept on ice and 250 ml aliquots were transferred to chilled Corning bottles.

Bottles were centrifuged at 4000 × *g* for 25 minutes at 4°C, the supernatant was poured off and the pellet was gently resuspended in 200 ml ice-cold demineralised water (dH_2O) [Millipore] before another round of centrifugation at 4000 × *g* for 25 minutes at 4°C. Once the supernatant was removed, the cells were resuspended in 100 ml ice-cold dH_2O and centrifuged at 4000 × *g* for 25 minutes at 4°C. Bottles were placed on ice, the supernatant was removed and the cell pellet was resuspended in 20 ml ice-cold 10% [v/v] glycerol and centrifuged at 4000 × *g* for 25 minutes at 4°C. After the supernatant was removed, each cell pellet was very gently resuspended in 1 ml 15% [v/v] glycerol and 2% [w/v] sorbitol. The cell suspension was kept on ice, aliquoted into 1.5 ml eppendorf tubes and stored at -80°C. One microliter of pUC19 vector DNA (120 ng/µl) was used to test the electro-competency of the cells [section 2.2.9a].

2.2.9 <u>Transformation of *E. coli* cells</u>

<u>a. Electroporation</u>

Aliquots of 50 µl of electrocompetent cells were thawed on ice. DNA was added directly to cells and incubated on ice for 5 minutes. The mixture was pipetted into pre-cooled electroporation cuvettes (Bio-Rad Laboratories, CA, USA). Electroporation was performed using the following conditions; 1.8 kV, 25 µF, 200 Ω. Nine hundred and fifty microliters of SOC was immediately added to the cuvette and once mixed, transferred to sterile tubes. The mixture was incubated for 1 hour at 37°C with agitation and aliquots were plated on LBA plates supplemented with the appropriate antibiotic and grown overnight.

<u>b. Heat shock</u>

Plasmid DNA was added directly to 20 µl of competent cells, incubated on ice for 5 minutes and heat-shocked at 45°C for 30 seconds. Cells were incubated on ice for a further 2 minutes and 80 µl of SOC was added and the cells were incubated for 1 hour at 37°C. The transformation mix was plated on media supplemented with appropriate antibiotic and incubated overnight at 37°C.

2.2.10 Cell lysis using Bugbuster reagent

Cell cultures were incubated to an OD_{600} of 0.6 at 30°C or 37°C. The culture was centrifuged at 13 000 × *g* and the pellets were resuspended in Bugbuster (Novagen, USA) [5 ml for every 1 g of pelleted cells] and Benzonase nuclease (Novagen, USA) [1 U/ml Bugbuster] and incubated at room temperature for 30 minutes with gentle agitation. The lysed cells were centrifuged at 13 000 × *g* for 5 minutes and the supernatant was transferred to a sterile eppendorf tube.

2.2.11 His-tag purification

His-Bind resin (Novagen, USA) was completely resuspended by gentle inversion. Two milliliters of the slurry was transferred to a purification column and packed by gravitational flow to a final bed volume of 1 ml. To charge and equilibrate the column the following sequence of washes was used;

1. 3 ml sterile demineralised water
2. 5 ml of 1 × charge buffer (8 × = 400 mM NiSO4)
3. 3 ml of 1 × binding buffer (8 × = 4 M NaCl, 160 mM Tris-HCl, 40 mM imidazole [pH 7.9])

After draining of the Binding buffer, prepared extract was added to the column. The column was washed with 10 ml of 1 × binding buffer, 6 ml of 1 × wash buffer (8 × = 4 M NaCl, 480 mM imidazole, 160 mM Tris-HCl [pH 7.9]), 6 ml 1 × elute buffer (4 × = 4 M imidazole, 2 M NaCl, 80 mM Tris-HCl [pH 7.9])and finally 6 ml of 1 × strip buffer (4 × = 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% EtOH at 4°C.

2.2.12 SDS-PAGE

Vertical SDS-PAGE gels were cast with a 10-12% stacking gel (1.5 M Tris-HCl [pH 8.8], 20% [w/v] SDS, 30% [w/v] acrylamide, 0.8% [w/v] bis-acrylamide, 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], 20% [w/v] SDS, 30% [w/v] acrylamide, 0.8% [w/v] bis-acrylamide, 10% [w/v] ammonium persulphate, 0.1% [v/v] TEMED). Samples were mixed with an equal volume of 2 × loading dye (80 mM Tris-HCl [pH 6.8], 10% [v/v] mercaptoethanol, 2% [v/v] SDS, 10% [v/v] glycerine, bromophenol blue), vortexed and heated to 95°C for 5-10 minutes.

Samples were loaded on gels and electrophoresed at 60 V in 1 × running buffer (0.25 mM Tris-HCl, 2 M glycine, 1% [w/v] SDS) for 30 minutes through the stacking gel. Electrophoresis continued through the separating gel at 100 V for ~ 2 hours. The gel was stained with coomassie stain (0,125% [w/v] Coomassie blue R250, 50% [v/v] methanol, 10% [v/v] acetic acid) for 45 minutes and de-stained overnight with SDS destain (50% [v/v] methanol, 10% [v/v] acetic acid). The size of the proteins was determined according to their migration in the gel as compared to that of the protein ladder used (Pageruler™ unstained protein ladder [Fermentas]).

2.2.13 Acetone precipitation of proteins

Equal volumes of ice-cold acetone were added to the culture supernatant, incubated on ice for 1 hour and centrifuged at 10000 × *g* for 10 minutes. The pellet was resuspended in sterile demineralised water.

2.2.14 Bradford assay for determination of protein concentration

Ten microliters of sample were mixed with 200 µl of Bradford's reagent and 790 µl of sterile demineralised water and incubated at room temperature for 20 minutes. Optical density measurements were performed at 595 nm and plotted against a 1-20 µg BSA standard curve (Bradford, 1976).

2.2.15 Enzyme assays using *p*-nitrophenyl esters

Nine hundred and seventy microliters of buffer (0.1 M NaCl, 0.1 M NaH_2PO_4, 1% [v/v] acetonitrile), 10 µl of 50 mM substrate (dissolved in acetonitrile) and 10 µl of 1% [v/v] Triton-X 100 was pipetted into a 1 ml cuvette, mixed thoroughly by inversion and the absorbance measured at 405 nm over a period of 3 minutes (Winkler *et al*., 1979). 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.

2.2.16 PCR (Polymerase chain reaction)

a. PCR amplification using lipolytic gene specific primers

PCR reactions (20-50 µl) contained ~ 10 ng of template DNA, 5 × HF PCR buffer, 0.2 mM of each dNTP, 0.5 pmol of each primer (Table 4, page 45) and 0.5 µl Phusion™ *Taq* DNA polymerase (Finnzymes, Finland) DNA polymerase. For control purposes, forward primer, reverse primer and a negative control (a reaction mixture containing all reagents except template) was routinely included. Following PCR, an aliquot of each reaction mixture was analysed using gel electrophoresis as described (section 2.2.3).

b. PCR amplification using 16S rDNA primers

PCR reactions (50 µl) contained ~ 200 ng of template DNA, 10 × NEB PCR buffer (200 mM Tris-HCl [pH 8.8], 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% [v/v] Triton X – 100), 0.2 mM of each dNTP, 0.5 pmol of each primer (Table 4, page 45) and 0.5 µl recombinant *Taq* DNA polymerase [Desai *et al.*, 1995]). A positive control containing either *E. coli* genomic DNA, or an archaeal 16S rRNA gene fragment as template DNA, was included. PCR reaction mixtures were placed in an Applied Biosystems thermocycler Gene Amp® 2700. Forward primer and reverse primer controls, as well as a negative control (a reaction mixture containing all reagents except template) were routinely included. An aliquot of each reaction mixture was analysed using gel electrophoresis as described (section 2.2.3).

2.3. Metagenomic library construction

2.3.1 <u>Sample collection</u>

Environmental samples were collected by members of Antarctica New Zealand Event K021 during January 2006 from the Miers Valley in Eastern Antarctica. Samples were collected at 5 cm intervals along a perpendicular axis, centred with respect to crab-eater seal carcasses. All samples were collected from the top 2 cm of soil using aseptic techniques, stored in sterile RNA/DNA-free 15 ml and 50 ml containers, frozen in the field and stored at – 80°C until analysed. Table 5 gives the geographical position of seal carcasses in the Miers Valley from which samples were collected and used in this study.

Table 5 Geographical position of soil samples taken from under seal carcasses in the Miers Valley, Antarctica.

Sample code (MVS)	Geographical position	Altitude (m)
5	S 78 04.113 E 163 51.158	586
6	S 78 04.007 E 163 51.517	651
7	S 78 03.782 E 163 51.478	651
8	S 78 04.782 E 163 51.478	651
11	S 78 04.024 E 163 51.644	699
14	S 78 04.004 E 163 51.739	715
21	S 78 04.000 E 163 51.899	763
22	S 78 03.996 E 163 51.917	768
25	S 78 04.000 E 163 51.970	788

2.3.2. DNA extraction

The extraction of the total community DNA was performed according to the modified Zhou protocol (Stach *et al*., 2001) [section 2.2.1a]. Samples were thawed on ice. Using a sterile spatula, 20 g dry-weight was measured and placed in sterile 50 ml Falcon tubes. Extraction buffer [section 2.2.1a] was prepared in a sterile UV hood. 20 mg/ml protease K (25 µl per 5ml of buffer) was added to the extraction buffer. Samples were suspended in an equal volume; i.e. 20 ml, of extraction buffer.

The negative control contained sterile demineralised water only. The DNA was extracted by phenol: chloroform: isoamyl alcohol (25:24:1) and ethanol precipitation [section 2.2.2]. Following an overnight isopropanol precipitation at room temperature, the DNA pellet was dried in the laminar flow for 5 minutes and re-suspended in 0.5 M TE buffer (pH 8). The DNA concentration was measured by fluorometery [section 2.2.4]. Five microliters of DNA was loaded onto a 0.7% agarose gel and electrophoresed at 30 V for 18 hours in order to assess the size range of extracted DNA obtained [section 2.2.3]. A total mass of 150 g of soil from the samples listed in Table 5 was used in the DNA extraction. The extracted DNA was pooled in order to generate the fosmid library.

2.3.3. Size fractionation and purification

A 0.7% agarose gel was prepared; a plug was cut out of the gel using a sterile scalpel and was filled with 0.7% low melting point (LMP) agarose. Total extracted high molecular weight

(HMW) DNA was loaded onto the gel alongside a Lambda-*Hind*III ladder and fosmid control DNA (Epicentre). Electrophoresis was carried out at 30 V for 18 hours after which marker containing lanes were cut from the gel and stained in TAE-EtBr buffer (EtBr 5 mg/ml, 0.5 × TAE) for 20 minutes. Markers were viewed at 302 nm UV in the AlphaImager 3400 (Alpha Innotech Corporation, San Leandro, CA), the 23 kb and 40 kb fragments were marked using a sterile toothpick. The gel was reassembled; sample DNA in the desired range for fosmid cloning was cut from the low melt plug and transferred to pre-weighed 1.5 ml eppendorf tubes. Agarose gel electrophoresis at 30 V for 18 hours through a 0.7% agarose gel was found to be sufficient for the clear resolution of the HMW DNA. The methodology employed for gel electrophoresis of the DNA avoided EtBr contamination and possible damage to DNA by UV and therefore allowed the correct size range of DNA to be excised from the gel. The DNA was recovered from the gel using GELase [section 2.2.5 a]. The casting of thinner gels also allowed more efficient purification of the DNA when using the GELase enzyme. Precipitated DNA was centrifuged at 13000 × *g*, the supernatant was removed and the pellet was washed with 70% ice-cold EtOH. After air-drying the pellet, the DNA was resuspended in 0.1 M TE buffer (pH 8) and the DNA concentration was quantified by fluorimetry [section 2.2.4].

2.3.4 <u>Cloning of high molecular weight DNA</u>

DNA was end-repaired to generate 5'-phosphorylated blunt-ended DNA fragments. The following reagents were combined on ice to a total volume of 80 µl; Sterile demineralised water, 8 µl end-repair buffer (330 mM Tris-acetate [pH 7.8], 660 mM potassium acetate, 100 mM magnesium acetate, 5 mM DTT), 8 µl 2.5 mM dNTP mix (2.5 mM each of dATP, dCTP, dGTP, dTTP), 8 µl 10 mM ATP, 4 µl End-repair enzyme mix (including T4 DNA Polymerase and T4 Polynucleotide Kinase), up to 20 µg insert DNA.

The reaction was incubated at room temperature for 45 minutes and enzymes were inactivated by incubation at 70°C for 15 minutes. DNA was precipitated using 0.1 vol 3 M NaOAc (pH 7) and 2.5 vol of ice-cold absolute ethanol. After centrifugation at 13000 × *g* for 20 minutes, the pellet was resuspended in 9 µl of TE buffer (pH 8) and the DNA was quantified by fluorimetry [section 2.2.4]. Six microliters of insert DNA at a minimum concentration of 250 ng was used in the ligation reaction. The following reagents were added to a total volume of 10 µl in the order given with gentle mixing after each addition; 1 µl 10 × Fast-Link ligation buffer, 1 µl 10 mM ATP, 1 µl CopyControl pCC1FOS vector (0.5 µg/ µl), 6 µl concentrated DNA insert, 1 µl Fast-Link DNA ligase. The reaction was incubated overnight at room temperature, transferred to 70°C for 15 minutes to inactivate the ligase enzyme and samples were stored at -20°C.

2.3.5 Packaging

The EPI300-T1^R plating stain is supplied as a glycerol stock. Cells were streaked onto an LBA plate. These cells were grown overnight and stored at 4°C. A single colony was inoculated into 5 ml LB broth and grown overnight, the day before the packaging reaction.

For the packaging reaction, 50 ml LB broth supplemented with 10 mM $MgSO_4$ was inoculated with 5 ml of an EPI300-T1^R overnight culture. This culture was incubated at 37°C until an OD_{600} of 0.8 – 1.0 was obtained. One tube of MaxPlax Lambda Packaging extract per ligation reaction was thawed on ice. Twenty-five microliters of packaging extract was added to a sterile 1.5 ml eppendorf tube, kept on ice and the rest was returned to -80°C. The 10 µl ligation reaction was added to the thawed extract and mixed carefully so as not to introduce any air bubbles. This reaction was incubated at 30°C for 90 minutes after which time a further 25 µl of extract was added. The reaction was incubated for a further 90 minutes at the same temperature. Phage dilution buffer (10 mM Tris-HCl [pH 8.3], 100 mM NaCl, 10 mM $MgCl_2$) was added to a final volume of 1 ml. Twenty-five microliters of chloroform was added and the reaction was stored at 4°C.

The phage was titered by making serial dilutions of the 1 ml packaged phage particles in PDB (Phage dilution buffer). The three dilutions were as follows;

1. 1:1 (10 µl of phage, no PDB)
2. 1:10^2 (10 µl phage in 990 µl PDB)
3. 1:10^4 (10 µl of 1:10^2 dilution in 990 µl PDB)

Ten microliters of each dilution was added individually to 100 µl of prepared EPI300-$T1^R$ host cells and the reaction was incubated 37°C for 30 minutes. One hundred microliters was plated onto LBA-cam and incubated overnight. The colonies were enumerated and the theoretical number of CFUs was determined using the following equation;

$$\text{Titre} = \frac{\text{(\# of colonies)(dilution factor)(1000 µl/ml)}}{\text{Volume phage plated (µl)}}$$

Phage particles were added to EPI300-T1^R cells at a ratio of 100 µl of cells for every 10 µl of phage particle and adsorption occurred at 37°C for 20 minutes. Infected bacteria were plated on LBA-cam and incubated overnight to select for CopyControl fosmid clones. Two milliliters of LB broth was pipetted onto a plate and cells were lifted off using a sterile spreader. The re-suspension was transferred to the next plate and cells were lifted off. This procedure was repeated on all plates and the final re-suspension was aliquoted into 2 ml screw-cap eppendorf tubes. Fifty percent [v/v] glycerol was added to a final concentration of 20% [v/v] and tubes were stored at -80°C. Five milliliters of the final re-suspension was stored at 4°C for screening. The control DNA supplied with the fosmid production kit was prepared according to the manufacturers specifications and the cloning efficiency was determined to be 3.02 x 10^6 CFU/ml.

2.3.6 Library verification

a. End-sequencing

Fosmids were extracted from 6 selected clones [section 2.2.6]. DNA was quantified by fluorimetry [section 2.2.4] and clones were sent with the T7–promoter primer (5' TAATACGACTCACTATAGGG 3') to the University of Stellenbosch sequencing facility for end-sequencing using the ABI PRISM 377 automated DNA sequencer.

b. Restriction analysis

Fosmids were extracted from clones conferring lipolytic activity as well as from randomly selected clones. Fifteen microliters of sample DNA was digested using *Eco*RI and *Hin*dIII [section 2.2.7]. The reaction was incubated for 2 hours at 37°C and digested products were electrophoresed overnight at 30 V on a 0.7% gel in order to determine different patterns of the clones and to estimate the average insert size of clones in the library [section 2.2.3].

2.4 Gene discovery

2.4.1 Functional screening of the library for lipolytic activity

Tributyrin agar (LBA, 1% [v/v] tributyrin, 1% [w/v] gum arabic, 0.01% [w/v] arabinose) supplemented with chloramphenicol was used to screen for clones expressing lipolytic activity. The library was plated onto the media and incubated at room temperature for 3 days and at 4°C for a further 5 days. Clones with zones of clearing around the colonies were selected and glycerol stocks were prepared. To verify that lipolytic activity was conferred by the fosmids and not the strain itself, fosmids were extracted [section 2.2.6] and electroporated [2.2.9a] into competent EPI-300 *E. coli* cells [section 2.2.8]. Halo formation was monitored by growth of clones on Tributyrin agar-cam for 3 days at room temperature and for a further 5 days at 4°C.

2.4.2 Transposon Mutagenesis

Tributyrin hydrolysing clones were mutated at random locations by Transposon mutagenesis (GPS® -Mutagenesis system, New England Biolabs, UK). The following reagents were mixed in a total volume of 18 µl; 2 µl 10 × GPS buffer, 1 µl pGPS3 donor, 100 ng Target DNA, variable volume sterile demineralised water.

One microliter of TnsABC Transposase was added to each tube and gently mixed. This assembly reaction was incubated for 10 minutes at room temperature. One microliter start solution was added and this strand transfer reaction was incubated at 37°C for 1 hour. After an inactivation step at 75°C for 15 minutes, the following reagents were added; 5 µl 10 × PI-Sce 1 buffer, 0.5 µl BSA, 18.5 µl dH_2O, 6 µl PI-Sce 1 (VDE, 6 units).

The reaction was incubated at 37°C for 1 hour and inactivated at 75°C for 15 minutes. One microliter was electroporated [section 2.2.9a] into competent EPI-300 *E. coli* cells [section 2.2.8] and 100 µl was grown on Tributyrin agar-cam and kan (20 µg/ml). The negative control for electroporation was 1 µl of a clone with a control insert. The plates were incubated overnight and then transferred to 4°C for 3 weeks in order to select lipase knock-out mutants. Glycerol stocks of these mutants were prepared and stored at -80°C.

2.4.3 Obtaining the full length gene by gene walking

Fosmids were extracted [section 2.2.6] from 3 knock-out mutants that were selected on tributyrin agar supplemented with cam (12.5 µg/ml) and kan (20 µg/ml). These were then sent to the University of Stellenbosch sequencing facility along with 4 µl of primer N and primer S (Table 4, page 45) at 3.2 pmol/µl (supplied in mutagenesis kit) to sequence upstream and downstream of the transposon insertion site. Once sequence information was obtained, TM3-N was used to design primers TM3-F and TM3-R (Table 4, page 45). The original fosmid was sent for sequencing and the sequence obtained was used to construct contigs using overlapping fragments. Another primer, LD1-RUS (Table 4, page 45) was designed to continue walking upstream and sequence overlap was used to further construct the contig. Two more primer walking steps were required; the first with LD1-RUS2 (Table 4, page 45) and the final primer used was LD1-RUS3 (Table 4).

2.4.4 Sequence analysis

Once the final nucleotide contig was constructed, EXPASY (Gasteiger *et al*., 2005) was used to translate the nucleotides into protein. GENEMARK (Besemer *et al*., 1999) was used to predict ORFs and genes in the sequence and were compared to other proteins in the database using BLASTp (Altschul *et al*., 1997). 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 based 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 predicts transmembrane 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 Codon Caltor was used to predict rare codon content. The translated nucleotide sequence was used for homology modeling using the Swiss-protein modeler 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*., 2002). The model was superimposed onto the template using the PyMol program.

2.5 Cloning of the lipolytic gene *LD1*

Primers (LD1-R-X1 and LD1-F-N1) [Table 4, page 45] were designed with restriction cut sites and were used to PCR amplify the gene from the original fosmid using a gradient PCR with a 1°C increase over a 5°C window [section 2.2.16.a]. The PCR reaction mixtures were placed in a Merck mastercycler gradient eppendorf machine and the following cycling conditions were used; Initial denaturation at 98°C for 30s, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 52-56°C for 30 s, and extension at 72°C for 30 s. The final elongation step was performed at 72°C for 2 minutes.

The PCR product was purified using the Illustra™ GFX ™ PCR DNA and gel band purification kit [section 2.2.5b] and digested with *Xho*I and *Nde*I [section 2.2.7]. One microgram of pET 28a vector was digested with the same restriction enzymes and the gene and vector were ligated, using a ratio of 1:2 vector to insert. The following reagents were used for the ligation reaction which was performed overnight at 18°C; 1 U of T4 DNA ligase in appropriate ligation buffer and 0.2 mM of each dNTP. The amount of insert required was calculated by the following equation assuming that 50 ng of vector DNA was used in the ligation reaction.

$$\text{ng insert required} = \frac{\text{ng vector x kb insert}}{\text{kb vector}} \text{ x } 2$$

Three microliters of the ligation mix was dialysed on a 0.02 nm nitrocellulose filter (Millipore). One and a half microliters of this mixture was electroporated into competent Genehog *E. coli* cells [sections 2.2.8 and 2.2.9a]. These cells were grown on LBA-kan and tributyrin agar-kan plates overnight. A negative control of circular, uncut pET28a was included. Colony PCR using primers LD1-R-X1 and LD1-F-N1 and restriction digestion of randomly selected clones from both the + and – plates was used to confirm clones containing insert. PCR reactions were placed in an Applied Biosystems thermocycler Gene Amp® 2700. and the following cycle conditions were used; Initial denaturation at 98°C for 30 s , followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. The final elongation step was performed at 72°C for 2 minutes. Glycerol stocks of clones containing the insert were prepared and stored at -80 °C.

Plasmids were extracted using the Invisorb® Spin plasmid mini two kit (Invitek, Berlin) according the manufacturers specifications, and sent for sequencing using the T7 promoter and T7 terminator primers, as well as primers (LD1-TM3 F, LD1-TM3 R, LD1-RUS1, LD1-RUS2, LD1-RUS3) [Table 4, page 45] designed for sequence elucidation by gene walking.

2.6 Expression of lipolytic gene *LD1*

Ten nanograms of plasmid LD1-pET +3 as well as 10 ng of circular pET28a vector was electroporated into competent ArcticExpress (DE3) *E. coli* cells and were grown overnight on tributyrin agar supplemented with gentomycin [20 µg/ml] and kanamycin [30 µg/ml]. Random transformants were selected from these plates for a small scale expression study. Colonies were grown overnight in 5 ml LB gent (20 µg/ml) and kan (30 µg/ml). Five hundred microliters of this culture was used to inoculate 10 ml LB gent (20 µg/ml) and kan (30 µg/ml) and grown at 30 ^{o}C until an OD_{600} of 0.6-1.0 was obtained. One millilitre was transferred to a sterile eppendorf tube and centrifuged at 10000 × *g* for 15 minutes. Cell pellets were air-dried and stored at -20 ^{o}C. The remaining culture was split in two equal parts and one was induced with 1mM IPTG. Both the induced and un-induced cultures were grown for 24 hours at 16°C after which OD_{600} measurement were recorded and 1 ml was centrifuged, dried and stored at -20°C. The volume of sample to be loaded on a 10 % SDS-PAGE gel was calculated using the following equation;

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

Samples were separated on a 10 % SDS-PAGE gel [section 2.2.12]. Ten nanograms of LD1-pET +3 as well as 10 ng of the circular vector was also transformed into Rosetta (DE3) pLysS and BL21 pLysS expressions strains [section 2.2.9a]. Twenty microliters of the transformation mix was plated on tributyrin agar supplemented with kan (30 µg/ml) and cam (34 µg/ml) and grown for 3 days at room temperature.

2.7 Crude enzyme assays

ArcticExpress (DE3) cell cultures were grown to OD_{600}- 0.6 at 30°C, induced with 0.5 mM IPTG and grown for 36 hours at 16°C. Ten milliliters of each culture was centrifuged at 13 000 × *g* and the pellet was lysed using Bugbuster [section 2.2.10]. The cell pellet was resuspended in 20 mM Tris-HCl (pH 8.5). For the enzyme assays, pellet and cell free extract were assayed with *p*-NP esters C_3 (propionate), C_8 (caprylate) and C_{12} (laurate) [section 2.2.10]. Increasing volumes (20 µl and 40 µl) of the supernatant was used with C_3 and C_8 with buffer volume adjusted and AU/min was measured. Ten microliters of the supernatant was boiled for 2 minutes at 95°C and activity with C3 was measured. The Bradford assay was used to calculate total protein content [section 2.2.14] and the rate was calculated from the data obtained.

2.8 Large scale expression of lipolytic gene *LD1*

A single culture of LD1-pET ArcticExpress (DE3) and LD1-pET Rosetta (DE3) pLysS was inoculated into 5 ml LB broths supplemented with appropriate antibiotic and grown overnight at 37°C. The culture was transferred to 500ml LB broth and grown at 30°C until an OD_{600} of 0.6 was obtained. 0.5 mM IPTG was added and cells were incubated at 16°C for 5 days. 50 ml aliquots of these cultures were pelleted and stored at -20°C. The supernatant of these aliquots were acetone precipitated [section 2.2.13] and fractions were electrophoresed using 10 % SDS-PAGE [section 2.2.12]. Cells were pelleted by centrifugation at 10 000 × *g* for 10 minutes at 4°C. After removal of the supernatant, pellets were lysed using bugbuster [section 2.2.10] and the resulting cell pellet, cell-free extract and cell lysate fractions were analysed by SDS-PAGE. The supernatant was sterilised through a 0.22 µm filter and used for His-tag purification [section 2.2.11]. All fractions were collected and 10 µl of each was electrophoresed on 10 % SDS-PAGE [section 2.2.12].

The eluted fraction was transferred to a dialysis cassette and dialysed overnight against 2L of buffer A (20 mM Tris-HCl [pH 8.5], 1% [v/v] DTT). The dialysis cassette was moved to a second buffer B (20 mM Tris-HCl [pH 8.5], 10% [v/v] glycerol) and dialysed overnight at 4°C. The recovered fraction was stored at 4°C and used in a subsequent preliminary enzyme assay with buffer B as a control.

The cell pellet of a 50 ml aliquot was treated with bugbuster [section 2.2.10]. The pellet was weighed, gently resuspended in 10 ml of wash buffer (20mM Tris-HCl [pH7.5], 100mM EDTA, 10% [v/v] Triton-X-100) and centrifuged at 10000 × *g* for 10 minutes. After repeating the wash step the pellet was gently resuspended in 2 ml solubilisation buffer (500 mM CAPS [pH 11], 20% [v/v] N-Lauroylsarcosine) and incubated at room temperature for 1 hour. The supernatant was transferred to a dialysis cassette and dialysed in buffer A and B as previously described. Ten microliters of the recovered fraction was electrophoresed on a 12% SDS-PAGE [section 2.2.12]. The recovered fraction was stored at 4°C and used in a subsequent preliminary enzyme assay with buffer B as a control.

2.9 Preliminary enzyme assays

The His-tag purified fraction of LD1-pET Rosetta and the re-solubilised fraction from the LD1-pET Rosetta pellet [section 2.8] were used in the preliminary enzyme assays [section 2.2.15] and the protein concentration was quantified using the Bradford assay [section 2.2.14]. Buffer B was used as a control as both fractions were dialysed in this buffer. After addition of 10µl of resuspended pellet or supernatant, the formation of product was measured. The volume of buffer was adjusted accordingly when increased volumes of enzyme or substrate were used. Activity was calculated using the following equation.

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

ε is the extinction co-efficient of p-nitrophenol and was experimentally determined by Dr. C. Heath and X. P. Hu as 13 900 $mM^{-1}.cm^{-1}$. A is the rate of the enzyme reaction based on Vmax and the volume of enzyme used in the 1ml assay. L is the path length of light through the cuvette and has a value of 1.

2.10 Prokaryotic diversity study

2.10.1 PCR amplification of bacterial 16S rRNA

One aliquot of the stored library was thawed overnight at 4°C, inoculated into 10 ml LB-cam12.5 and incubated overnight. Fosmids were extracted according to method previously described [section 2.2.6] and the DNA was quantified using fluorimetry [section 2.2.4]. DNA was extracted from EPI-300 *E. coli* cells using a crude boiling method [section 2.2.1b] and used as a positive control for the PCR. 16S rRNA universal bacterial primers (341 F-GC and 534r) used for DGGE analysis were used to amplify possible 16S rRNA gene sequences in the library [section 2.2.16.b]. The following PCR cycle conditions were used: Initial denaturation at 95°C for 4 minutes, followed by 20 cycles of denaturation at 94°C for 45 s, annealing at 65°C for 45 s (touch-down from 65°C to 55°C), and extension at 72°C for 1 minute. Another 20 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 minute. The final elongation step was performed at 72°C for 20 minutes.

2.10.2 PCR amplification of archaeal 16S rRNA

DNA prepared for bacterial 16S rRNA was used for an archaeal study. An archaeal 16S rRNA gene fragment cloned into an *E. coli* stain was used as the positive control for PCR with universal archeal primers [section 2.2.16.b]. Instead of extracting the DNA from the control, single colonies from a streak plate incubated overnight and suspended in 5 µl ultra high quality Millipore water, were used.

a. Primers Ua1204R and A571F

The first primer set was used to amplify possible nanoarchaea signals in the fosmid library. 1 mg/ml of BSA and 1 mM $MgSO_4$ was added to the reaction mixture. The following PCR cycle conditions were used: Initial denaturation at 94°C for 5 minutes, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 45 s. The final elongation step was performed at 72°C for 7 minutes.

b. Primers A3FA and AB927R

PCR mixtures were prepared as described. The following PCR cycle conditions were used: Initial denaturation at 94°C for 4 minutes, followed by 25 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 45 s, and extension at 72°C for 75 s. this was followed by a further 10 cycles denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 75 s. of The final elongation step was performed at 72°C for 5 minutes.

2.10.3 DGGE (denaturing gradient gel electrophoresis) of PCR products

Low (30%) and high (70%) denaturing solutions were prepared according to the manufacturers specifications from 0% (2% [v/v] 50 × TAE, 20% [v/v] acrylamide:bis-acrylamide [37.5:1]) and 100% denaturing (2% [v/v] 50 × TAE, 20% [v/v] acrylamide:bis-acrylamide [37.5:1], 42% [w/v] urea, 40% [v/v] formamide) stock solutions for 9% polyacrylamide gels. Sixteen microliters of TEMED and 160 µl of 10% [w/v] APS was added to each gel solution and mixed by inverting. The mixer was used to pour the gel which was polymerised for 1-2 hours.

Twenty microliters of each sample was mixed with 5 µl of loading dye. Samples were loaded into the wells and electrophoresed for 16 V/hours at 60°C. The gel was removed from the core unit and stained for 15 minutes in 1× TAE buffer containing 0.5 µg/ml EtBr, de-stained in 1 × TAE for 20 minutes and viewed in the UV illuminator.

Chapter 3: Results and discussion

3.1.1 Metagenomic fosmid library construction

Soil samples were collected from beneath the mummified carcasses of several seals and used for extraction of total community DNA and subsequent fosmid library production. Table 5 gives the geographical locations of these carcasses in the Dry Valleys of Antarctica. At each step in production of the fosmid library, large quantities of DNA were lost. In this study, efficient library construction therefore involved the isolation of large quantities of high molecular weight DNA from the environment of interest. This required a relatively large amount of sample and an efficient method of cell lysis and DNA extraction. The Zhou method used to extract total community DNA from the soils is a chemical-based extraction method that limits the shearing of DNA as compared to robust mechanical methods (Bertrand, 2005). Electrophoretic analysis of extracted DNA showed that chemical lysis liberated high molecular weight DNA (Figure 9). More efficient recovery of DNA from the agarose gel was obtained when thinner gels were cast, possibly due to a higher DNA to agarose ratio. The library was constructed using the fosmid CopyControl pCC1FOS vector in the EPI300-T1^R *E. coli* strain. A library of 7900 transformants was prepared.

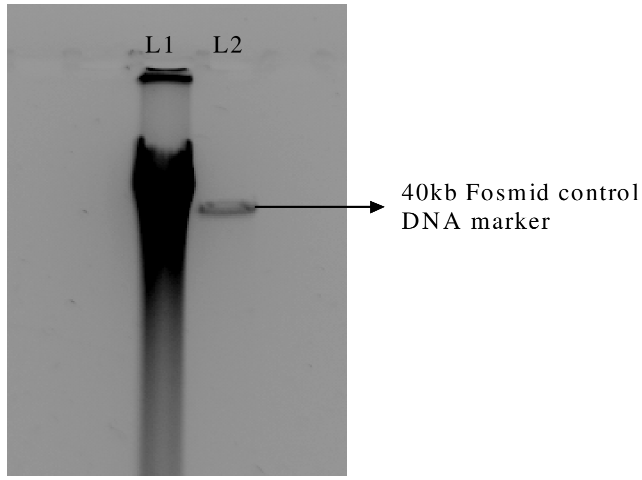


Figure 9 Agarose gel electrophoresis of extracted metagenomic DNA. Lane 1) Sample of extracted HMW DNA. Lane 2) Fosmid control DNA (40kb).

3.1.2 Library verification

Fosmid extractions were performed on six selected clones (section 2.2.6). End-sequencing (section 2.3.6 a) showed that the inserts cloned into the fosmid vectors were of prokaryotic origin (Table 6). The bacterial classes represented in these sequences include the γ-Proteobacteria (*Pseudomonas*, *Xanthomonas* and *Psychrobacter*) and the Actinobacteria (*Kineococcus*). This is consistent with a recent study of bacterial diversity of soils beneath seal carcasses in the Antarctic Dry Valleys, which showed similar dominant microbial classes (Robson, Cowan, Cary. Unpublished). Restriction endonuclease digestion (section 2.2.7) of 17 randomly selected clones was used to assess the average insert size of the metagenomic library, using *Eco*RI and *Hind*III (Figure 10). Inserts ranged in size from 21 kb to 52 kb with an average insert size estimated to be 29 kb (n=16). Although fosmid vectors generally contain inserts of uniform size, there appears to be some flexibility in the size range of DNA that was cloned.

Library verification by restriction enzyme digestion is a relatively crude method that is dependant on adequate separation of DNA fragments through low percentage agarose gels. Efficient estimates of insert size were obtained by these methods 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.

The fosmid library generated in this study contained approximately 7900 clones with an average insert size of 29 kb. This compares well to other studies where metagenomic fosmid libraries were generated from environmental samples; e.g. in a fosmid library constructed by Pang and co-workers using forest topsoil, 3624 clones were obtained (Pang *et al*., 2008). Similarly, 7000 clones were obtained in a fosmid library using Baltic Sea sediment as the sample (Hårdeman *et al*., 2007).

Coverage of 2.29×10^7 bp, equivalent to 74 prokaryotic genomes, was obtained in the library constructed in this study. Based on the assumption that a given environment contains 2000 genomes with an average genome size of 3100 kb (Gabor *et al*., 2004; Sandaa *et al*., 1999), approximately 3.7% of the total metagenome was represented in the library. Although the representation was low, a three-fold greater coverage (when compared to that of current culturing methods) was obtained. The construction of large-insert libraries is beneficial, as they cover more sequence space than conventional plasmid libraries. In order to obtain the same coverage, approximately 7.7×10^4 clones with an insert size of 3 kb would be required in a plasmid library. This value neglects to take into account the greater number of partial ORFs that would be cloned from small inserts. Therefore, the actual number of clones that would be required would be substantially larger.

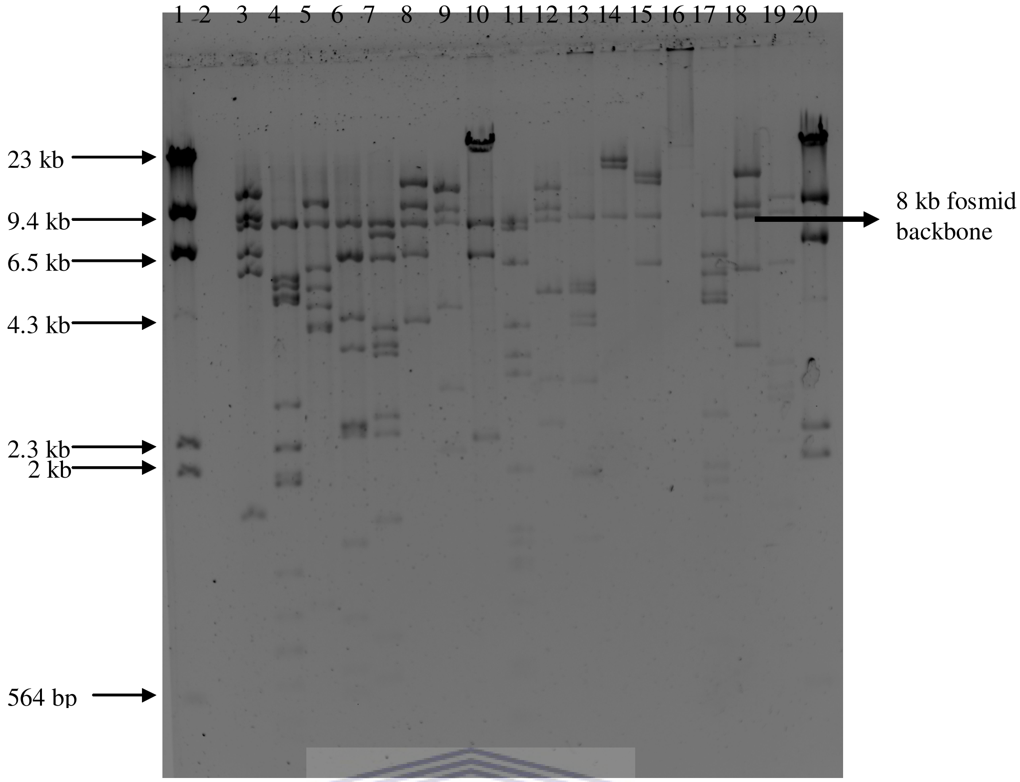


Figure 10 Restriction endonuclease digestion of 17 randomly selected fosmid clones to estimate average insert size. Lane 1 and 20) Lambda-*Hind*III DNA marker. Lanes 3-19) Recombinant fosmids digested with *Eco*RI and *Hind*III. The band sizes of the molecular weight marker (phage lambda DNA digested with *Hind*III) and the 8 kb fosmid backbone are indicated.

Table 6 Nucleotide end-sequences of selected fosmid clones and identity of the closest match.

Clone name	**Nucleotide sequence**	**% identity**	**I.D. of nearest match** (Accession number)
MVS-Ld1	CATGGAGCGTTCTGCATCACGTAAGCGTTGTGCGGCGCCGCTGTAATTTTTATTGTCTATATCTAATAAGGCTAAACCAAAATGGCCGT AAGCAGCGCGGTGGTCTCCCCAACGCAAGGCTTCTTGCAGACCCTCGGTGAAATAGATATAGGCTTGATGGTTATGGCCTTGGCGTAG TTTTAGTTGGCCGAGGCGCAATTGTAAGCGGCCTAACACAGGGCTTTGAATAGTGTGATCTTCGCCTAATAATGCACAGGCTCTTTCA AGAATGCTCTCTGCATCTTTAAATGAGCCTCTATGCTCCAATAGTTGAGCATGATCCAGCTCTAAATAGGCTTCAAATAATAGGCTCTT TTGTTGGCGAGAAAGACGCAATGCAGTGCGGTTAAGAGTGCGGGCTAAGTTTAAGTCACCGTTGAGCATAGCTTGCTGAGTGAGGGC GCTATAACAGGTTAAGCGTATTTCCCAAAAATCATCAGATAACCATTGAATTGCGGCATGTAAATGTTTATTGGCAACATCAGTTATC CCTCGCAGGTGTGCGATCCAGCCAGTAAATCCTTGCC	52%	*Pseudomonas fluorescens* Pf0-1 ATP-dependent transcriptional regulator (CP000094.1)
MVS-Ld3	CCTATGCCGGCCCGATCGGTGCGGAGTTCATGCACATCGCCGATGCGCCGCAGCGCCGTTGGCTGTATGAGCGCATGGAAAAGGCCG CCGGCCAGTACGGTGTCAGCAAAGACGACAAGCGCCGGATCCTGGATCGCCTGACCGCGGCCGAAGGTCTGGAGCGCTACCTGCACA CCAAGTACGTGGGTCAGAAGCGGTTCTCGCTGGAAGGCGGGGATGCACTGATCCCGTTGATGGACACCACCATCCGCCGCGCCGGCG AGCAGGGGGCCAAGGACGTGGTCATCGGCATGGCCCACCGTGGTCGGCTGAACGTGCTGGTCAACACCCTGGGCAAGTCGCCACGCA AGCTGTTTGACGAGTTCGAAGGCAAGTTCGACCTCAACGAGCTGGCCCACGCCGGTGACGTGAAATACCACATGGGTTTCAGTGCGG ACGTGGCCACACCCGGCGGTCCGGTCCACCTGGCGCTGGCATTCAACCCCTCCCATCTGGAGATCGTCAATCCGGTGGTGGCCGGTTC GGTGCGCTCGCGCCAGACCCGCCGTGGCGGGGATGACAGCCGCAAGCAGGTGAT	77%	*Xanthomonas campestris* pv. *campestris* strain B100 Oxoglutarate dehydrogenase (AM920689.1)
MVS-Ld4	TGCGTTATTAACCAGTACCACACTAGCCACGCCGCAACTCATCAGTCAATTGATGGAGCTGGAATTATTAGGCGCTATACATGAGCAA GGCGGGCGTTATTTACGCATTTAACATGAATGCATTCGATTCAATAGCGTTTTTGCCAGTGTTTAAGTTTTTAGTTTAGGTCTTTAGCTT AAGTTCTTAGCTCAAATTTTTAGATACCTTAACCTCCTTATCACCTGTATCATTAATATCTACTTTTTCATCATTCTAAATCATCTGACTT ATGCATATATCTGCGTCCCTCATTACTGACTCTGTCGTTCAGGCTGCTAATTGGCTCCAAAAAGGGCAACTACTTGCTTACCCTACTGA AAGCGTTTGGGGTATTGGCTGCAATGCTTATGACAAAGAAGCGGTTCAGCGAATTCTTGATATCAAACAGCGCCCACCAGCTAAAGG GATGATTGTAGTGACGGATAGTGCCGCTCGGATTGCACCGCTATTAGAACGCTTAAATGACCAACAGCGCCACACTGTATTAGAAAGC TGGAGTCATTCACCCCAAGCGCTAGCGCAACAAGCGCACACTTGGTTATTACCGATAAATGCACCCCTAAAAATACCGATTCCCTCTT GGGTCACAGGTGCACACGATAGTATTGCAGTACGGGTAATTGCTCATCCA	66%	*Psychrobacter arcticus* Conserved hypothetical protein (CP000082.1)
MVS-Ld5	CACTGACCGCGGCCGAGGTCTGGAGCGCTACCTGCACACCAAGTACGTGGGTCAGAAGCGGTTCTCGCTGGAAGGCGGGGATGCACT GATCCCGTTGATGGACACCACCATCCGCCGCGCCGGCGAGCAGGGGGCCAAGGACGTGGTCATCGGCATGGCCCACCGTGGTCGGCT GAACGTGCTGGTCAACACCCTGGGCAAGTCGCCACGCAAGCTGTTTGACGAGTTCGAAGGCAAGTTCGACCTCAACGAGCTGGCCCA CGCCGGTGACGTGAAATACCACATGGGTTTCAGTGCGGACGTGGCCACACCCGGSGGTCCGGTCCACCTGGCGCTGGCATTCAACCCC TCCCATCTGGAGATCGTCAATCCGGTGGTGGCCGGTTCGGTGCGCTCGCGCCAGACCCCGCCGTGGCGGGGATGACAGCCGCAAGCA GTGATCCCGATCAAGATCCATGGCGACGCGGCATTCGCAGCTAGGCGTGGTGATGGA	79%	*Xanthomonas campestris* pv. *campestris* Oxoglutarate dehydrogenase (AM920689.1)
MVS-Ld7	GTTGCCGCTTGAGCGAAGTGAGTCGGCCTTCTTGGAAGCGGTGAGTGCCTTGTGGCCCCACTCGTGGGAGGCGGCGATGTCGTCGGCA TGGTCGGCCTCGGCCAAGCGCAAGTTGCCGATGGTCTGGGCAACGGCGTCTTCGGCTTCGGCAATGCTGTCGGTGTAGTCGCGAACGA GCTGGTCGAGCATTTTCTCGGGATCCTCGGCGCGGTCGAGGGCCGCGTTGATGTTCGCGCGGGTCATTTGCGTGACGCGACCGAGGAT CGATTGCTTCTGGGCCATGAGGTCCCCTTCTATGAATGCCGAATGTATTACGTCGATCTTGGCGCATGCGGCGTCTTGACACAATGACG GGGCAACGAACGTCGACGGCTCACGCTCAACTGAGCTCTTCCTCCGAAATGACACCGGCGCGGCGTACTCCGTCGGTGATGGCCAGA GTGCGCACGATGTCGGCGGGATTTCCGTTGCCCGAGAGGTCCAGATAGACCGAGGTGACGGCCGCGTCACCGTCGGGGGTCTCATGC	77%	*Kineococcus radiotolerans* Phage shock protein A (CP000750.2)
MVS-Ld13	CGTTACCTATTTTGTCGCGTCTTGATGCTGCGCGTCGTGAAGCGCAAGTGTTGGTGCTCGCGCCGACCCGTGAGTTGGCTCTGCAAGTG GCCACTGCGTTTGAAAGCTTTGCTGCACAGATGCCTAGCGTAAATGTTGTTGCTATCTACGGTGGCGCGCCAATGGGCCCGCAACTGA AAGCAATCCGTAATGGCGCGCAAGTGATTGTTGCAACGCCAGGTCGTTTAGTTGACCACTTGAGCCGTAATGGTGGCTTACTCTCGAC GATTAAATTCTTAGTGCTTGATGAAGCTGACGAGATGTTAAAGCTCGGCTTTATGGATGACTTAGAAGTTATTTTCAACGCCATGCCTG ATGAGCGTCAGACTGCATTGTTCTCAGCAACATTGCCTGCATCGATTCGCGGCATTGCGGAAAAACACTTGCGTAACCCGCAGCAAGT TAAAATTGCTAGCAAAACGCAAACTGTTGCGCGTATTGATCAAGCTCACTTGATGGTGCATGCGGATCAAAAAGTAAATGCTATTTTA CGTTTGCTTGAAGTTGAAGACTTTGACGCAATGATCGGTTTTGTACGTACCAAACAGGCGACTTTAGATATCGCAGCAGCACTTG	70 %	*Pseudomonas putida* W619 DEAD/DEAH box helicase domain protein (CP000949.1)

3.1.3 Prokaryotic diversity study

3.1.3.1 PCR amplification of archaeal 16S rRNA

No archaeal 16S rRNA gene amplification was obtained from the fosmid library, even though two different primer pairs were used [Ua1204 R, A571Fb and AB927R, A3FA] (Table 4, page 45) and despite the archaeal controls showing a positive result. Archaeal DNA may in fact be present, but in quantities too low to be detected.

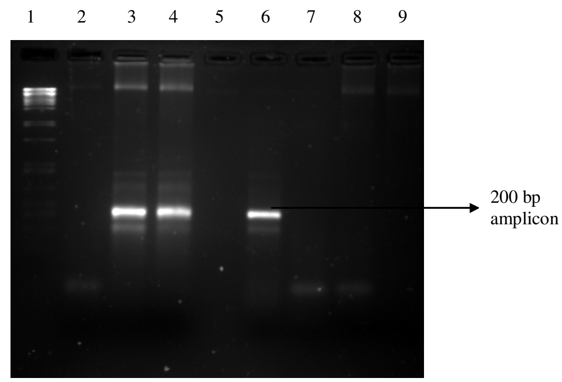
3.1.3.2 PCR amplification of bacterial 16S rRNA

Bacterial diversity of the fosmid library was assessed by the PCR amplification of 300 ng of template DNA using primers 341FGC and 534r, designed specifically for DGGE. These primers were designed to target conserved regions at positions 341 and 534 of the bacterial 16S rRNA gene, yielding a 200 bp PCR product. As seen in Figure 11, a single amplicon of approximately 200 bp was observed following agarose gel electrophoresis. No amplification was observed in the negative control.

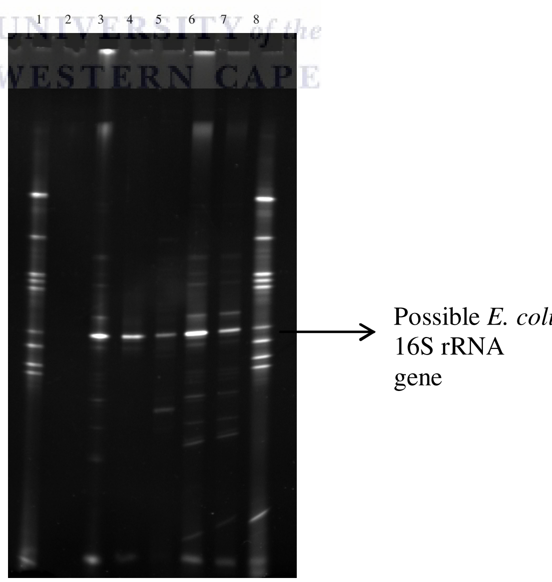
3.1.3.3 Denaturing gradient gel electrophoresis (DGGE)

Denaturing gradient gel electrophoresis was performed using the products obtained from PCR amplification of bacterial 16S rRNA genes (341FGC and 534r). Figure 12 shows the different operational taxonomic units based on the different denaturation profiles observed. Twelve dominant phylotypes were observed by DGGE analysis (assuming that each band represents one OTU).

However, it is important to note that one organism could contain multiple 16S rRNA genes with variable sequence. Since no sequence analysis was performed, the identities remain unknown and they can not be linked to any particular fosmid insert in the library.



<u>Figure 11</u> PCR amplification of bacterial 16S rRNA using primers 341 F-GC and 534 r. Lane 1) DNA molecular marker, lambda-PstI digested DNA. Lanes 3 and 4) 16S rDNA amplicons from the fosmid library. Lane 6) 16S rDNA amplicon from *E. coli* genomic DNA. Lane 7) Negative control. Lane 8) Reverse primer control. Lane 9) Forward primer control.



<u>Figure 12</u> DGGE profile of 16S rRNA gene content of the Dry Valley soil metagenomic library. Lanes 1 and 8) DGGE marker. Lane 2) Negative control. Lanes 3, 5, 6 and 7) 16S rRNA gene component of the fosmid library obtained from different PCR reactions. Lane 4) 16S rRNA gene of the *E. coli* host.

3.1.4 Activity-based screening for lipolytic clones

The metagenomic library was screened for functional tributyrin hydrolysing activity on tributyrin indicator plates as described in chapter 2.4.1. The plates were routinely monitored during the three day incubation at 4°C for the presence of zones of clearing around the colonies (Figure 13). Thirteen clones (designated LD1 through LD13) formed halos on the indicator plates and each of these were streaked to obtain pure cultures on Luria-Bertani agar supplemented with chloramphenicol (12.5 μg/ml). Fosmids were extracted from these clones and were analysed with *Eco*RI and *Hind*III restriction enzymes in order to eliminate replicate clones and estimate the average insert sizes of the lipolytic clones (Figure 14). Glycerol stocks (20% [v/v] glycerol) of the clones were also prepared and stored at -80°C.

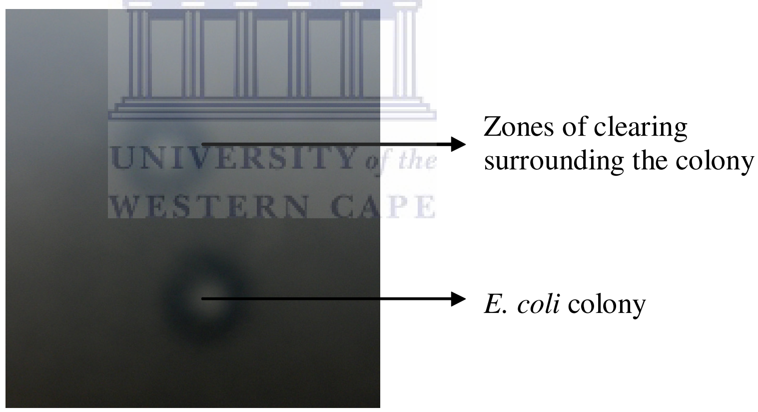


Figure 13 Growth of recombinant *E. coli* colonies on tributyrin agar. A zone of clearing around the colonies indicates hydrolysis of the lipid substrate.

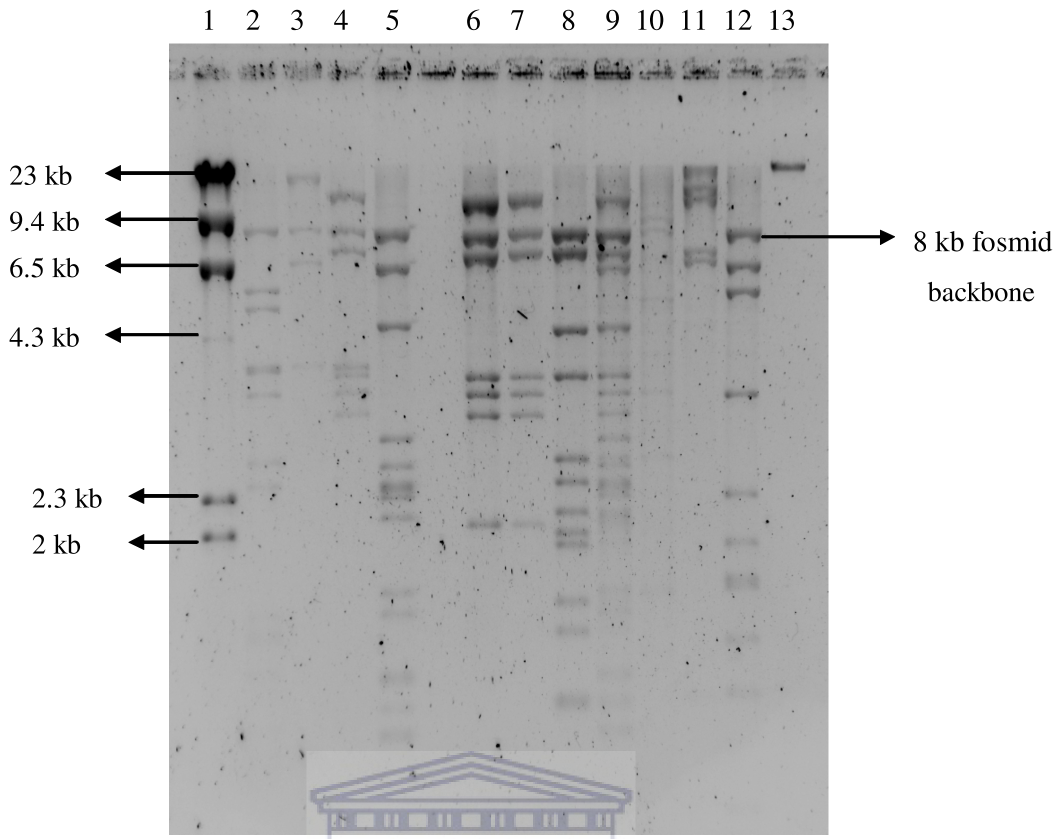


Figure 14 Restriction profiles of the tributyrin hydrolysing clones that formed halos during activity-based screening of the metagenomic library on tributyrin indicator plates. Lane 1: Lambda-*Hind*III DNA marker; Lanes 2-13: Halo-forming recombinant fosmids digested with *Eco*RI and *Hind*III.

The average insert size estimated for the halo-forming recombinant fosmid clones was 31.6 kb. This relates well to the 29 kb average insert size obtained from restriction digestion of randomly selected clones. Clones LD5 and LD6 (Figure 14; lanes 6 and 7, respectively) showed identical profiles, indicating that these two clones contained the same insert and were probably duplicate clones. The other 11 clones showed different patterns, indicating that these clones contained different inserts. However, this does not necessarily confirm that the genes responsible for halo formation on the indicator plates were not replicates.

The following equation was derived by Gabor *et al*, (2004) to assess the number of clones (N_p) required in a metagenomic library in order to recover a target gene with a certain probability (P).

$$N_p = \frac{\ln(1-P)}{\ln\left[1 - \frac{I\ X}{G.\ z.\ c}\right]}$$

Where I is average insert size in the library, X is the size of the gene of interest and G is the average genome size [approximately 3100 kb] (Gabor *et al*., 2004). The correction factor, c, is dependant on the three possible modes of expression (vector independent gene expression, vector dependent expression or expression as a transcriptional fusion). z is the number of different species in a sample, assuming even distribution.

Assuming the following for the fosmid library generated; where vector independent gene expression occurs, such a in large insert libraries (Lorenz *et al*., 2005), a correction factor of 1 is used; the average insert size was estimated at 29 kb; an average size of a lipolytic gene is 1 kb; the average genome size is 3100 kb and the number of species is 2000 (Gabor *et al*., 2004; Sandaa *et al*., 1999). The number of bacterial species z is, however, an estimate and would need to be experimentally determined for the Antarctic soil samples used in this study. In order to obtain one positive hit (with probability of 0.9) during functional screening, the library should theoretically contain 5.6×10^5 clones. However, clones conferring lipolytic activity were still obtained even though the theoretical number of clones required was clearly larger than the number obtained in the fosmid library.

This could be attributed to the sample; seal carcasses contribute very important nutrients, such as proteins and lipids, to the microbial communities in the Dry Valleys of Antarctica (Smith *et al*., 2006). Although decomposition of these carcasses occurs at a very slow rate, lipids are nonetheless leached into the soils. This process may serve as a form of natural pre-enrichment for lipolytic activity in these soils thereby increasing the hit-rate. Microbes which have the ability to utilise these lipids as a nutrient source would have a clear advantage in the habitat. For the purpose of this study, the lipolytic clone LD1 was chosen for further analysis. This clone was the first to produce a halo on the indicator plates after only 1 day incubation at 4°C. The insert size of this particular clone was estimated to be 21 kb.

The extracted fosmid from LD1 as well as a fosmid, designated C12, extracted from the control library were prepared, electroporated into competent EPI300-T1^R *E. coli* cells (sections 2.2.8 and 2.2.9a) and inoculated onto the tributyrin indictor plates in order to verify that the fosmid insert was responsible for the observed activity. These plates were incubated in the same manner as in the initial screening (section 2.4.1). As shown in Figure 15, the control fosmid did not produce any halos on the indicator plates whereas clone LD1 formed clear zones around the colonies. This indicated that tributyrin hydrolysing ability was conferred by the fosmid insert of clone LD1.

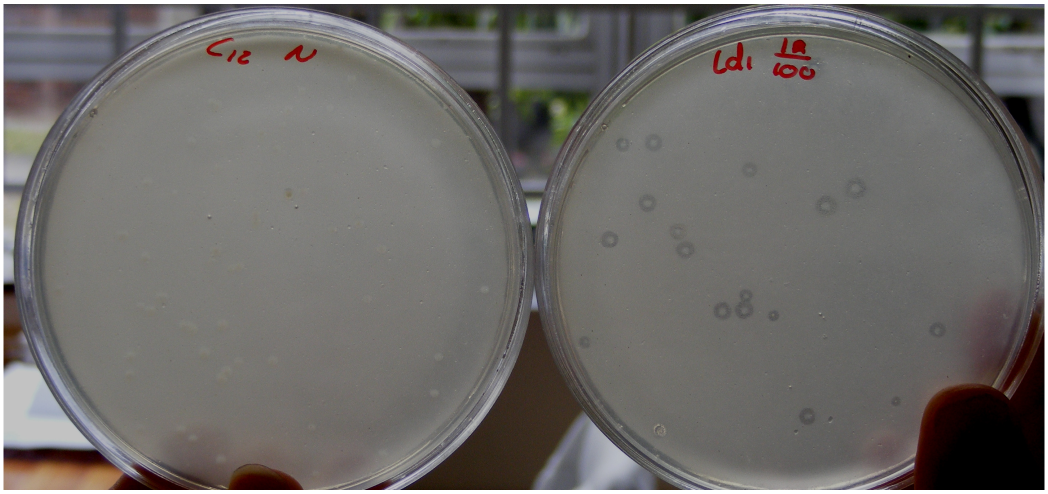


Figure 15 Fosmid clone with lipolytic activity (right) and control fosmid clone with no lipolytic activity (left) on tributyrin agar indicator plates.

3.1.5 Transposon mutagenesis

Transposon mutagenesis is an efficient method that may be used as a starting point to obtain a gene sequence. The 1 kb transposon inserts randomly into the DNA of the host and the vector. If critical DNA elements (such as antibiotic resistance genes) are mutated, the host will not grow. The transposon also confers kanamycin resistance, allowing the simple selection of mutants on indicator plates containing the correct concentration of antibiotic. Using the supplied primers, selected knock-out mutants are sequenced thereby allowing subsequent primer walking to obtain the full length gene.

The fosmid of lipolytic clone LD1 was extracted (section 2.2.6) and randomly mutated using the GPS® -Mutagenesis system (section 2.4.2). Several knock-out mutants which lost their ability to form halos on tributyrin agar were selected and stored as glycerol stocks at -80 °C. Fosmids from three of these were extracted (designated TM2, TM3 and TM4) and were sequenced using the primers N (5´ACTTTATTGTCATAGTTTAGATCTATTTTG 3') and S (5´ATAATCCTTAAAAACTCCATTTCCACCCCT 3’). The sequences that were obtained for these three clones contained overlapping nucleotide fragments and this showed that in these three mutants, the transposon had mutated in similar regions of the insert DNA, resulting in a loss of lipolytic function.

3.1.6 Obtaining the gene sequence of lipolytic clone LD1

One of the sequences (designated TM3-N) obtained from mutagenesis was used to design primers LD1-TM3 F and LD1-TM3 R (Table 4, page 45) which target the nucleotide sequence flanking the transposon. Sequencing was performed on the original fosmid using these primers and the contig was assembled from overlapping fragments. New primers were designed from each DNA fragment obtained and the original fosmid was sequenced with the synthesised primers. Several rounds of single stranded primer walks were required and in order to eliminate any error, a minimal overlap of 100 bp was used to obtain the sequence of lipolytic gene *LD1*. Figure 16 shows the full length sequence (including additional upstream and downstream data) obtained by primer walking as well as the binding sites of the designed primers. The full length sequence obtained was 2475 bp with a G+C content of 45.94 %.

3.1.7 Sequence analysis

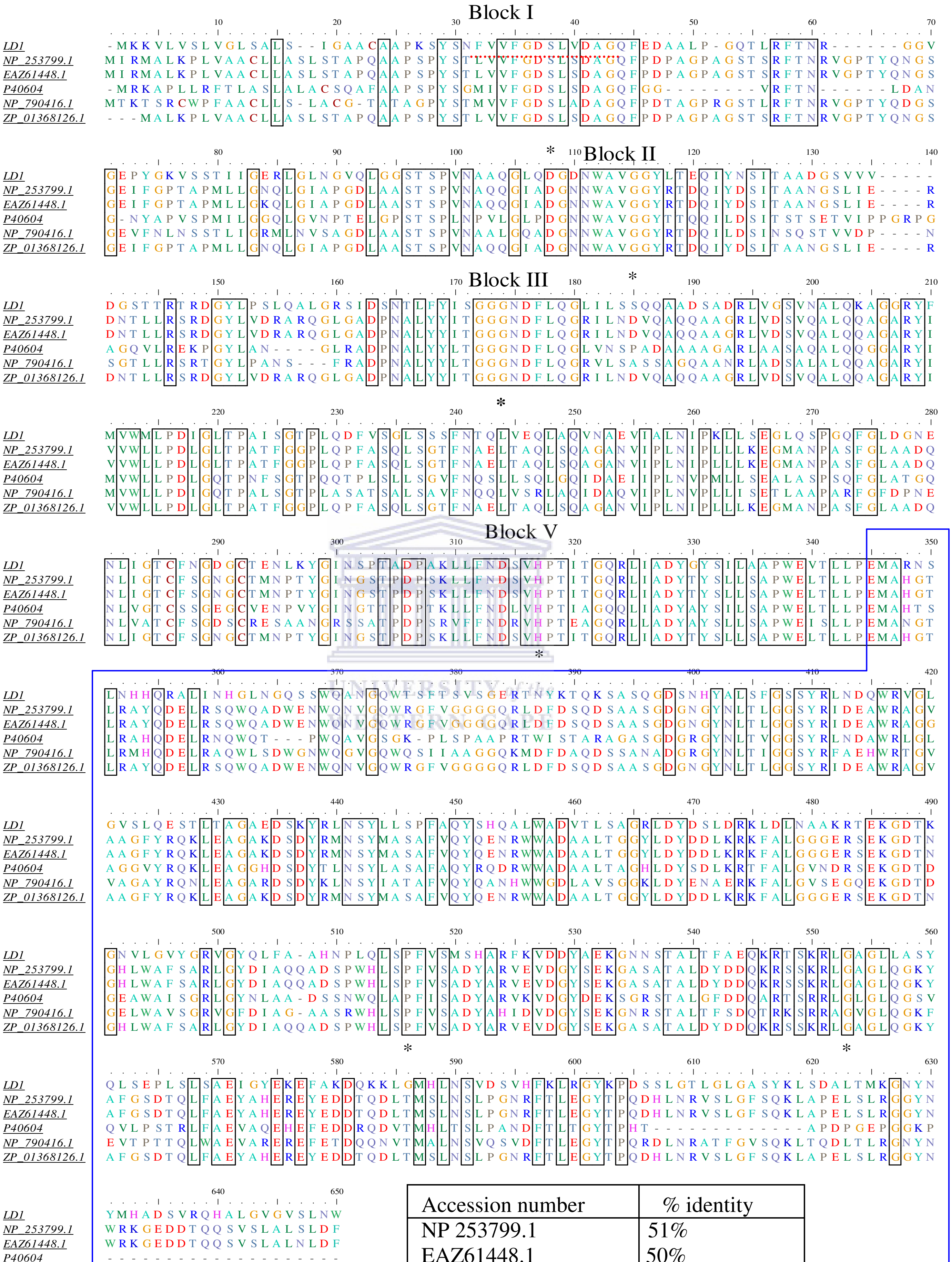
Metagenomic gene discovery using functional screening generally employs the use of bioinformatics tools to predict basic structural information based on sequences obtained. Using more than one tool is advantageous as it allows more accurate predictions and comparisons.

A range of bioinformatics tools were used to analyse the sequence obtained (section 2.4.4). The EXPASY translate tool (www.expasy.org) was used to translate the nucleotide sequence to an amino acid sequence. ORFs within the contig were predicted using GENEMARK (v 2.6)

Heuristic model. One ORF was predicted in the sequenced region. This 1902 bp sequence encoded a putative protein 634 amino acids in length on the positive strand of the sequence. BLASTp was used to compare this sequence to other proteins in the database. The putative protein showed 51% amino acid identity to Esterase A of *Pseudomonas aeruginosa* PAO1, a opportunistic pathogen commonly isolated from temperate soils. Using hits identified by BLASTp analysis, a five protein multiple sequence alignment was created using ClustalW. Conserved regions were identified based on this sequence alignment (Figure 17).

```
5’ TGCACTAAAATATTCTTTGTAGTAATTAATGCTCAGTAATAAAAATAATAATACGGAGTTA
CGCTATGAAGAAGGTACTGGTAAGTTTGGTAGGTCTTTCGGCTTTAAGTATTGGTGCGGCATG
TGCCGCGCCTAAATCCTATAGTAATTTTGTTGTGTTTGGCGATAGTTTGGTTGATGCAGGACAG
TTCGAAGATGCGGCTCTGCCGGGGCAAACGTTACGCTTCACCAACCGAGGGGGGGTTGGTGA
GCCGTATGGCAAAGTGTCATCCACTATAATTGGTGAGCGTTTAGGCTTAAATGGTGTTCAATT
AGGAGGTTCTACTTCTCCAGTTAATGCCGCGCAAGGTTTGCAGGATGGTGATAACTGGGCAGT
TGGTGGTTATCTTACGGAGCAAATTTACAACTCAATTACAGCTGCTGATGGTTCTGTCGTTGTA
GACGGTAGTACAACACGAACGCGTGATGGGTATTTACCAAGCTTGCAGGCACTAGGGCGCTC
GATTGACTCCAATACATTGTTTTATATCAGTGGCGGCGGTAATGATTTTTACAAGGATTGATT
TTAAGTTCTCAGCAAGCTGCAGATTCAGCTGATCGATTGGTGGGCAGTGTCAACGCTTTGCAA
AAAGCAGGGGGTCGCTATTTTATGGTGTGGATGCTGCCTGATATTGGGCTTACTCCGGCTATC
TCTGGCACGCCTTTGCAGGATTTTGTCTCTGGGTTGTCGAGCAGTTTTAATACTCAACTTGTTG
AGCAATTAGCACAGGTAAATGCTGAGGTTATTGCGCTGAATATTCCAAAGTTGCTCAGTGAAG
GGCTGCAGAGTCCAGGACAGTTTGGTTTAGATGGCAATGAGAATCTTATCGGCACTTGCTTTA
ATGGCGACGGCTGTACGGAGAATCTAAAGTACGGCATTAACAGCCCTACAGCTGATCCAGCT
AAGCTGTTATTTAATGATAGTGTGCATCCCACTATTACCGGACAGCGTTTGATTGCTGATTATG
GCTATTCAATTTTAGCTGCACCTTGGGAAGTTACTTTGTTACCAGAGATGGCGCGCAACTCACT
CAATCATCATCAAAGAGCTCTGATAAATCATGGTCTTAATGGCCAGTCCAGCTGGCAAGCCAA
TGGACAATGGACGAGCTTTACGTCTGTCAGTGGAGAGCGTACCAACTATAAAACGCAAAAAA
GTGCCAGCCAAGGTGATAGTAATCACTATGCACTGTCGTTTGGCTCAAGCTATCGCCTCAATG
ATCAATGGCGTGTAGGTTTAGGTGTTAGTTTACAAGAAAGCACTTTAACAGCGGGTGCCGAAG
ATTCTAAATATCGTTTAAACAGTTATTTACTCAGTCCTTTTGCGCAGTACAGTCATCAAGCGCT
GTGGGCCGATGTAACGCTGAGCGCAGGGCGCTTAGATTACGATAGTTTGGATCGCAAACTCG
ATTTGAATGCTGCCAAACGCACTGAGAAAGGAGACACTAAGGGTAATGTTTTAGGGGTTTATG
GACGCGTTGGTTATCAACTATTTGCTGCCCACAATCCACTGCAGCTTTCGCCATTTGTATCGAT
GAGTCACGCGCGCTTTAAAGTGGATGACTATGCTGAGAAAGGCAATAACTCAACGGCTTTAA
CCTTTGCTGAGCAAAAGCGCACATCTAAACGCTTAGGTGCGGGCCTGTTAGCAAGCTATCAGT
TAAGTGAGCCGTTGAGTTTGAGTGCTGAAATCGGTTACGAAAAAGAGTTTGCTAAGGATCAG
AAGAAGCTAGGTATGCACCTTAATTCGGTAGATTCTGTGCATTTTAAATTGCGAGGTTATAAA
CCCGATAGTTCTTTAGGAACGCTGGGTTTAGGCGCGAGTTATAAGTTGTCGGATGCATTAACT
ATGAAAGGCAATTACAACTACATGCATGCTGATTCAGTCCGTCAGCATGCACTGGGTGTCGGT
GTAAGCCTTAACTGGTAAATGTGGCTGTGTATAAAATTAAGGCGCTTTACAGCGCCTTAATTT
TTAATCGCCCCAGCGTCCAGGCGAATAAGAAAATACCGGCAGAGATAGCTTCCAGCGGATGG
CTGCCAAGCGTAAGGCTAGGCCGAAGCTGAAGGATACGGCGGTATTGATATCGTCAGATACG
CCTTGCCACATCAGCACCAGGTAGACAATTGCGACCACCAACGACACGCTGGCGTAGAGTTC
ATGACGCAGCACTTGTGGCGTACGGTTGCACATAATATCGCGCAGAATACCGCCAAAGATAC
CGGTGGTAATCCCCGCCATAATAACTACAGGTGTCTCGTAGCCCAGCTTTAAGGCGACATTAC
AGCCAATGATGGTAAAGGCGACTAAGCCCATGGCATCCAAAACCAAGAAAACTTGATTGAGC
TTTTGCATAAAGCGCGCGACCAACATGGTCGCGAGGCCTGAGCCAATGGTCAGATAGATATA
GGGCGGGTGTTGTGTCCAAGTGACT 3’
```

Figure 16 Full length sequence of lipolytic clone LD1 obtained by primer walking. Primer binding sites are indicated in yellow (LD1 RUS 3 [5'CGTCTACAACGACAG AACCATCAGC 3']), pink (LD1 RUS 2 [5’GCTGAGGTTATTGCGCTG 3’]), green (LD1 US2 [5’ GCGCGCCATCTCTGGTAAC 3’]), red (LD1 RUS1 [5’ GCACTGTCGTTTGGCTCAAGC 3’]), blue (LD1 TM3R [5’ CATCTAAACGCTT AGGTGC 3’]) and turquoise (LD1 TM3F [5’ CCTTAACTGGTAAATGTGG 3’]). A putative ribosome binding site 7 bp upstream of the start codon is underlined.



Accession number	% identity
NP 253799.1	51%
EAZ61448.1	50%
P40604	50%
NP 790416.1	48%
ZP 01368126.1	51%

Figure 17 Multiple sequence alignment of the *LD1* gene sequence with hits generated from BLASTp. Boxes indicate sequence similarity with a threshold value of 90%. The PROSITE motif predicted in LD1 is underlined in red. * beneath the conserved residues in the N- and C- terminals are indicated. Blocks I, II, III and V are shown. The C-terminal autotransporter is blocked in blue. Accession numbers in the figure denote the following; NP_253799.1: Esterase EstA [*Pseudomonas aeruginosa* PAO1], EAZ61448.1: Esterase EstA [*Pseudomonas aeruginosa* 2192], P40604: Uncharacterized protein in trpE-trpG intergenic region precursor [*Pseudomonas putida*], NP_790416.1: Autotransporting lipase, GDSL family [*Pseudomonas syringae* pv. *tomato* str. DC3000], ZP_01368126.1: Hypothetical protein PaerPA_01005281 [*Pseudomonas aeruginosa* PACS2]. Percent identity of LD1 to the selected protein sequences are indicated in the table.

Pfam was used to find matches to the sequence obtained using protein family domains. Two significant matches were found; the first match was to the GDSL-like Lipase/Acylhydrolase family, with a bit score of 4.9e^{-6}, and the second match was to the Autotransporter beta-domain, with a bit score of 7e^{-31} (Figure 18). The PROSITE motif search also identified motifs in the gene sequence that matched to the GDSL-like Lipase family ([LIVMFYAG](4) - G - D - S - [LIVM] - x(1,2) - [TAG] - G) [Upton *et al*., 1995].

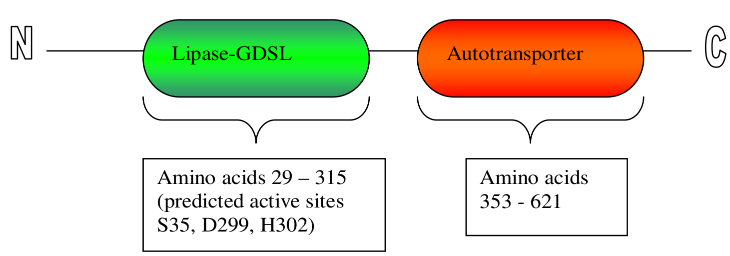


Figure 18 The graphic provided by Pfam which shows the arrangement of matches on the sequence obtained for LD1.

The SignalP server was used to identify a 21 amino acid N-terminal signal sequence in the predicted gene with the most likely cleavage site between amino acid positions 21 and 22 (ACA*AP) [Figure 19]. The Rare Codon Caltor predicted a total of 45 rare codons (Table 7) in the sequence. Eight percent of the sequence consists of rare codons with the majority [5%] being glycine (GGG and GGA) and threonine (ACG). This could lead to difficulties in expression of the gene in the heterologous host, *E. coli*.

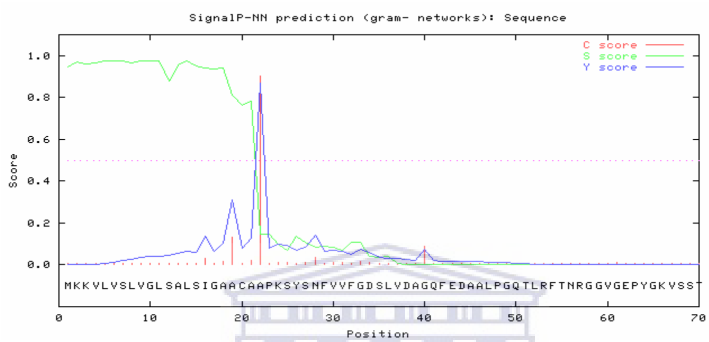


Figure 19 Prediction of N-terminal signal peptide cleavage site in polypeptide LD1.

Table 7 Rare codons and their frequency in the nucleotide sequence obtained for lipolytic clone LD1 as predicted by Rare Codon Caltor.

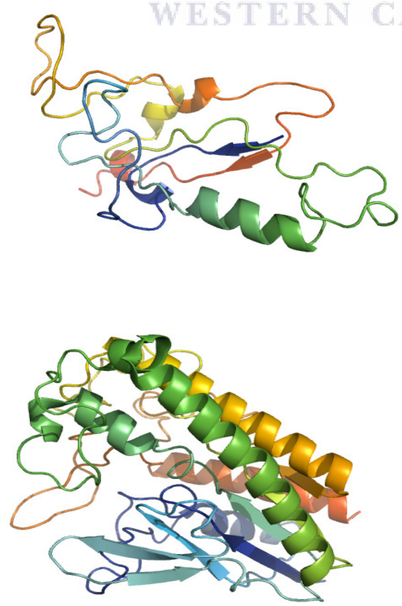
Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	4
	CGG	0
	AGG	0
	AGA	1
Glycine	GGA	10
	GGG	11
Isoleucine	AUA	2
Leucine	CUA	4
Proline	CCC	2
Threonine	ACG	11

The rare codons predicted by 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). Further analysis would be required in order to assess and compare the codon bias observed the lipolytic gene *LD1* to other microorganisms.

GDSL lipases/ esterases were discovered relatively recently and are representatives of the SGNH hydrolase superfamily (Upton *et al*., 1995; Akoh *et al*., 2004). Their structural properties do not resemble those observed for lipases or esterases belonging to the α-β hydrolase family. Lipolytic enzymes in this family generally have a catalytic serine in the conserved GXSXG sequence, located in the centre in the gene. However, GDSL lipolytic enzymes have the catalytic serine in the GDS(L) consensus sequence. This sequence is also located closer to the N-terminal of the gene (Akoh *et al*., 2004). The *LD1* gene contained this motif located near to the N-terminus. It also contained the three strictly conserved residues Gly-Asn-His found in blocks II, III and V, respectively. Two of these residues form the putative catalytic triad along with the catalytic serine residue found in block I. Furthermore, homology modelling of the N-terminus of the LD1 protein using the Swiss model server showed 13% sequence identity to the enantioselectivity region of esterases (Figure 20).

A

B



<u>Figure 20</u> (A) Homology model built by the Swiss model server using amino acids 28 to 165 of the N-terminus GDSL lipolytic enzyme, LD1. (B) The enantiorecognition site of esterases [1 ESC] was used as the template.

The gene *LD1* contained all the structural motifs of a typical autotransporter protein, the N-terminal signal sequence, a passenger domain [also referred to as the N-terminal domain or α-domain] (conferring lipolytic activity in this case) and the C-terminal domain [also called the β-domain or autotransporter domain]. The autotransporter domain is composed of an α-helical linker followed by β-sheet structures (Henderson *et al*., 2004; Lee *et al*., 2003; Jacob-Dubuisson *et al*., 2004). The C–terminal domain was predicted to contain 14 β-strands. These are thought to form the β-pore through which the N-terminal passenger domain is translocated (Jacob-Dubuisson *et al*., 2004). PSI-PRED results also showed these β-sheets were preceded by a helix structure (Figure 22).

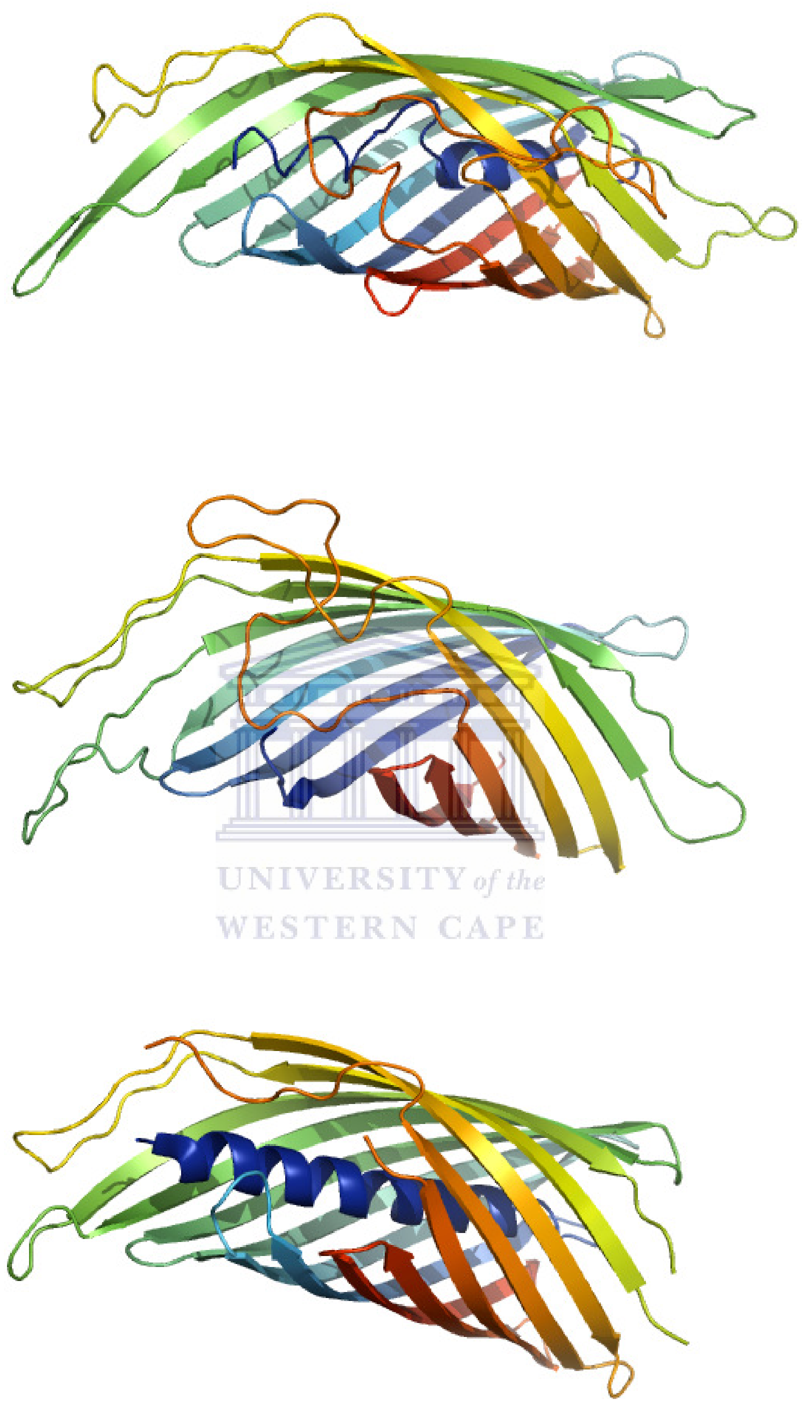
This helix is a possible linker, observed in all known autotransporters, which anchors the passenger domain by covalent bonding. Modelling of the protein using 3D-JIGSAW and the Swiss model server showed homology of the C-terminus to the translocation unit of NalP, an autotransported serine protease found in *Neisseria meningitis* [Figure 21] (Henderson *et al*., 2004).

Over 700 proteins belong to the autotransporter superfamily and they are generally ubiquitous among the Proteobacteria (Pallen *et al*., 2003). Autotransporter proteins are predominately found in pathogenic bacteria and have a wide variety of functions including cytotoxicity, serum resistance, protease and lipases/esterase activity (conferred by the passenger domain) and adhesion (Henderson *et al*., 2004; Jain *et al*., 2006; Ieva *et al*., 2008; Rosenau, 2000). Since autotransporter proteins have also been observed in non-pathogenic microbes it cannot be assumed that the lipolytic enzyme LD1 is a virulence factor in the organism from which it originates (van Ulsen *et al*., 2006).

A

B

C

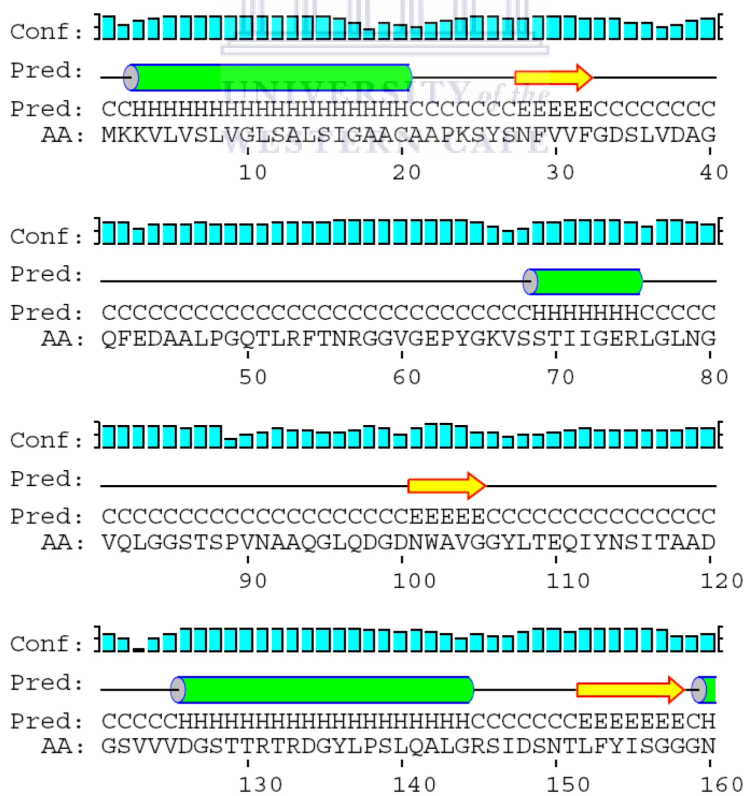


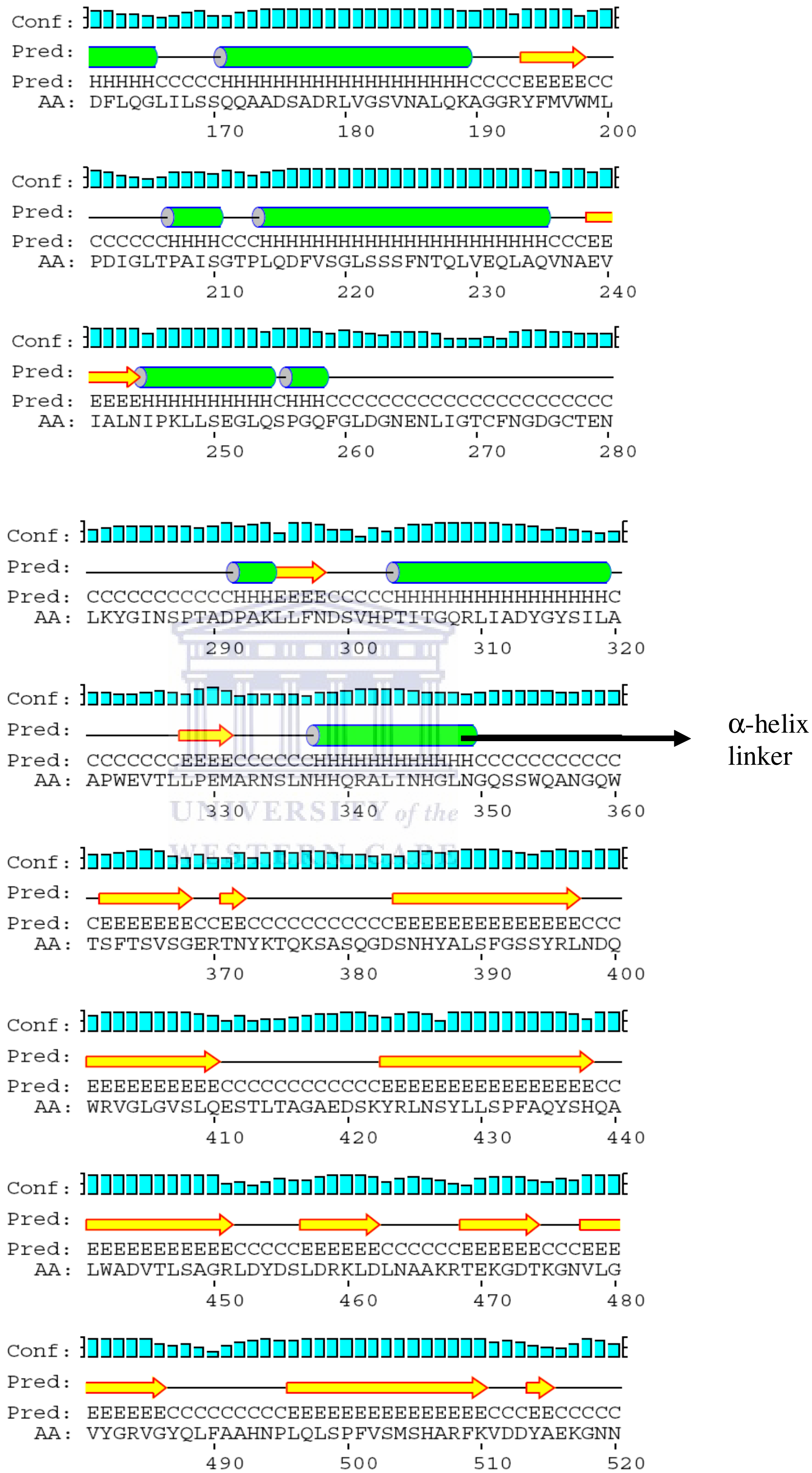
<u>Figure 21</u> Homology models of the C-terminus of LD1 built by the 3D JIGSAW (A) [amino acids 327 to 634] and the Swiss model server [amino acids 379 to 634] (B). The autotransporter domain of NalP from *N. meningitis* [1 UYN] was used as a template (C).

The degree of folding of the passenger domain that is achieved in the periplasmic space is still under investigation, but analysis of the crystal structures of NalP (*N. meningitis*), EspP (*E. coli*) and Hia (*H. pylori*) has shown that the narrow β-domain pore cannot support the translocation of passenger polypeptides that have tertiary structure (Ieva *et al*., 2008). Translocation of autotransporter passenger domains is an energy-independent process, and it is believed that partial folding of the passenger domain in the periplasm and subsequent complete folding at the surface of the cell may drive translocation (Ieva *et al*., 2008; Henderson *et al*., 2004). Proteolytic cleavage resulting in the separation of the protein into two polypeptides occurs during the course of secretion (Jacob-Dubuisson *et al*., 2004; Ieva *et al*., 2008). The β-domain remains bound to the outer membrane and is stable within the membrane, while the passenger is exposed at the cell surface or released into the medium (Desvaux *et al*., 2004).

Final localisation is related to the biological function of the N-terminal domain (Henderson *et al*., 2004; Jacob-dubuisson *et al*., 2004). It may be advantageous for an organism to retain an enzyme in close contact to the cell surface, especially in nutrient-poor systems (Rosenau, 2000). Antarctica, the sample source, is such an environment and the seal carcasses offer a relatively large pool of nutrients in this otherwise nutrient poor habitat (Smith *et al*., 2006). Even with this addition of nutrients to the environment, enzymatic reactions that are catalysed in close proximity to the cell may facilitate quicker uptake of nutrients and prevent utilisation of nutrients by other cells in the niche. The following mechanism is hypothesised for lipolytic gene *LD1* based on literature and experimental evidence.

1. The full length protein is transported across the inner membrane of the cell by Sec translocation (presence of an N-terminal signal peptide).
2. In the periplasm, folding of the protein occurs starting from the C-terminus (Henderson *et al*., 2004; Desvaux *et al*., 2004).
3. Partial folding of the passenger (GDSL esterase) occurs after formation of the β-barrel pore in the outer membrane (Henderson *et al*., 2004; Ieva *et al*., 2008).
4. The catalytic esterase is translocated through the pore into the extracellular space, but remains bound to the C-terminal domain via the α- helix linker (PSI-PRED results showing the linker region).
5. The enzyme is not likely to be released into the medium as no activity was observed in the culture supernatant, however, this observation may be due to dilution of the enzyme in the supernatant.




Conf:
Pred:
Pred: HHHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCCEEEEEECC
AA: DFLQGLILSSQQAADSADRLVGSVNALQKAGGRYFMVWML
170 180 190 200
Conf:
Pred:
Pred: CCCCCCHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCEE
AA: PDIGLTPAISGTPLQDFVSGLSSSFNTQLVEQLAQVNAEV
210 220 230 240
Conf:
Pred:
Pred: EEEEHHHHHHHHHHHCHHHCCCCCCCCCCCCCCCCCCCCCC
AA: IALNIPKLLSEGLQSPGQFGLDGNENLIGTCFNGDGCTEN
250 260 270 280
Conf:
Pred:
Pred: CCCCCCCCCCCHHHEEEECCCCCHHHHHHHHHHHHHHHHC
AA: LKYGINSPTADPAKLLFNDSVHPTITGQRLIADYGYSILA
290 300 310 320
Conf:
Pred:
Pred: CCCCCCCEEEECCCCCCHHHHHHHHHHHCCCCCCCCCCC
AA: APWEVTLLPEMARNSLNHHQRALINHGLNGQSSWQANGQW
330 340 350 360
α-helix
linker
Conf:
Pred:
Pred: CEEEEEEECCEECCCCCCCCCCCEEEEEEEEEEEEEECCC
AA: TSFTSVSGERTNYKTQKSASQGDSNHYALSFGSSYRLNDQ
370 380 390 400
Conf:
Pred:
Pred: EEEEEEEEEECCCCCCCCCCCCEEEEEEEEEEEEEEEEECC
AA: WRVGLGVSLQESTLTAGAEDSKYRLNSYLLSPFAQYSHQA
410 420 430 440
Conf:
Pred:
Pred: EEEEEEEEEEECCCCCEEEEEEECCCCCCEEEEEEECCCEEE
AA: LWADVTLSAGRLDYDSLDRKLDLNAAKRTEKGDTKGNVLG
450 460 470 480
Conf:
Pred:
Pred: EEEEEECCCCCCCCCEEEEEEEEEEEEEEECCCEECCCCC
AA: VYGRVGYQLFAAHNPLQLSPFVSMSHARFKVDDYAEKGNN
490 500 510 520

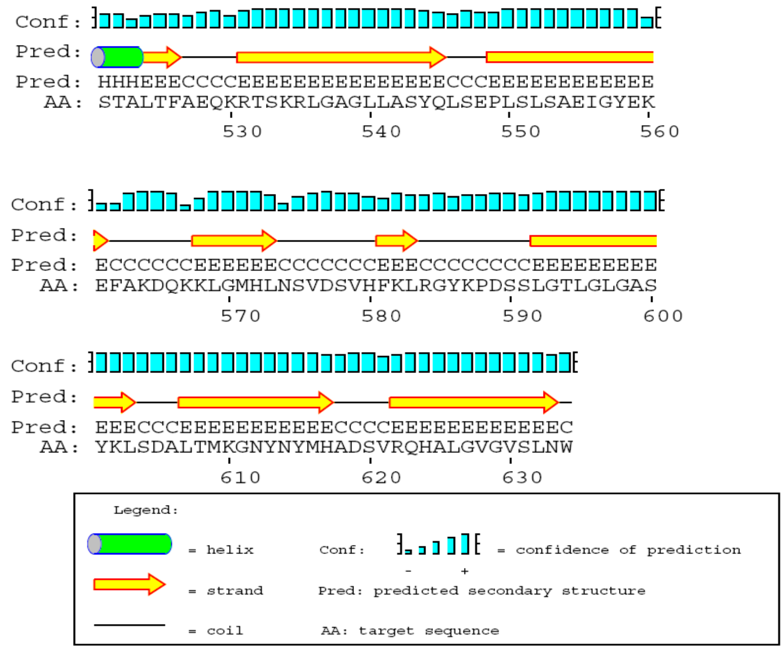
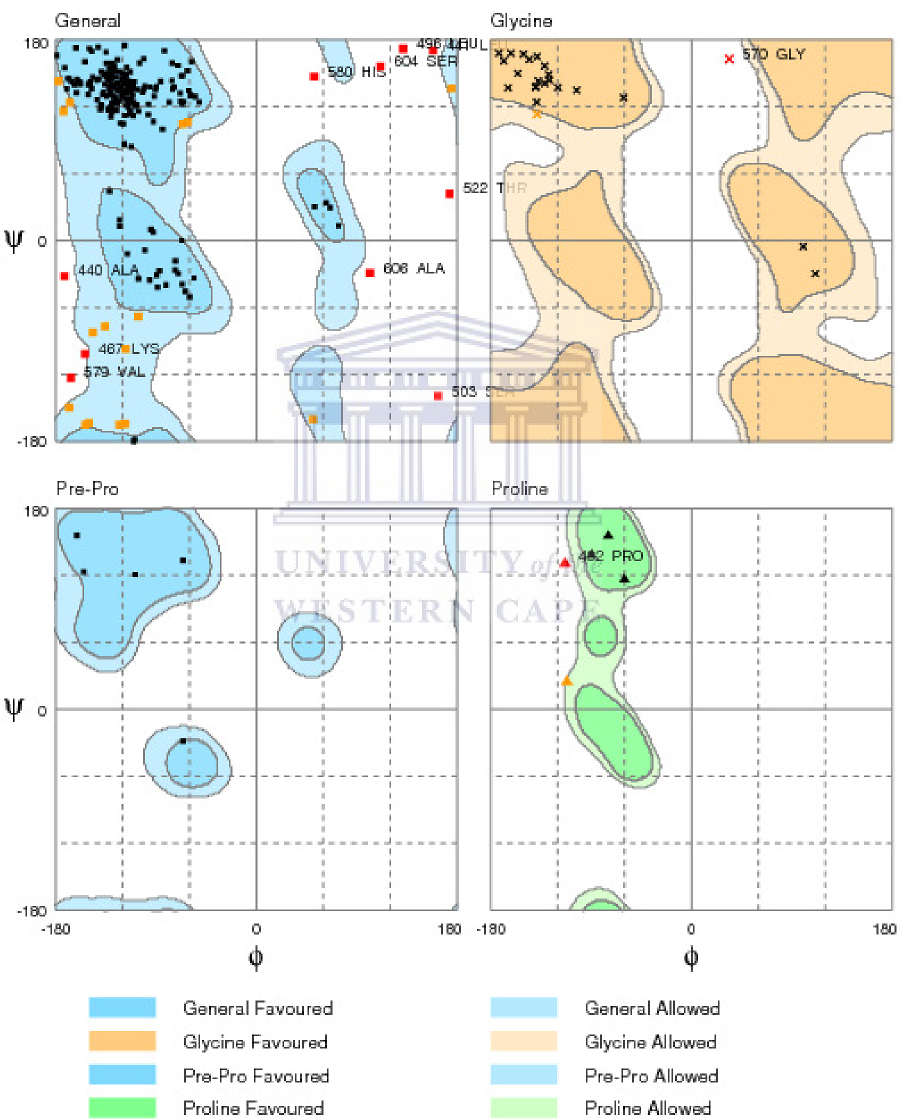


Figure 22 Secondary structure predicted by PSIPRED (Jones 1999; McGuffin, Bryson *et al*., 2000) for the amino acid sequence obtained for lipolytic clone LD1. The C-terminal domain consists mainly of β-strand structures while the N-terminal domain has a higher proportion of helical structures.

RAMPAGE was used to visualise and assess the Ramachandran plots of the predicted protein structures of the N- and C-terminal domain of LD1. For a model to be accurate, 98% of residues are expected in the favoured region and 2% of residues in the allowed region. The model of the C-terminus of LD1 built by the Swiss model server contained 88.2% residues in favoured region, 7.1% residues in allowed region and 4.7% residues in outlier region (Figure 23). The C-terminal model of LD1 generated by 3D-JIGSAW contained only 87.5% of residues in the favoured region, 6.5% of residues in the allowed region and 6.1% of residues in the outlier region (Figure 24). This indicates that the model built by the Swiss model server was more accurate. This might be due to the fact that fewer residues were used to construct the model. However, both models of the autotransporter domain of LD1 were considered to be inaccurate due to the number of residues present in outlier regions. Furthermore, the catalytic N-terminus of LD1 contained 80.9% of residues in favoured region, 8.8% of residues in allowed region and 10.3% of residues in outlier region (Figure 25). It was concluded that accuracy of this model of LD1 was relatively low. Considering that cold-adapted enzymes may have structural modifications and that comparisons of this nature generally utilise mesophilic counterparts, a number of outliers could be expected.



<u>Figure 23</u> Ramachandran plot for the model of the C-terminal autotransporter of LD1 built by the Swiss model server.

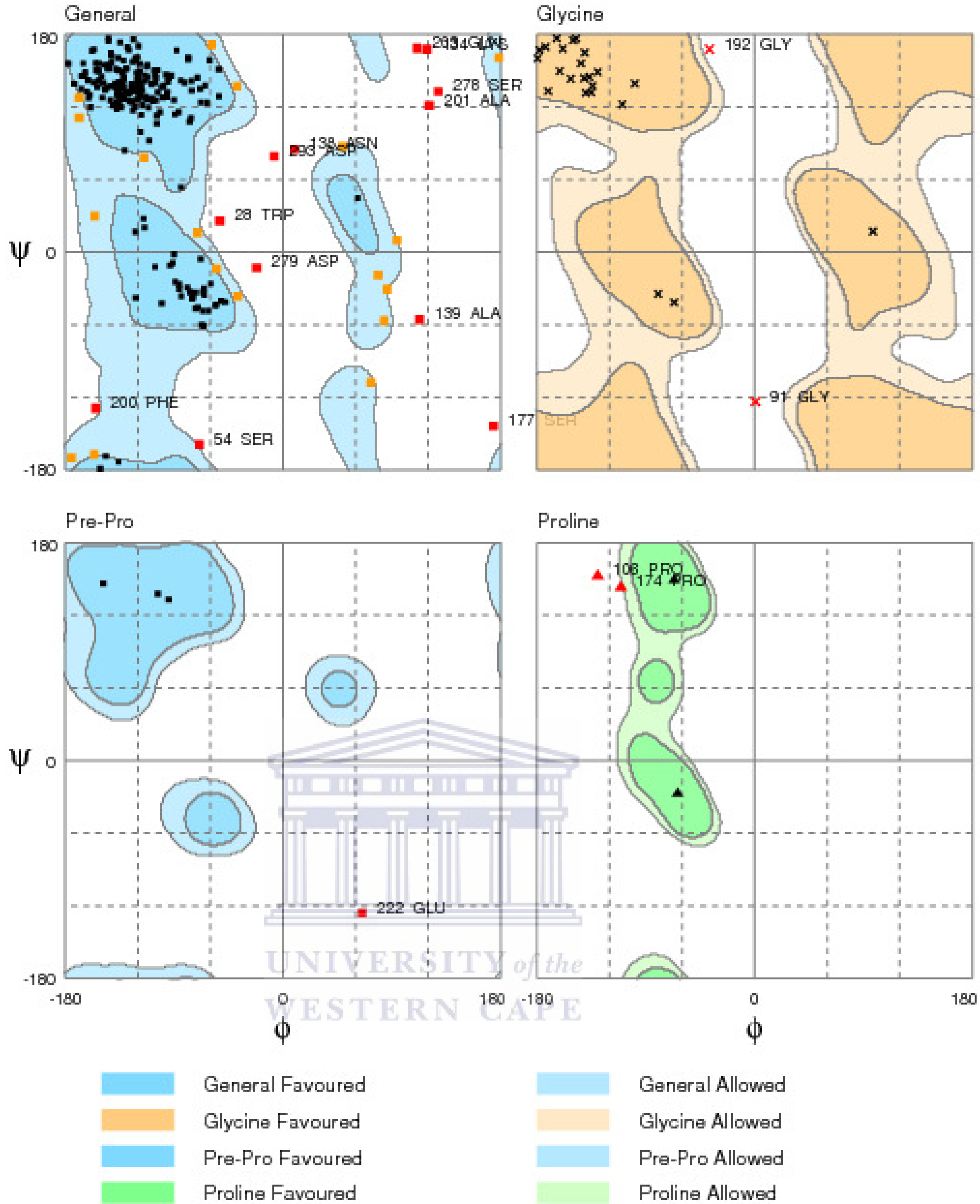
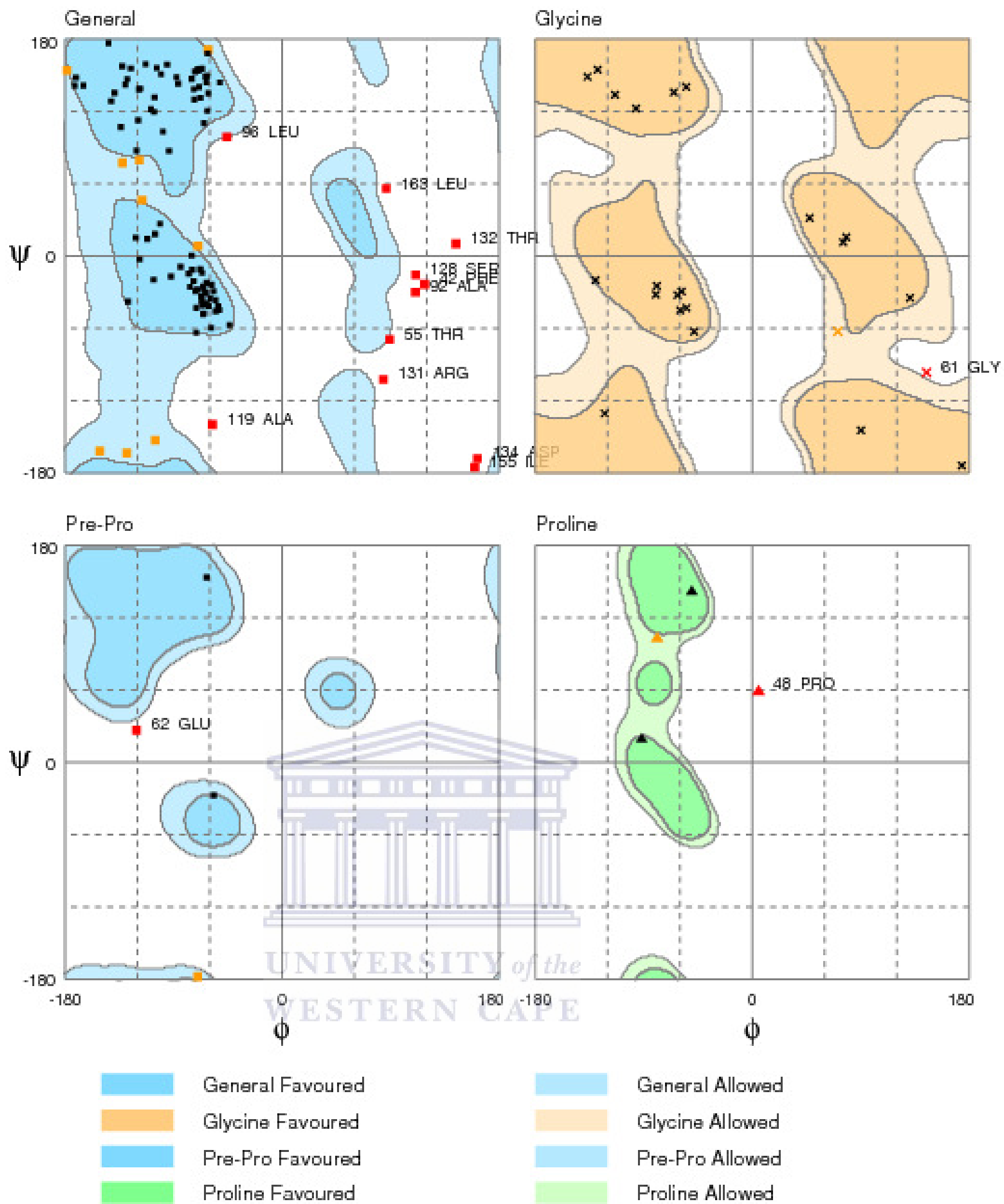


Figure 24 Ramachandran plot for the model of the C-terminal autotransporter of LD1 built by 3D-JIGSAW.



<u>Figure 25</u> Ramachandran plot for the model of the N-terminus of LD1 built by the Swiss model server.

Since the model of LD1generated by 3D-JIGSAW covered more amino acid sequence, PyMol was used to superimpose the template and the model (Figure 26). The residues found in the outlier region were mostly located in the external loop regions of the C-terminus LD1 model and did not appear to effect the β-barrel conformation of the C-terminal domain.

It can be seen in Figures 21 and 26, that the α- helix linker of LD1 was not completely modelled, suggestion fundamental structural differences between LD1 and the autotransporter region of NalP.

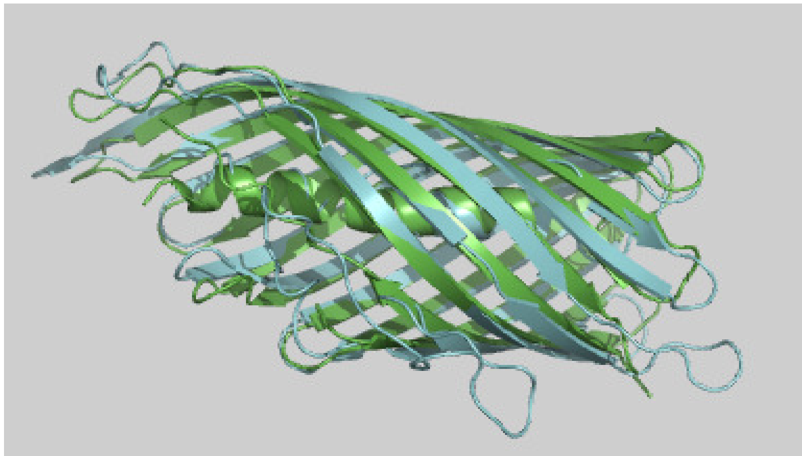
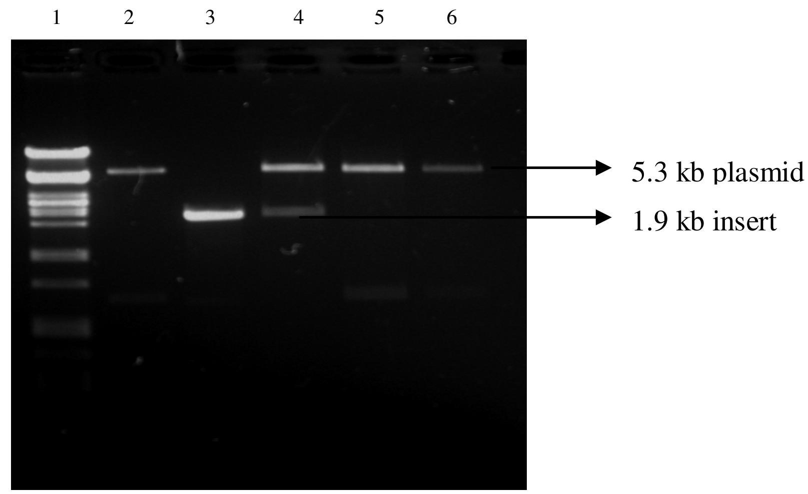


Figure 26 PyMol superimposed model with the autotransporter domain of NalP (green) and the C-terminal domain structure of LD1 predicted by 3D-JIGSAW (light blue).

3.1.8 Cloning of the lipolytic gene *LD1* into an expression vector

DNAMAN was used to predict restriction enzyme recognition sites occurring in the *LD1* DNA sequence. It was established that *Nde*I and *Xho*I did not cut within the gene sequence and a primer pair [LD1-R-X1 and LD1-F-N1] (Table 4, page 45) was designed with these sites introduced at the ends of each primer. To obtain the full length gene of 1.9kb, gradient PCR was performed. Gel electrophoresis of the resulting PCR products showed that a fragment corresponding to the expected size was successfully amplified. The fragment was cloned into the pET28a vector (section 2.6). Successful cloning was confirmed by colony PCR (using LD1-R-X1 and LD1-F-N1), restriction digestion using *Nde*I and *Xho*I (Figure 27), and sequencing using the T7 promoter and terminator primers. The clone designated LD1-pET +3 conferred lipolytic phenotype on the host when transformed.



<u>Figure 27</u> Restriction enzyme digestion of clone LD1-pET +3. Lane 1) DNA molecular marker, lambda-PstI digested DNA. Lane 2) Linearised pET28a vector DNA. Lane 3) PCR amplified gene of LD1 digested with *Nde*I and *Xho*I. Lane 4) Plasmid extracted from clone LD1-pET +3 and digested with *Nde*I and *Xho*I. Lane 5) Plasmid extracted from a clone randomly selected from the control plate and digested with *Nde*I and *Xho*I. Lane 6) Uncut pET28a vector DNA.

3.1.9 Expression of lipolytic gene *LD1*

Clone LD1-pET +3, as well as the pET28a circularized vector, were transformed into the *E. coli* expression strains ArcticExpress (DE3) and Rosetta (DE3) pLysS. Small-scale expression experiments were performed at 16°C for 24 or 36 hours and were analysed by SDS-PAGE. LD1 over-expression was not detected during the small scale expression studies. A large scale expression study was performed under conditions described in chapter 2.8. Analysis by SDS-PAGE of the cytoplasmic fraction of LD1-pET ArcticExpress, bound and eluted from His-bind resin, showed no additional protein band. However, the eluted fraction from LD1-pET Rosetta (DE3) pLysS showed a protein band of approximately 70 kDa (Figure 28). This was the expected size of the full length protein.

The recovery of this full length protein in the soluble fraction was unexpected. If the signal peptide (and therefore the histidine tag) was cleaved in the *E. coli* host following translocation across the inner membrane, this protein could not be a periplasmic proprotein. Since insertion of the C-terminal domain into the outer membrane is believed to occur rapidly (Henderson *et al*., 2004), it is possible that the His-tag purified protein was a cytoplasmic proprotein. Addition of the histidine tag to the N-terminus might have affected the efficiency of translocation across the inner membrane via the Sec-dependent pathway. The eluted fraction also showed lipolytic activity when assayed with p-nitrophenyl esters, which could be attributed to partial folding of the full length protein, thereby allowing some catalytic function.

To investigate this further, both the soluble and insoluble fractions were analysed by SDS-PAGE. Analysis of the membrane fraction showed over-expression of two protein bands which was consistent with the predicted molecular masses of the individual N- and C- terminal proteins (34.4 kDa and 33.5 kDa, respectively). Following the solubilisation and refolding experiment (chapter 2.8), a pure preparation of both these bands was observed by SDS-PAGE analysis [Figure 29]. Additionally, esterase activity was observed when this fraction was used in preliminary enzyme assays.

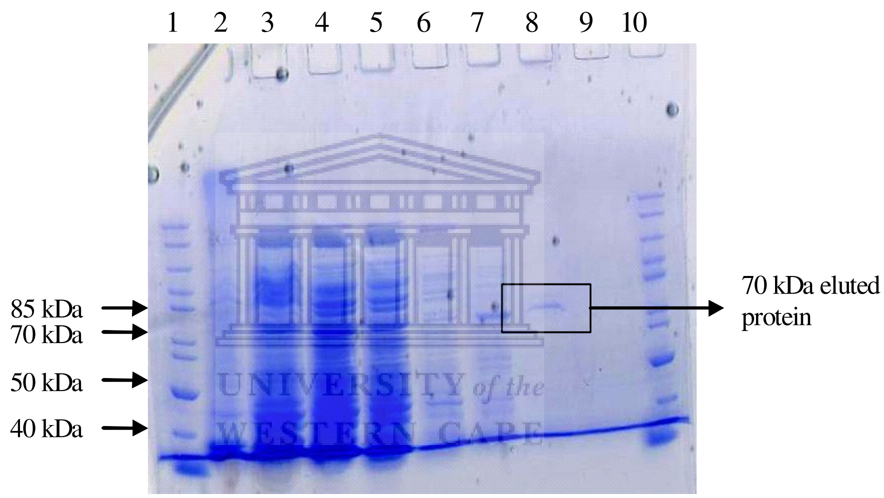
The β-barrel structures of autotransporters are relatively stable and cellular localisation studies have demonstrated the persistence of these structures in the cell membrane (Henderson *et al*., 2004; Charles *et al*., 1994). According to the model of translocation hypothesised for LD1, the catalytic domain would be kept in close association with the β-pore by covalent bonding. The observation of a cleaved pro-protein was therefore not expected. A possible explanation could be that the protein was toxic to the *E. coli* expression host, leading to the proteolysis of the full length protein by an unknown mechanism. However, this would have to be determined experimentally.

In a recent study of esterase [EstE] from *Xanthomonas vesicatoria*, the protein was found exclusively in the insoluble fraction of the cell lysate (Talker-Huiber *et al*., 2003). Furthermore, the authors predicted that the N-terminal catalytic domain anchored by the β-barrel membrane protein was unusually directed towards the periplasm and not the extracellular medium (Talker-Huiber *et al*., 2003). The authors also concluded that the C-terminal extension on the protein was unlikely to be an autotransporter, but rather a porin-like membrane protein which could direct substrate through the outer membrane and into the periplasm where the catalytic esterase domain would function (Talker-Huiber *et al*., 2003).

If this were true for LD1, one could expect that cleaved polypeptides would be found in the insoluble fraction if the N-terminal domain was accessible to the periplasmic proteases in the *E. coli* expression host. The hypothesised mechanism of translocation would therefore, be incorrect. Differential fractionation experiments would need to be performed in order to determine the subcellular localisation of LD1 and external protease accessibility experiments could be used to determine the orientation of the N-terminal catalytic domain.

The ArcticExpress (DE3) strain is designed for expression of cold-active proteins due to the co-expression at 16 °C of the cold-adapted chaperonins, Cpn10 and Cpn60, from the psychrophilic bacterium, *Oleispira antarctica*. These chaperonins decrease inclusion body formation by preventing recombinant protein aggregation at low temperatures, thereby increasing the yield of soluble protein (manufacturers manual).

However, in this study, Rosetta (DE3) pLysS appeared to produce better expression yields. This strain contains a plasmid that supplies tRNAs for several rare codons. Frameshift errors that ultimately lead to low levels of expression and the formation of non-functional proteins is observed when rare codons such as leucine (CAU), arginine (AGG/ AGA), proline (CCC) and isoleucine (AUA) are present in proteins (Wu *et al.*, 2004). Since all of these rare codons are present in gene *LD1*, rare tRNAs might be required to obtain sufficient expression of the enzyme. The reduction of codon bias appeared to be a major factor in the efficient expression of recombinant LD1 protein in the heterologous hosts.



<u>Figure 28</u> SDS-PAGE analysis of His-tag purification of LD1-pET + 3 in Rosetta (DE3) pLysS induced with 0.5 mM IPTG and grown for 5 days at 16°C. The protein band corresponding to a size of 50 kDa and the ~70 kDa eluted protein is indicated. Lanes 1 and 10) Protein molecular weight ladder. Lane 2) Total cell extract of LD1-pET. Lanes 3 and 4) Soluble fraction of LD1-pET. Lane 5) Eluate from loading. Lane 6) Eluate from the binding step. Lane 7) Eluate from the wash step. Lane 8) Eluted LD1 protein (molecular weight ~ 70 kDa).

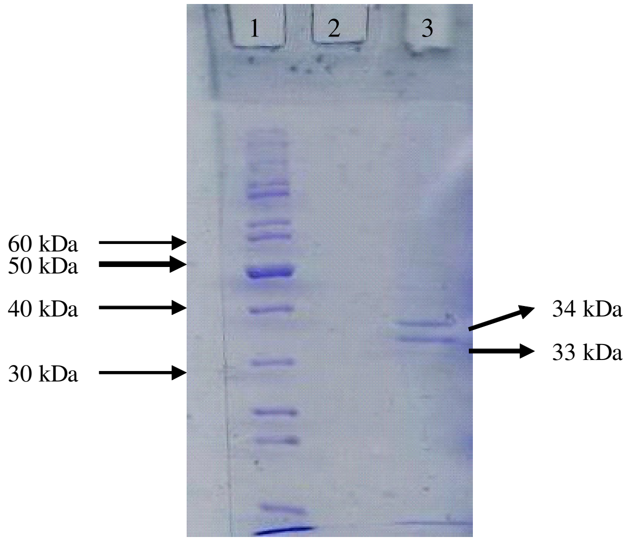


Figure 29 SDS-PAGE analysis of the folded LD1 protein described in section 2.8. Lane 1) Protein molecular weight ladder. The 50 kDa protein band is indicated. Lane 3) Re-solubilised membrane fraction of LD1-pET in Rosetta (DE3) pLysS.

3.1.10 Initial enzyme assays using crude extract

A crude enzyme assay was performed using the soluble and insoluble fractions of LD1-pET +3 and the pET28a control (chapter 2.7). No enzyme activity was observed when the insoluble membrane fraction was assayed with ester substrates C_3 (*p*-nitrophenyl propionate), C_8 (*p*-nitrophenyl caprylate) and C_{12} (*p*-nitrophenyl laurate). The soluble fraction showed activity when assayed with C_3, C_8 and C_{12} [Figure 30]. The control showed no activity with any of the substrates used in the assay. Since a gene encoding an esterase is found in the *E. coli* genome (Blattner *et al*., 1997), the expression strain was used as a control in these assays. The specific activity calculated for LD1-pET +3 was 0.12U/mg of total protein.

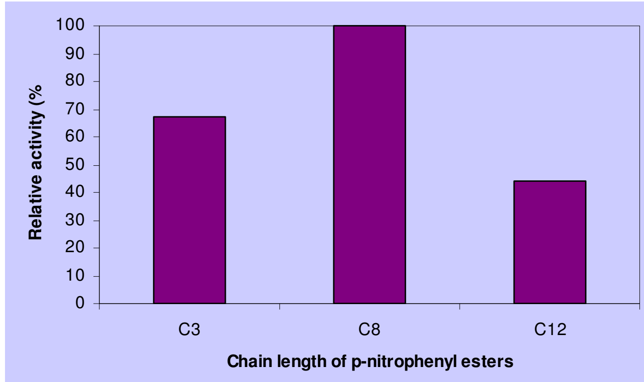


Figure 30 Activity of the crude extract toward *p*-nitrophenyl esters of varying chain lengths. C_8 is taken as 100%.

3.1.11 Preliminary kinetic analysis

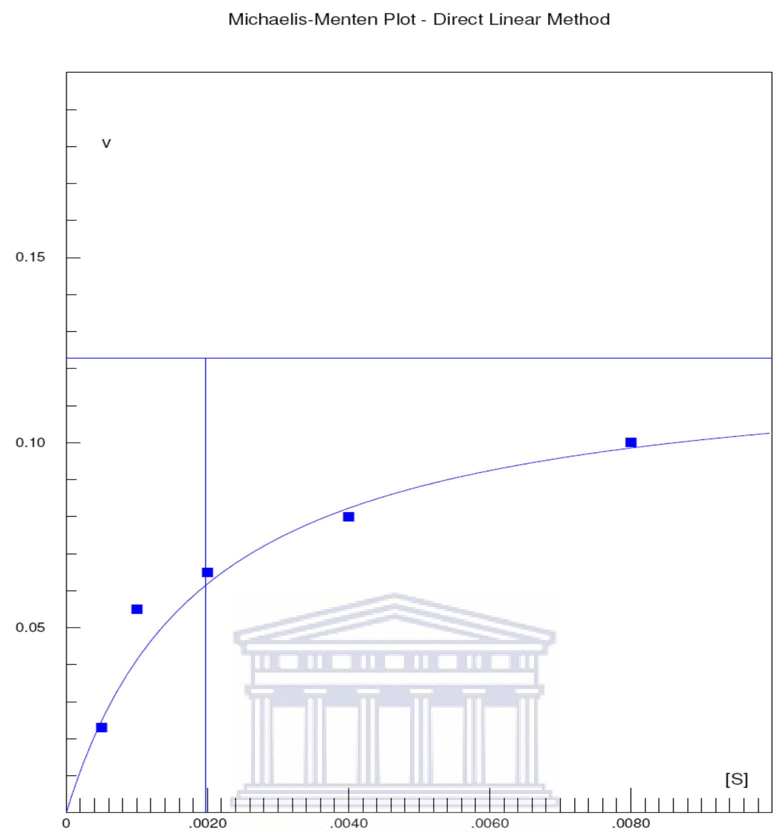
Following dialysis of soluble (His-tag purified) and insoluble fractions (section 2.8), preliminary enzyme assays were performed (section 2.2.15). No activity was observed with the C_3, C_8 and C_{12} substrates. Minimal activity was detected when *p*-nitrophenyl acetate (C_2) was used as a substrate, but only when large amounts of the enzyme fractions were used (50 µg and 450 µg, respectively). This may be attributed to enzyme instability, or incorrect and/or partial folding of the enzymes in the dialysis buffer. For all substrates tested, the rate of formation of *p*-nitrophenol was directly proportional to the amount of enzyme used in the assay. There was a hyperbolic dependence of the rate on substrate concentration. No enzyme activity was observed after heat inactivation of samples at 60°C for 5 minutes, giving some indication that LD1 is heat labile. The control also showed no activity towards any substrates used in the assays. The data from these preliminary assays was analysed by the direct linear plot generated by the Enzpack program (Figure 31). The K_m for enzyme LD1 (His-tag purified) was determined to be 1.97 mM. A V_{max} value of 8.2×10^{-2} U/mg of enzyme was determined for the His-tag purified LD1 enzyme.

The results obtained from both the crude enzyme and preliminary assays do not verify the exact substrate specificity or the kinetic parameters of enzyme LD1. In the crude enzyme assay, LD1 showed activity to short-, medium-, and long chain fatty acid substrates indicating lipase-like activity for LD1. However, in the preliminary enzyme assays, no activity was detected on substrates greater than *p*-nitrophenyl octanoate (C_{10}). This suggests that LD1 is an esterase, rather than a lipase (further supported by structural analysis and homology modelling) [section 3.1.7].

V_{max} estimates for some cold-active lipolytic enzymes range from 100- 160 U/mg of protein (Choo *et al*., 1998; Lee *et al*., 2003; Hårdeman *et al*., 2007). These values are much larger than that obtained for LD1. One possible reason for the result observed for LD1 may be the improper folding of the enzyme thereby limiting accessibility of substrate to the active site of the enzyme. Differences observed in the crude enzyme assay and the preliminary enzyme assays might be explained by the presence of accessory proteins that could assist in correct folding of LD1 in the crude extract. These proteins would be absent from purified preparations of LD1 and the folding of the enzyme would be subsequently affected.

It must be noted that characterisation of the enzyme is still preliminary. A main goal of future studies is to obtain a purified and active protein, which will allow accurate determination of substrate specificity, pH optimum, temperature optimum and thermostability.

a)



b)

Curve No	Km	68% Confidence Limits	Vmax	68% Confidence Limits
1	.00197	.00106 - .00267	0.123	0.104 - 0.133

<u>Figure 31</u> a) Preliminary Michaelis-Menton direct linear plot of rate, v (AU/min) versus substrate concentration (M) for LD1. **b)** K_m and V_{max} estimates computed by the Enzpack program.

Concluding remarks

It has been shown that the Antarctic Dry Valley soils contain high levels of diversity and that much of this microbial diversity is novel and as yet uncultured (Cowan *et al*., 2002; Smith *et al*., 2006). Lifecycle strategies and biological adaptations employed by microbes inhabiting extreme environments such as Antarctica are of major biotechnological interest (Vincent, 2000; Peck *et al*., 2006). The application of metagenomic technology to extreme environments allows researchers to gain access to the genetic information of 'unculturable' microorganisms, thereby providing unique opportunities for novel gene discovery (Cowan *et al*., 2004; Ferrer *et al*., 2007; Lämmle *et al*., 2007; Streit *et al*., 2004). In this study, high molecular weight DNA was extracted from soils obtained from beneath seal carcasses in the Miers Valley, Eastern Antarctica. This DNA (in the size range of 23-40 kb) was used to construct a large insert metagenomic DNA fosmid library. The coverage obtained in this library was estimated to be 2.29×10^7 base pairs, equivalent to 74 prokaryotic genomes.

Screening of the library for bacterial and archaeal 16S rRNA genes using a sequence-based method resulted in detection of bacterial 16S rRNA, although, no archaeal signals were detected by PCR. Additionally, the seal carcasses might have contributed eukaryotic DNA to the metagenomic DNA extract which may have been cloned into the library. DGGE analysis of bacterial 16S rRNA amplicons obtained from PCR identified 12 dominant phylotypes. Results indicated less than 1% representation of bacterial 16S rRNA genes in the fosmid library.

An activity-based method with tributyrin as substrate was used to screen the library for clones conferring lipolytic activity. Twelve tributyrin hydrolysing clones were obtained and one of these clones was selected for further study. Transposon mutagenesis and subsequent primer walking was used to obtain the 1902 bp putative gene sequence of the lipolytic clone (designated LD1). A number of bioinformatics tools were used for the prediction of structural features present in the gene. This gene showed structural homology with the C-terminus to the autotransporter family and conserved regions located at the N-terminus were homologous to the GDSL family of lipases and esterases. Furthermore, LD1 showed amino acid identity to other esterase sequences in the NCBI database (51% amino acid homology to *P. aeruginosa* EstA). Based on identification of conserved residues and sequence motifs, lipolytic enzyme LD1 was classified as a member of the SGNH super family. This enzyme was classified further into Family II (GDSL family) of lipolytic enzymes with Cluster 10 autotransporter homology.

Following cloning of the gene into expression stains, ArcticExpress (DE3) and Rosetta (DE3) pLysS, preliminary enzyme assays were performed using *p*-nitrophenyl esters of varying chain length. Lipolytic enzyme LD1 exhibited esterase activity, as indicated by a preference to short chain ($>C_{10}$) *p*NP esters. The enzyme showed no activity after heating at 60°C for 5 minutes, thereby showing some thermolability, consistent with its possible designation as a cold-active enzyme.

A survey of current literature suggests that this is the first report of a bacterial autotransporting GDSL esterase homolog isolated from a cold habitat soil metagenome.

The number of lipolytic enzymes being discovered and characterised is steadily increasing (Jaeger *et al*., 1999; Joseph *et al*., 2007). This could possibly be due to the renewed interest in these enzymes for biotechnological applications (Hasan *et al*., 2006; Joseph *et al*., 2008). A large number of cold-active microbial lipolytic enzymes have been kinetically characterised (Table 3), although, very few have resolved 3D crystal structures (Roy *et al*., 2007). This has been attributed to the difficulty of crystallisation of psychrophilic enzymes, mainly due to the flexibility and instability exhibited by these cold-adapted enzymes (Russell, 2000). A main goal in future studies includes the expression, purification and complete structural and functional characterisation of LD1. The low level of structural homology exhibited by LD1 may also make this enzyme an interesting candidate for crystallisation studies. Comparisons between LD1 and its mesophilic and thermophilic counterparts may help to clarify the mechanisms of cold-adaptation in proteins. Eight of the tributyrin hydrolysing clones obtained from functional screening of the metagenomic fosmid library, including LD1, are currently being sequenced at the University of the Western Cape, using SOLEXA technology. The elucidation of ORFs in these inserts and the characterisation of genes encoding the lipolytic activity will be the focus of future studies.

Congress contributions

International

Anderson, D. E., C. Heath., C. Cary and D. A. Cowan (2008). A novel cold-active lipolytic enzyme from an Antarctic metagenomic library. Extremophiles 2008. Cape Town, South Africa, September 2008.

References

1. Abdou, A. M. (2003). "Purification and partial characterisation of psychrotrophic *Serratia marcescens* lipase." *Journal of Dairy Science* **86**: 127-132.

2. Adams, B. J., R. D. Bardgett, *et al*. (2006). "Diversity and distribution of Victoria Land biota." *Soil Biology and Biochemistry* **38**: 3003-3018.

3. Aislabie J.M., K. Chhour, *et al*. (2006). "Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica." *Soil Biology and Biochemistry* **38**: 3041-3056.

4. Akoh, C., G. Lee, *et al*. (2004). "GDSL family of serine esterases/lipases." *Progress in Lipid Research* **43**: 534-552.

5. Altschul, S. F., J. L. Madden, *et al*. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Research* **25**: 3389-3402.

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

7. Arpigny, J. L., J. Lamotte, *et al*. (1997). "Molecular adaptation to cold of an Antarctic bacterial lipase." *Journal of Molecular Catalysis B: Enzymatic* **3**: 29-35.

8. Baker, G. C., J. J. Smith, *et al*. (2003). "Review and re-analysis of domain-specific 16S primers." *Journal of Microbial Methods* **55**: 541-555.

9. Balks, M. and I. Campbell (2001). *Ross Sea region 2001: A state of the environment report for the Ross Sea region of Antarctic.* Christchurch, New Zealand Antarctic Institute.

10. Barwick, R. E. and R. W. Balham (1967). "Mummified seal carcasses in a deglaciated region of South Victoria land, Antarctica." *Tuatara* **15**: 165-180

11. Bates, P. A., L. A. Kelley, *et al*. (2001). "Enhancement of Protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM." *Proteins: Structure, Function and Genetics* **5**: 39-46.

12. Béjà, O. (2004). "To BAC or not to BAC: marine ecogenomics." *Current Opinion in Biotechnology* **15**: 187-190.

13. Béjà, O., M. T. Suzuki, *et al*. (2000). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." *Environmental Microbiology* **2**: 516-529.

14. Bertrand, H., F. Poly, *et al*. (2005). "High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction." *Journal of Microbial Methods* **62**: 1-11.

15. Besemer, J. and M. Borodovsky (1999). "Heuristic approach to deriving models for gene finding." *Nucleic Acid Research* **27**: 3911-3920.

16. Blattner, F. R., G. Plunkett, *et al*. (1997). "The complete genome sequence of Escherichia coli K-12." *Science* **277**: 1453-1474.

17. Bohannan, B. J. M. and J. Hughes (2003). "New approaches to analyzing microbial biodiversity data." *Current Opinions in Microbiology* **6**: 282–287.

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

19. Boubakri, H., M. Beuf, *et al*. (2006). "Development of metagenomic DNA shuffling for the construction of a xenobiotic gene." *Gene* **375**: 87-94.

20. Boyd, W. L. and J. W. Boyd (1963). "Soils microorganisms of the McMurdo sound area, Antarctica." *Applied Microbiology* **11**: 116-121.

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

22. Brookes, P. (2001). "The soil microbial biomass: Concept, measurement and applications in soil ecosystem research." *Microbes and Environments* **16**: 131-140.

23. Cavicchioli, R., K. Siddiqui, *et al*. (2002). "Low temperature extremophiles and their applications." *Current Opinion in Biotechnology* **13**: 253-261.

24. Charles, L., N. Fairweather, *et al*. (1994). "Expression of the *Bordetella pertussis* P.69 pertactin adhesin in *Escherichia coli*: fate of the carboxy-terminal domain." *Microbiology* **140**: 3301-3308.

25. Chattopadhyay, M. K. (2006). "Mechanism of bacterial adaptation to low temperature." *Journal of Biosciences* **31**: 157-165.

26. Choo, D., T. Kurihara, *et al*. (1998). "A clod-adapted lipase of an Alaskan psychrotroph, *Pseudomonas* sp. strain B11-1: gene cloning and enzyme purification and characterisation." *Applied and Environmental Microbiology* **64**: 486-491.

27. Coleman, D. C. and W. B. Whitman (2005). "Linking species richness, biodiversity and ecosystem function in soil systems." *Pedobiologia* **49**: 479-497.

28. Cowan, D., Q. Meyer, *et al*. (2005). "Metagenomic gene discovery: Past, present and future." *Trends in Biotechnology* **23**: 321-329.

29. Cowan, D. A. and L. Ah-Tow (2004). Endangered Antarctic environments. *Annual Review of Microbiology*. **58:** 649-690.

30. Cowan, D. A., A. Arslanoglu, *et al*. (2004). "Metagenomics, gene discovery and the ideal biocatalyst." *Biochemical Society Transactions* **32**: 298-302.

31. Cowan, D. A., N. J. Russell, *et al*. (2002). "Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass." *Extremophiles* **6**: 431-436.

32. Dahllöf, I. (2002). "Molecular community analysis of microbial diversity." *Current Opinion in Biotechnology* **13**: 213-217.

33. D'Amico, S., P. Claverie, *et al*. (2002). "Molecular basis of cold adaptation." *Philosophical Transaction of the Royal Society London B* **357**: 917-925.

34. Daniel, R. (2004). "The soil metagenome - A rich resource for the discovery of novel natural products." *Current Opinion in Biotechnology* **15**: 199-204.

35. Daniel, R. (2005). "The metagenomics of soil." *Nature Reviews Microbiology* **3**: 470-478.

36. Desvaux, M., N. J. Parham, *et al*. (2004). "Type V protein secretion: Simplicity gone awry?" *Current Issues in Molecular Biology* **6**: 111-124.

37. De Vries, A. L., S. K. Komatsu, *et al*. (1970). "Chemical and physical properties of freezing point-depressing glycoproteins from Antarctic fishes." *Journal of Biological Chemistry* **245**: 2901-2908.

38. Dort, W. J. (1982). "The mummified seals of Southern Victoria Land, Antarctica." *Antarctic Research Series* **30**: 123-154.

39. Ekelöf, E. (1908). *Bakteriologische studien während der Schwedischen Südpolar-expedition 1901-1903*. Stockholm.

40. Elend, C., C. Schmeisser, *et al*. (2006). "Isolation and Biochemical Characterization of Two Novel Metagenome-Derived Esterases." *Applied Environmental Microbiology* **72**: 3637-3645.

41. Emanuelsson, O., S. Brunak, *et al*. (2007). "Locating proteins in the cell using TargetP, SignalP, and related tools." *Nature Protocols* **2**: 953-971.

42. Eyers, L., I. George, *et al*. (2004). "Environmental genomics: Exploring the unmined richness of microbes to degrade xenobiotics." *Applied Microbiology and Biotechnology* **66**: 123-130.

43. Ferrer, M., V. Olga, *et al.* (2005). "Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora." *Environmental Microbiology* **7**: 1996-2010.

44. Ferrer, M., O. Golyshina, *et al*. (2007). "Mining enzymes from extreme environments." *Current Opinion in Microbiology* **10**: 207-214.

45. Finn, R. D., J. Tate, *et al*. (2008). "The Pfam protein families database." *Nucleic Acids Research* **36**(Database Issue): D281-D288.

46. Fojan, P., P. H. Jonson, *et al*. (2000). "What distinguishes an esterase from a lipase: A novel structural approach." *Biochimie* **82**: 1033-1041.

47. Franzmann, P. D. (1996). "Examination of Antarctic prokaryotic diversity through molecular comparisons." *Biodiversity and Conservation* **5**: 1295-1305.

48. Gabor, E., K. Liebeton, *et al*. (2007). "Updating the metagenomics toolbox." *Biotechnology Journal* **2**: 201-206.

49. Gabor, E. M., W. B. L. Alkema, *et al*. (2004). "Quantifying the accessibility of the metagenome by random expression cloning techniques." *Environmental Microbiology* **6**: 879-886.

50. Gandhi, N. (1997). "Applications of lipase." *Journal of the American Oil Chemists' Society* **74**: 621-634.

51. Gardy, J. L., M. R. Laird, *et al*. (2005). "PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis." *Bioinformatics* **21**: 617-623.

52. Gasteiger, E., C. Hoogland, *et al*. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *Humana Press*.

53. Gerday, C., M. Aittaleb, *et al*. (1997). "Psychrophilic enzymes: a thermodynamic challenge." *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1342**: 119-131.

54. Gerday, C., M. Aittaleb, *et al*. (2000). "Cold-adapted enzymes: from fundamentals to biotechnology." *Trends in Biotechnology* **18**: 103-107.

55. Gianese, G., P. Argos, *et al*. (2001). "Structural adaptation of enzymes to low temperatures." *Protein Engineering* 14: 141-148.

56. Gilbert, J. A., P. J. Hill, *et al*. (2004). "Demonstration of antifreeze protein activity in Antarctic lake bacteria." *Microbiology* **150**: 171-180.

57. Gilham, D. and R. Lehner (2005). "Techniques to measure lipase and esterase activity in vitro." *Methods* **36**: 139-147.

58. Gillespie, D. E., S. F. Brady, *et al*. (2002). "Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA." *Applied Environmental Microbiology* **68**: 4301-4306.

59. Green, B. D. and M. Keller (2006). "Capturing the uncultivated majority." *Current Opinion in Biotechnology* **17**: 236-240.

60. Griffiths, B. S., K. Ritz *et al*. (2000). "Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity–ecosystem function relationship." *OIKOS* **90**: 279–294.

61. Grocock, R. J. and P. M. Sharp (2002). "Synonymous codon usage in *Pseudomonas* PAO1." *Gene* **289**: 131-139.

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

63. Handelsman, J. (2004). "Metagenomics: Application of genomics to uncultured microorganisms." *Microbiology and Molecular Biology Reviews* **68**: 669-685.

64. Handelsman, J. (2005). "Metagenomics or megagenomics?" *Nature Reviews Microbiology* **3**: 457-458.

65. Handelsman, J. (2005). "Sorting out metagenomes." *Nature Biotechnology* **23**: 38-39.

66. Hårdeman, F. and S. Sjöling (2007). "Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment." *FEMS Microbiology Ecology* **59**: 524-534.

67. Hasan, F., A. A. Shah, *et al*. (2006). "Industrial applications of microbial lipases." *Enzyme and Microbial Technology* **39**: 235-251.

68. Hebraud, M. and P. Poitier (1999). "Cold shock response and low temperature adaptation in psychrotrophic bacteria." *Journal of Molecular Microbiology and Biotechnology* **1**: 211-219.

69. Henderson, I. R., F. Navarro-Garcia, *et al*. (2004). "Type V Protein Secretion Pathway: the Autotransporter Story." *Microbiology and Molecular Biology Reviews* **68**: 692-744.

70. Hogg, I. D., C. Cary, *et al*. (2006). "Biotic interactions in Antarctic terrestrial ecosystems: Are they a factor?" *Soil Biology and Biochemistry* **38**: 3035-3040.

71. Horn, G., R. Hofweber, *et al*. (2007). "Structure and function of bacterial cold shock proteins." *Cellular and Molecular Life Sciences* **64**: 1457-1470.

72. Hulo, N., A. Bairoch, *et al*. (2007). "The 20 years of PROSITE." *Nucleic Acids Research*

73. Hunter-Cevera, J. C. (1998). "The value of microbial diversity." *Current Opinion in Microbiology* **1**: 278-285.

74. Ieva, R., K. M. Skillman, *et al*. (2008). "Incorporation of a polypeptide segment into the β-domain pore during the assembly of a bacterial autotransporter." *Molecular Microbiology* **67**: 188-201.

75. Jacob-Dubuisson F., R. Fernandez, *et al*. (2004). "Protein secretion through autotransporter and two-partner pathways." *Biochim Biophys Acta* **1694:** 235-57.

76. Jaeger, K. E., B. W. Dijkstra, *et al*. (1999). Bacterial biocatalysts: Molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annual Review of Microbiology* **53:** 315-351.

77. Jaeger, K. E. and T. Eggert (2002). "Lipases for biotechnology." *Current Opinion in Biotechnology* **13**: 390-397.

78. Jaeger, K. E., S. Ransac, *et al*. (1994). "Bacterial Lipases." *FEMS Microbiology Reviews* **15**: 29-63.

79. Jaeger, K. E., B. Schneidinger, *et al*. (1997). "Bacterial lipases for biotechnological applications." *Journal of Molecular Catalysis - B Enzymatic* **3**: 3-12.

80. Jain, S., P. van Ulsen, *et al*. (2006). "Polar Localization of the Autotransporter Family of Large Bacterial Virulence Proteins." *Journal of Bacteriology* **188**: 4841-4850.

81. Joseph, B., P. W. Ramteke, *et al*. (2008). "Cold active microbial lipases: Some hot issues and recent developments." *Biotechnology Advances* **26**: 457-470.

82. Joseph, B., P. W. Ramteke, *et al*. (2007). "Standard review cold-active lipases: a versitile tool for industrial applications." *Biotechnology and Molecular Biology Review* **2**: 39-48.

83. Jurgens, G., K. Lindström, *et al*. (1997). "Novel group within the kingdom Crenarchaeota from boreal forest soil." *Applied Environmental Microbiology* **63**: 803-805.

84. Knietsch, A., T. Waschkowitz, *et al*. (2003). "Construction and Screening of Metagenomic Libraries Derived from Enrichment Cultures: Generation of a Gene Bank for Genes Conferring Alcohol Oxidoreductase Activity on Escherichia coli." *Applied Environmental Microbiology* **69**: 1408-1416.

85. Kowalchuk, G. A., A. G. C. L. Speksnijder, *et al*. (2007). "Finding the needles in the metagenome haystack." *Microbial Ecology* **53**: 475-485.

86. Krsek, M. and E. M. H. Wellington (1999). "Comparison of different methods for the isolation and purification of total community DNA from soil." *Journal of Microbiological Methods* **39**: 1-16.

87. Lämmle, K., H. Zipper, *et al*. (2007). "Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning." *Journal of Biotechnology* **127**: 575-592.

88. Langer, M., E. M. Gabor, *et al*. (2006). "Metagenomics: an inexhaustible access to nature's diversity." *Biotechnology Journal* **1**: 815-821.

89. Larkin, M. A., G. Blackshields, *et al*. (2007). "ClustalW2 and ClustalX version 2." *Bioinformatics* **23**: 2947-2948.

90. Lee, H., M. Ahn, *et al*. (2003). "Purification and characterisation of cold active lipase from psychrotrophic *Aeromonas* sp. LPB 4." *Journal of Microbiology* **41**: 22-27.

91. Lee, H. W. and S. M. Byun (2003). "The pore size of the autotransporter domain is critical for the active translocation of the passenger domain." *Biochemical and Biophysical Research Communications* **307**: 820-825.

92. Lee, S. W., K. Won, *et al*. (2004). "Screening for novel lipolytic enzymes from uncultured soil microorganisms." *Applied Microbiology and Biotechnology* **65**: 720-726.

93. Liebeton, K. and J. Eck (2004). "Identification and Expression in E. coli of Novel Nitrile Hydratases from the Metagenome." *Engineering in Life Sciences* **4**: 557-562.

94. Lorenz, P. and J. Eck (2005). "Metagenomics and industrial applications." *Nature Reviews Microbiology* **3**: 510-516.

95. Lorenz, P. and C. Schleper (2002). "Metagenome—a challenging source of enzyme discovery." *Journal of Molecular Catalysis B: Enzymatic* **19-20**: 13-19.

96. Lovell, S. C., I. W. Davis, *et al*. (2002). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." *Proteins: Structure, Function & Genetics* **50**: 437-450.

97. McGuffin, L. J., K. Bryson, *et al*. (2000). "The PSIPRED protein structure prediction server." *Bioinformatics* **16**: 404-405.

98. Morgan-Kiss, R. M., J. C. Priscu, *et al*. (2006). "Adaptation and acclimation of photosynthetic microorganisms to permanently cold environment." *Microbiology and Molecular Biology Reviews* **70**: 222-252.

99. Muyzer, G. (1999). "DGGE/TGGE a method for identifying genes from natural ecosystems." *Current Opinion in Microbiology* **2**: 317-322.

100. Muyzer, G., E. C. de Waal, *et al*. (1993). "Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA." *Applied Environmental Microbiology* **59**: 695-700.

101. Nannipieri, P., J. Ascher, *et al*. (2003). "Microbial diversity and soil functions." *European Journal of Soil Science* **54**: 655-670.

102. Nichols, D. S., J. Bowman, *et al*. (1999). "Developments with Antarctic microorganisms: Culture collections, bioactivity screening, taxonomy, PUFA production and cold-adapted enzymes." *Current Opinion in Biotechnology* **10**: 240-246.

103. Nichols, D. S., P. D. Nichols, *et al*. (1993). "Polyunsaturated fatty acids in Antarctic bacteria." *Antarctic Science* **5**: 149-160.

104. Øvreås, L. (2000). "Population and community level approaches for analysing microbial diversity in natural environments." *Ecology Letters* **3**: 236-251.

105. Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." *Science* **276**: 734–740.

106. Pang, M., N. Abdullah, *et al*. (2008). "Isolation of High Molecular Weight DNA from Forest Topsoil for Metagenomic Analysis." *Asia Pacific Journal of Molecular Biology and Biotechnology* **16**: 35-41.

107. Peck, L. S., M. S. Clark, *et al*. (2005). "Genomics: Applications to Antarctic ecosystems." *Polar Biology* **28**: 351-365.

108. Peck, L. S., P. Convey, *et al*. (2006). "Environmental constraints on life histories in Antarctic ecosystems: Tempos, timings and predictability." *Biological Reviews of the Cambridge Philosophical Society* **81**: 75-109.

109. Prosser, J. I. (2002). "Molecular and functional diversity in soil micro-organisms." *Plant and Soil* **244**: 9-17.

110. Ranjard, L., F. Poly, *et al*. (2001). "Characterization of Bacterial and Fungal Soil Communities by Automated Ribosomal Intergenic Spacer Analysis Fingerprints: Biological and Methodological Variability." *Applied and Environmental Microbiology* **67**: 4479-4487.

111. Ranjard, L., F. Poly, *et al*. (2000). "Monitoring complex bacterial communities using culture-independent molecular techniques: Application to soil environment." *Research in Microbiology* **151**: 167-177.

112. Rappé, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." *Annual Reviews of Microbiology* **57**: 369–94.

113. Rashid, N., Y. Shimada, *et al*. (2001). "Low temperature lipase from psychrotrophic *Pseudomonas* sp. strain KB700A." *Applied and Environmental Microbiology* **67**: 4064-4069.

114. Ray, M. K., G. S. Kumar, *et al*. (1998). "Adaptation to low temperature and regulation of gene expression in Antarctic psychrotrophic bacteria." *Journal of Bioscience* **23**: 423-435.

115. Ritz, K., M. McHugh, *et al*. (2003). *Biological diversity and function in soils: contemporary perspectives and implications in relation to the formulation of effective indicators*. OECD Expert meeting on soil erosion and soil biodiversity indicators, Rome.

116. Robson, L., R. L. Farrell, *et al*. "Bacterial diversity in mineral soils associated with mummified seals in the Miers Valley, Eastern Antarctica." *Unpublished*.

117. Rodrigues, D. F. and J. M. Tiedje (2008). "Coping with our cold planet." *Applied and Environmental Microbiology* **74**: 1677-1686.

118. Rodriguez-Valera, F. (2004). "Environmental genomics, the big picture?" *FEMS Microbiology Letters* **231**: 153-158.

119. Rondon, M. R., P. R. August, *et al*. (2000). "Cloning the soil metagenome: a strategy for assessing the genetic and functional diversity of uncultured microorganisms." *Applied and Environmental Microbiology* **66**: 2541-2547.

120. Rondon, M. R., R. M. Goodman, *et al*. (1999). "The Earth's bounty: assessing and accessing soil microbial diversity." *Trends in Biotechnology* **17**: 403-409.

121. Rosenau, F. and K. E. Jaeger (2000). "Bacterial lipases from *Pseudomonas*: Regulation of gene expression and mechanisms of secretion." *Biochimie* **82**: 1023-1032.

122. Roy, D. and S. Sengupta (2007). "Structural features of a cold-adapted Alaskan bacterial lipase." *Journal of Biomolecular Structure and Dynamics* **24**: 463-470

123. Russell, N. (2000). "Towards a molecular understanding of cold activity of enzymes from psychrophiles." *Extremophiles* **4**: 83-90.

124. Sandaa, R.-A., A. Torsvik, *et al*. (1999). "Analysis of bacterial communities in heavy metal-contaminated soils at different levels of resolution." *FEMS Microbiology Ecology* **30**: 237-251.

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

126. Schmeisser, C., H. Steele, *et al*. (2007). "Metagenomics, biotechnology with non-culturable microbes." *Applied Microbiology and Biotechnology* **75**: 955-962.

127. Schwede, T., J. Kopp, *et al*. (2003). "SWISS-MODEL: an automated protein homology-modeling server." *Nucleic Acids Research* **31**: 3381-3385.

128. Sharma, R., Y. Chisti, *et al*. (2001). "Production, purification, characterization, and applications of lipases." *Biotechnology Advances* **19**: 627-662.

129. Shravage, B. V., K. M. Dayananda, *et al*. (2007). "Molecular microbial diversity of a soil sample and detection of ammonia oxidizers from Cape Evans, McMurdo Dry Valley, Antarctica." *Microbiological research* **162**: 15-25.

130. Singh, B. K., S. Munro, *et al*. (2006). "Investigating microbial community structure in soils by physiological, biochemical and molecular fingerprinting methods." *European Journal of Soil Science* **57**: 72-82.

131. Smith, J. J., L. Ah-Tow, *et al*. (2006). "Bacterial diversity in three different Antarctic cold desert mineral soils." *Microbial Ecology* **51**: 413-421.

132. Stach, J. E. M., S. Bathe, *et al*. (2001). "PCR-SCCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification techniques." *FEMS Microbiology Ecology* **36**: 139-151.

133. Steele, H. L. and W. R. Streit (2005). "Metagenomics: Advances in ecology and biotechnology." *FEMS Microbiology Letters* **247**: 105-111.

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

135. Suzuki, T., T. Nakayama, *et al*. (2001). "Cold-active lipolytic activity of psychrotrophic *Actinobacter* sp. strain no. 6." *Journal of Bioscience and Bioengineering* **92**: 144-148.

136. Targulian, V. O. and S. Goryachkin (2004). "Soil memory: Types of record, carriers, hierarchy and diversity." *Revista Mexicana de Ciencias Geologicas* **21**: 1-8.

137. Tehei, M. and G. Zaccai (2005). "Adaptation to extreme environments: Macromolecular dynamics in complex systems." *Biochim Biophys Acta* **1724**: 404-410.

138. Thanassi, D. G. and S. Hultgren (2000). "Multiple pathways allow protein secretion across the bacterial outer membrane." *Current Opinion in Cell Biology* **12**: 420-430.

139. Theron, J. and T. E. Cloete (2000). "Molecular Techniques for Determining Microbial Diversity and Community Structure in Natural Environments." *Critical Reviews in Microbiology* **26**: 37 - 57.

140. Torsvik, V., F. L. Daae, *et al*. (1998). "Novel techniques for analysing microbial diversity in natural and perturbed environments." *Journal of Biotechnology* **64**: 53-62.

141. Torsvik, V. and L. Øvreas (2002). "Microbial diversity and function in soil: From genes to ecosystems." *Current Opinion in Microbiology* **5**: 240-245.

142. Treusch, H., A. Kletzin, *et al*. (2004). "Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea." *Environmental Microbiology* **6**: 970-980

143. Uchiyama, T., T. Abe, *et al*. (2005). "Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes." *Nature Biotechnology* **23**: 88-93.

144. Ulusu, N. N. and E. F. Tezcan (2001). "Cold shock proteins." *Turkish Journal of Medical Science* **31**: 283-290.

145. Upton, C. and J. T. Buckley (1995). "A new family of lipolytic enzymes?" *Trends in Biochemical Sciences* **20**: 178-179.

146. van Ulsen, P. and T. J. (2006). "Protein secretion and secreted proteins in pathogenic *Neisseriaceae*." *FEMS Microbiology Reviews* **30**: 292-319.

147. Verger, R. (1998). "Interfacial inactivation of lipases: facts and artifacts." *Tibtech* **15**: 32-38.

148. Vincent, W. F. (2000). "Evolutionary origins of Antarctic microbiota: Invasion, selection and endemism." *Antarctic Science* **12**: 374-385.

149. Ward, N. (2006). "New directions and interactions in metagenomics research." *FEMS Microbiology Ecology* **55**: 331-338.

150. Wexler, M., P. L. Bond, *et al*. (2005). "A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase." *Environmental Microbiology* **7**: 1917-1926.

151. Wilson, E. O. (1997). *Understanding and protecting our biological resources*. Washington DC, Joseph Henry Press.

152. Winkler, U. K. and M. Stuckman (1979). “Glycogen, hyaluronate and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*.” *Journal of Bacteriology* **138**: 663–679.

153. Wu, X., H. Jörnvall, *et al*. (2004). "Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance." *Biochemical and Biophysical Research Communications* **313**: 89-96.

154. Wynn-Williams, D. D. (1996). "Antarctic microbial diversity: The basis of polar ecosystem processes." *Biodiversity and Conservation* **5**: 1271-1293.

155. Xu, J. (2006). "Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances." *Molecular Ecology* **15**: 1713-1731.

156. Yen, M. R., C. R. Peabody, *et al*. (2002). "Protein-translocating outer membrane porins of Gram negative bacteria." *Biochim Biophys Acta* **1562**: 6-31.