UNIVERSITY *of the*
WESTERN CAPE

Cloning and Characterization of Three Compost Metagenome-derived α-L-Arabinofuranosidases with Differing Thermal Stabilities

Brent Marc Fortune

A thesis submitted in fulfillment of the requirements for the degree of
MAGISTER SCIENTIAE (M. Sc)

Institute for Microbial Biotechnology and Metagenomics
Department of Biotechnology, University of the Western Cape,
Bellville, Cape Town
South Africa

Supervisor: Prof. Marla Tuffin
Co-supervisors: Dr. Robert Huddy and Dr. Mariette Smart

Declaration

I, Brent Marc Fortune, hereby declare that *Cloning and Characterization of Three Compost Metagenome-derived α-L-Arabinofuranosidases with Differing Thermal Stabilities* is my own original work and that I have accurately reported and acknowledged all sources, and that this document has not previously, in its entirety or in part been submitted at any university for the purpose of obtaining an academic qualification.

…… ……

(Signature) (Date)

Abstract

Second generation biofuels production requires a suite of lignocellulolytic enzymes, acting synergistically to liberate the fermentable monosaccharides contained within agricultural waste materials. The use of thermostable lignocellulosic enzymes in a high-temperature process represents a number of advantages over their respective mesophilic counterparts, including increased solubility of the polymeric lignocellulosic substrates, enhanced enzyme processivity and reduced risk of bacterial and/or phage contamination during fermentation (Turner *et al.*, 2007; Viikari *et al.*, 2007). Alpha-L-arabinofuranosidases (AFases) participate in the deconstruction of lignocellulosic materials by hydrolysing the arabinofuranosyl bonds contained within the hemicellulosic portion of lignocellulose. In this study, three AFases isolated from compost-derived metagenomic DNA were characterised. Three genes derived from the fosmid metagenomic library constructed from 70°C compost were cloned into the pET21a(+) expression vector and expressed in *E. coli* BL21. The heterologously expressed proteins, AFase_H4, AFase_E3 and AFase_D3, were subsequently purified and their biochemical characteristics determined. All three AFases were shown to be active between pH 4.0 and 6.0. AFase_H4 and AFase_E3 displayed the highest activity at 60^{o}C, while AFase_D3 had an optimum temperature at 25^{o}C. Furthermore, the three AFases had differing thermostability profiles. In particular, AFase_E3 maintains 100% residual activity following 60 min incubation at 80^{o}C and 24 hour incubation at 60^{o}C. All three AFases have activity upon p-Nitrophenyl-arabinofuranoside and none against a range of alternative p-Nitrophenyl-glycosidic substrates. Phylogenetic analysis of the catalytic domain, identified within the amino acid sequences of the AFases, suggests that these AFases belong to glycoside hydrolase (GH) family 51. The difference in the AFase primary amino acid sequence motifs were used to infer differences in thermostability. All three AFases indicated similar

biochemical and biophysical characteristics. AFase_E3 was determined to be the most thermostable amongst all three AFases. In the assessment of the three AFases respective suitability for inclusion in thermogenic bioethanol production processes, AFase_E3 was concluded as a suitable candidate for hydrolysis and synergistic testing on natural substrates. AFase_D3 possess the potential to be included in novel mesophilic industrialized process.

Congress Contributions

National Symposia

Brent Fortune, Rob Huddy and Marla Tuffin (2012). *Cloning and characterization of novel alpha-L Arabinofuranosidases from a thermophilic metagenomic library*. Biotechnology Research Open Day. The University of the Western Cape.

Brent Fortune, Rob Huddy, Mariette Smart and Marla Tuffin (2012). *Cloning and characterization of novel alpha-L Arabinofuranosidases from a thermophilic metagenomic library*. The Nautaral Science Faculty Research Open Day. The University of the Western Cape.

Brent Fortune, Mariette Smart, Rob Huddy and Marla Tuffin (2013). *Cloning and characterization of three compost metagenome-derived α-L-arabinofuranosidases with differing thermal stabilities*. Functional Metagenomics Conference. Zebra Lodge, Pretoria

Brent Fortune, Mariette Smart, Rob Huddy and Marla Tuffin (2013). *Cloning and characterization of three compost metagenome-derived α-L-arabinofuranosidases with differing thermal stabilities*. 18th Biennual Conference of the South African Society of Microbiology. Bela Bela, Limpopo Province.

International Symposia

Brent Fortune, Mariette Smart, Rob Huddy and Marla Tuffin (2013). *Cloning and characterization of three compost metagenome-derived α-L-arabinofuranosidases with differing thermal stabilities.* Biomicroworld conference. Madrid, Spain.

Acknowledgements

I am thankful to my almighty omnipotent God for blessing me with undeserved opportunity to be a part of a globally recognized research institution and University for achieving excellency.

Special Thanks and much appreciation to Prof Marla Tuffin for affording me the opportunity to be part of IMBM and your patience and exceptional guidance has allowed me to grow as a future researcher and leader in biotechnology.

To Dr Rob Huddy and Dr Mariette Smart, your valuable input in training me and consistently molding me into a grown researcher is highly appreciated. A huge Thank you.

Dr Colin Ohlhoff, your highly impactful Ph.D has allowed the off-spin of my project. Thank you for your pivot role in my project.

To Lonnie van Zyl for all laboratory technical assistants and intellectual guidance within the laboratory.

Thank you to Dr Heide Goodman for assisting me in all administrative operations and issues related to my course and project. It has certainly allowed a halt-free working environment.

To the National Research Foundation for providing me the Masters innovation bursary from 2012-2014 including all travel funds provided.

To Inqaba biotec for providing me the conference award prize for taking part in the annual Biomicroworld conference held in Madrid, Spain in 2013.

I am grateful and blessed to be part of the Department of Biotechnology and the Faculty of Natural Science. Many social-academic networking have been fostered throughout the duration of my M.Sc course. Thank you to every friend, buddy or random relevant acquaintance relationship. There are too many to mention individually.

Dedications

"To my **Family (Malcolm, Dawn, Carin, Kim and Liezel)**, enormous extended family and to all my friends back home. I LOVE you all and Special Thanks and dedication for supporting and loving me unconditionally through the process of achieving my M.Sc"

Thank You.

Abbreviations

AFase	Arabinofuranosidase
AmpR	Ampicillin Resistance
APS	Ammonium Persulfate
BLAST	Basic Local Alignment Search Tool
CaCl2	Calcium Chloride
CAF	Central Analytical Facility
CamR	Chloramphenicol Resistance
E.coli	*Escherichia coli*
EcoRI	*Escherichia coli* R I
G	Relative centrifugal force
HindIII	*Haemophilus influenzae* III
Kan	Kanamycin
KCl	Potassium Chloride
LA	Luria-bertani agar
Lac	Lactose
LB	Luria-bertani broth
MCS	Multiple Cloning Site
mL	Millilitres
mM	Millimolar
MTP	Microtitre plate
NaCl	Sodium Chloride
NdeI	*Neisseria denitrificans* I
O/N	Overnight
OD	Optical Density
ORI	Origin of Replictation
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction

PCR	Polymerase Chain Reaction
pFos	plasmid Fosmid
pNP	para-Nitrophenol
RE	Restriction Endonuclease
Rpm	Revolutions per minute
SDS	Sodium Dodecyl Sulphate
SOB	Super Optimal Broth
TAE	Tris-base Acetic acid EDTA
TEMED	Tetramethylethylenediamine
TM	Trademark
UV	Ultraviolet
v/v	Volume/volume percentage
w/v	Weight/volume percentage
XhoI	*Xanthomonas holcicola* I
λPstI	Lambda DNA digested with *Providencia stuartii* I
µg	Micrograms
µL	Microlitres
µM	Micromolar

Contents

Chapter 1

Chapter 2

Chapter 3

List of Figures

List of Tables

Chapter 1:

Thermophilic ethanologenesis from enzymatic sacchrificated lignocellulose biomass

1.1 Introduction

Global fossil fuel reserves are depleting at an alarming rate due to the increased demand for natural gas, oil and coal to power diverse industries (Agarwal, 2007). Conventional energy sources cannot sustain these increased energy demands and have a negative environmental and social-economic impact (Amigun *et al*., 2008; Amigun *et al*., 2011). Alternative energy sources may circumvent these adverse impacts without the need to alter or reinvent the mechanisms of existing industries (Demirbas, 2009). These alternative sources may include: geothermal; solar-thermal and wind energy. However, predictions propose that the market for these energy sources will decrease in future, while energy from other sources such as bioethanol and biodiesel, collectively known as biofuels, will increase by 2022 (Ragauskas *et al*., 2006; Limayem and Ricke, 2012). Biofuels are a popular alternative energy source, as it is highly renewable in comparison to fossil fuel derivatives (Shafiee and Topal, 2009). It is the type of fuel that is produced from organic material, because of carbon fixation via the sun's energy and therefore it's a clean renewable source of energy (Gray *et al.*, 2006).

Although the production and use of biofuels is more environmentally friendly compared to fossil fuels, many problems affect the biofuel industry. In Africa, where many countries are stricken with hunger and poverty, the price and availability pressure that the use of a food grade feedstock, such as those used in first generation biofuel production, will place on food security is unacceptable (Amigun *et al.*, 2008). A more socio-economically appealing alternative to a first generation production process is the use of plant derived agricultural waste for biofuel production, known as a second generation biofuels (Sims *et al.*, 2010). However, for second generation biofuels, the recalcitrant nature of the feedstock has to be

overcome to access the fermentable sugars contained within (Naik *et al.*, 2010). Large scale second generation biofuel production needs to be feasible in terms of costs and therefore improving the access to the fermentable sugars contained within the recalcitrant material requires a biotechnological research approach (Wyman, 2007).

1.2 Biofuels as an alternative energy source

1.2.1 First-generation biofuels

First-generation biofuel products are derived from staple food sources such as starch, sugar and vegetable oils (Naik *et al.*, 2010). These feedstocks contain high quantities of fermentable monosaccharides that can be, by fermentation, used to produce bioethanol (Demirbas, 2009). The liberated sugars are fermented by anaerobic or microaerophilic microorganisms such as yeast (*Saccharomyces* species), bacteria (*Zymomonas* species), and fungi (*Trichoderma reesei*) (Naik *et al.*, 2010). Bioethanol is used for transportation or an additive to conventional fuels (Agarwal, 2007). Using biofuels as an additive has shown to have advantages: it decreases the rate at which fuel is used and reduces carbon monoxide and nitrogen oxides emissions which gives an overall improvement in engine efficiencies (Agarwal, 2007). Several countries use bioethanol as a transportation energy source, shown in table 1.1 (Demirbas, 2009).

Table 1.1 Countries producing bio-ethanol (billions of gallons per year) (Adapted from Demirbas, 2009).

Country	Gallons per 2004	Gallons per 2005	Gallons per 2006
USA	3.54	4.26	4.85
Brazil	3.99	4.23	4.49
(the) PRC	0.96	1	1.02
India	0.46	0.45	0.5
France	0.22	0.24	0.25
Germany	0.07	0.11	0.2
Russia	0.2	0.2	0.17
Canada	0.06	0.06	0.15
Thailand	0.07	0.08	0.09

The industrial consumption of first-generation fuels affects staple food prices and security in most developing countries and the first generation feedstocks that are used for bioethanol production will have an effect on the price of the feedstocks allocated for staple food products (Demirbas, 2009).

Table 1.2 indicates that the first-generation biofuels cannot be industrialized in developing countries, owing to the increase in staple food commodities even if adverse climate change would be mitigated (Sasson, 2012). The estimated percentage increase in prices would create a food security crisis in most developing countries if first-generation bioethanol was to be produced as the crops to be used as feedstock would result in decreased staple food commodities (Srinivasan, 2009; Havlik *et al.*, 2010). Therefore, developing countries such as South Africa cannot benefit from first-generation biofuel production.

Table 1.2 An estimated percentage increase in staple food costs by 2030 (taken from Sasson, 2012).

Crop	Without climate change (%)	With climate change (%)
Paddy rice	72	107
Wheat	53	82
Maize	71	126
Processed rice	34	48

1.2.2 Second-generation biofuels

Biofuels that are produced from non-food based or lignocellulose biomass are referred to as second-generation biofuels (Rubins, 2008). Lignocellulose is a complex macromolecule (Figure 1.1) that's located within all plant cell walls and its function is to confer structural rigidity to the plant cells and protection from pathogenic invasion from bacterial, fungal and viral proteins; therefore it's highly recalcitrant to biological or chemical degradation, hence its function (Himmel *et al.*, 2007). Lignocellulose is composed of a fibrous biomatrix consisting predominantly of three polymer types; cellulose, hemicellulose and lignin (Deobald & Crawford, 1997) (Figure 1.1). The cellulose and hemicellulose polymer material is composed of monosaccharides that can be fermented to bioethanol (Rubins, 2008). For this reason lignocellulosic biomass is regarded as a suitable renewable source of natural biological material for bioethanol production.

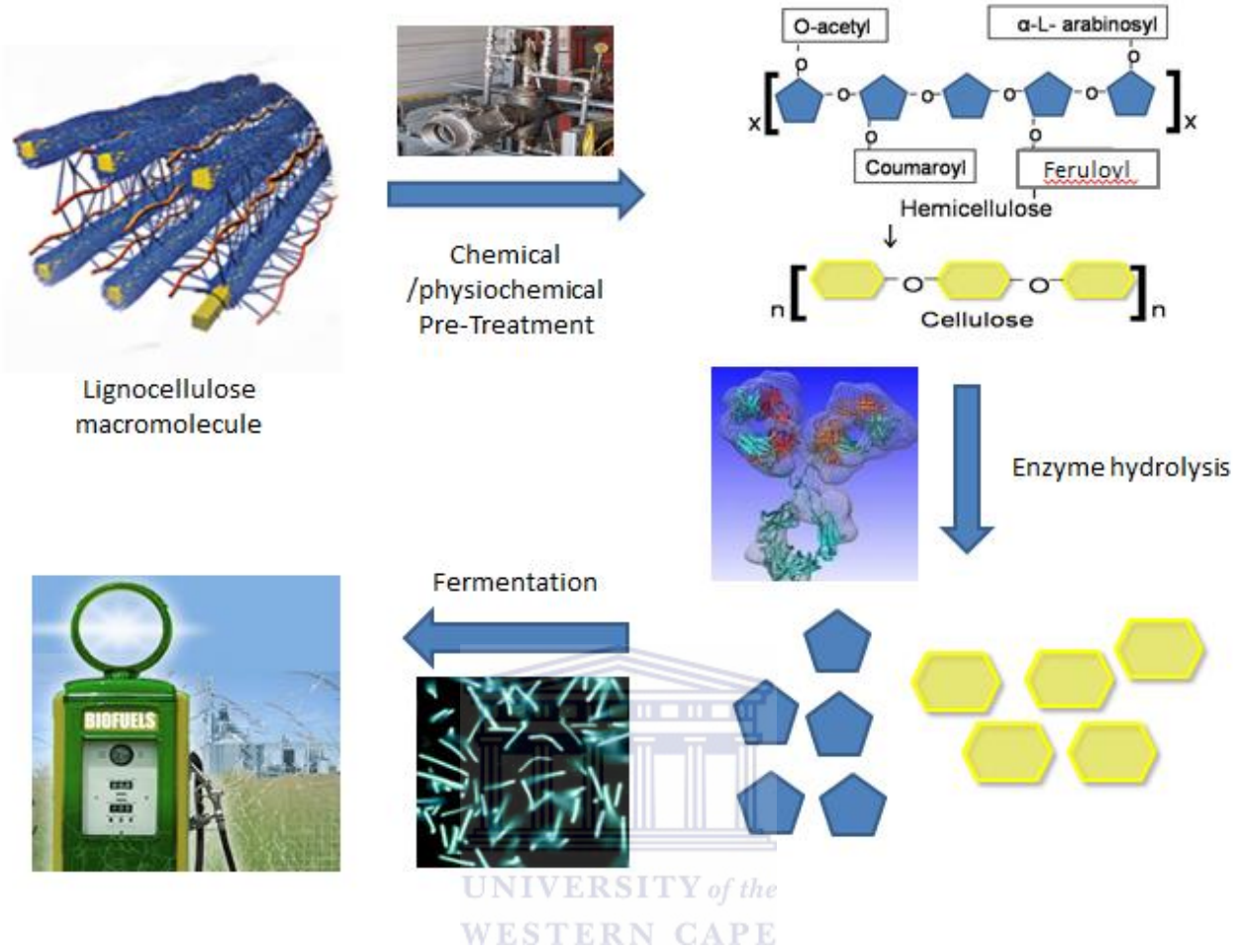


Figure 1.1 The flow diagram depicting the elementary process of second generation biofuels.

Bioethanol production from lignocellulose substrates also eradicates the need for land availability and the use of agricultural products intended for staple food commodities (Havlik *et al.*, 2010). However, using lignocellulose still includes many challenges that need to be overcome that hinder the industrial production of bioethanol, and which will negatively impact the economic feasibility of second generation biofuels in most developing countries (Koh and Ghazoul, 2008). Various process barriers need to be overcome, particularly in reducing the release of toxic phenolic compounds that can be inhibitory to the fermenting microbes or become factor to the loss of fermentable sugars (Rubins, 2008); and improving

enzyme technologies for lignocellulose degradation. Before elaborating on these, the structure of lignocellulose will first be presented.

1.3 Structure of lignocellulose

Lignocellulose is a macromolecule and therefore known to be highly complex and inert. It is composed of hexose and pentose monosaccharides that can be easily fermented. However, to make these monosaccharides accessible for fermentation it requires a suite of lignocelluosic degrading enzymes to hydrolyze the plant biomass in synergy. The suite of enzymes are categorized as cellulases, hemicellulases and accessory enzymes.

1.3.1 Cellulose

Cellulose is the most abundant biopolymer on earth and it's a complex macromolecule that is located within all plant cell walls (Taylor *et al.*, 2002). It accounts for approximately 40 % of the total weight distribution of the lignocellulose structure and the higher order of this structure is held together by covalent, hydrogen bonds and Van der Waals forces (Agbor *et al*., 2011). The cellulose component of the lignocellulose structure is made up of cellulose microfibrils, these microfibrils are composed of long covalently linked cellulose polymers that are bonded together. The extensively long cellulose polymers are formed by β-1,4-linked D-glucose residues as depicted in Figure 1.2 (Bayer *et al*., 1998). The overall cellulose structure is bound and encased in a network of hemicellulose and lignin to form the lignocellulose within the plant cell walls (Taylor *et al.*, 2002).

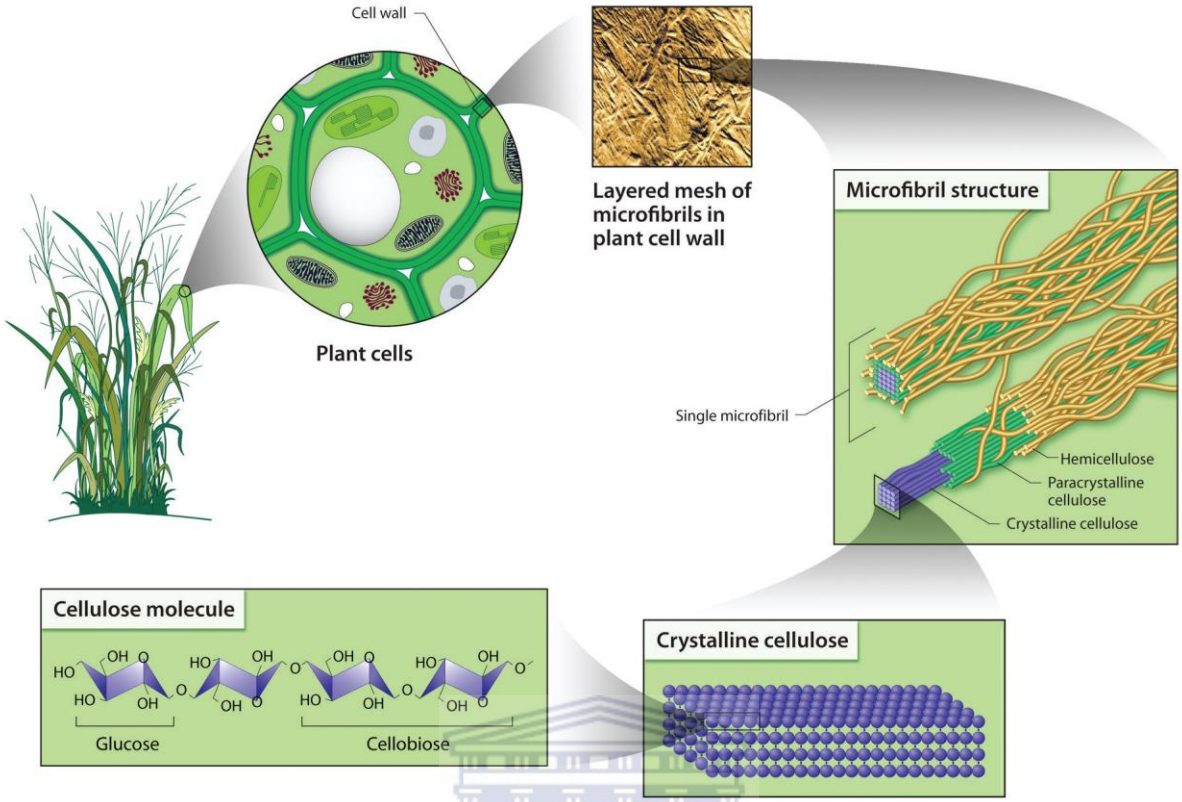


Figure 1.2 The cross-section dissecting arrangement of the monomers within cellulose (Adapted from Quiroz-Castañeda and Folch-Mallol, 2013).

1.3.2 Hemicellulose

Hemicellulose is the second most abundant heteropolymer on earth accounting for approximately 25 % of the lignocellulose weight distribution and generally consists of a xylan backbone which is composed of β-1,4-linked xylose residues as depicted in Figure 1.3 (Saha, 2003). Depending on the biomass source, the backbone can alternate between pentose (D-xylose and L-arabinose) and hexose sugars (D-galactose, L-galactose, D-mannose, L-fructose and L-rhamnose) (Peng *et al*., 2009; Scheller and Ulvskov, 2010). Attached to the xylan backbone are side chain residues which permits hemicellulose to interact with itself and the cellulose to form a strong inert complex macromolecule. The side chains present are arabinofuranosyl, glucuronic acid, 4-O-methyl ether, acetic, ferulic and *p*-coumaric acids

(Figure 1.4) (Saha, 2003; Peng *et al*., 2009; Scheller and Ulvskov, 2010), however the type and quantity of these side-chain residues are dependent on the source of the plant biomass (Saha, 2001).

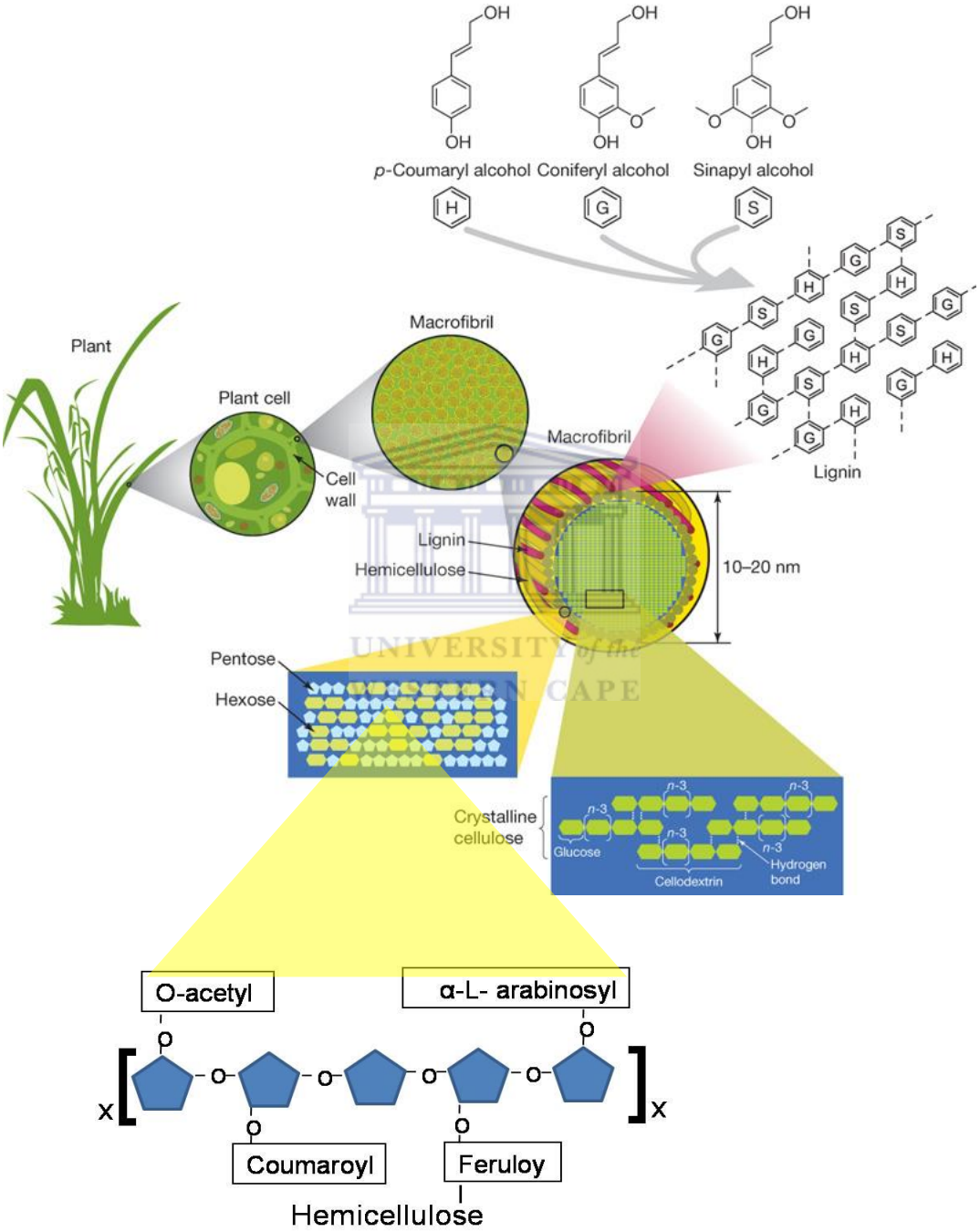


Figure 1.3 The dissected diagram of the sugar and side chain constituents of lignocellulose and indicating the molecular position and structure of hemicellulose (Adapted from Rubins, 2008).

1.3.3 Lignin

Although lignin has no clear defined function, it is assumed that its sole purpose is to form a final "rigid" character to the cell wall of the plant (Chen and Dixon, 2007). The extensive inert nature of this polymer's bond interaction with hemicellulose and cellulose provides the plant structural rigidity and resistance to enzymatic degradation and oxidative stress (Hendriks and Zeeman, 2009). The hemicellulose encompasses the cellulose and interacts with lignin through aryl ether bonds (Figure 1.5) to maintain the shape and function of the lignocellulose structure (Chen and Dixon, 2007). However, it has been reported that lignin is an amorphous heteropolymer, contributing to 20 % of the weight distribution, composed of three main phenolic monomers (phenyl propane units) known as *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol as illustrated in Figure 1.4 (Agbor *et al*., 2011). These monomers form strong ionic and Van der Waals forces that aids in the additional recalcitrant nature of lignocellulose (Ragauskas *et al.*, 2006).

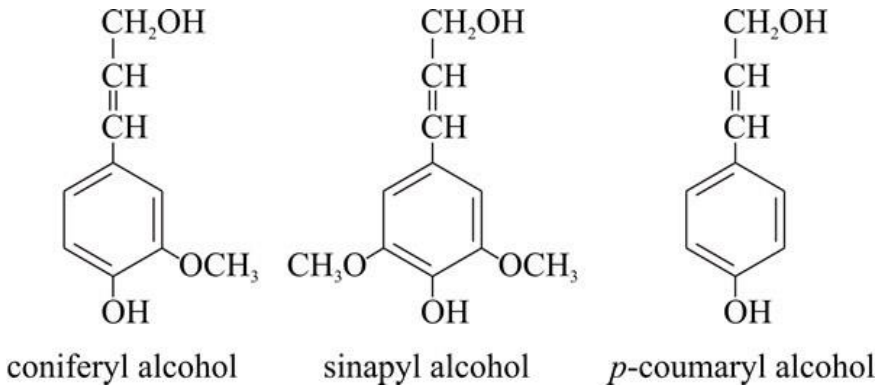


Figure 1.4 The three main phenolic monomer constituents that forms lignin within lignocellulose (Adapted from Moore *et al*., 2011).

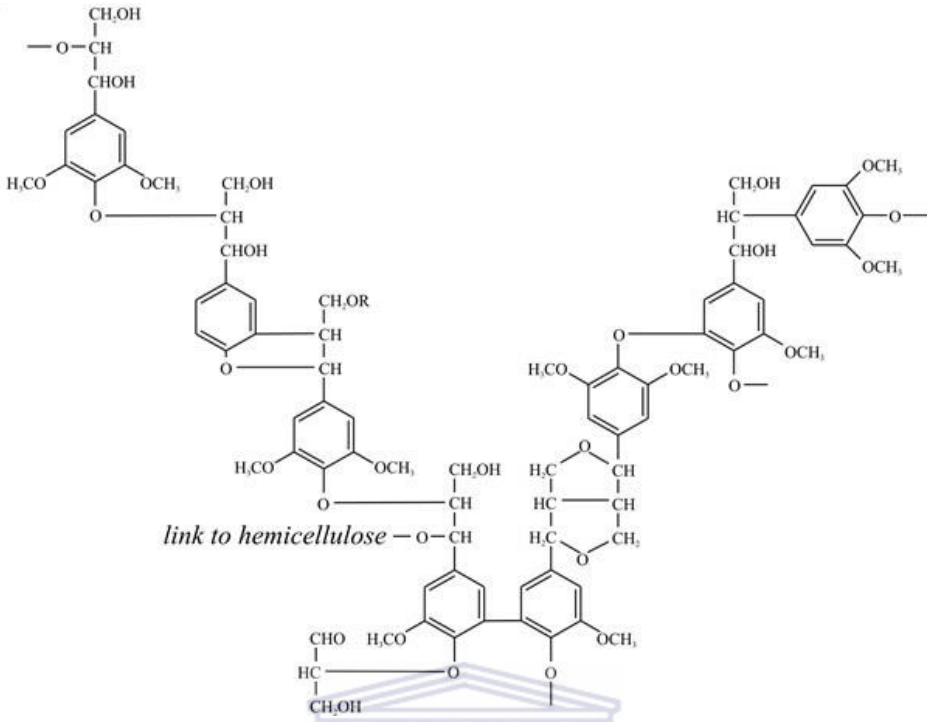


Figure 1.5 The illustration of an aryl ether bond formed between hemicellulose and one of the lignin monomers (Taken from Moore *et al*., 2011).

1.4 Lignocellulosic hydrolytic enzymes

1.4.1 Cellulases

The enzymes within this class of lignocellulosic enzymes act synergistically to catalyze the hydrolysis of the cellulose substrate to release glucose residues and are divided into three main types (Figure 1.6) (Howard et al., 2003). Endoglucanases catalyse the oxidation of the carboxy methyl of the cellulose crystal, yielding single chains of cellulose referred to as cello-oligosaccharides. Exoglucanases catalyse the oxidation of the non-reducing ends of the cellulose polymer, which liberates cellobiose. The β-glucosidases catalyse the hydrolysis of cellobiose to glucose (Karmakar and Ray, 2011).

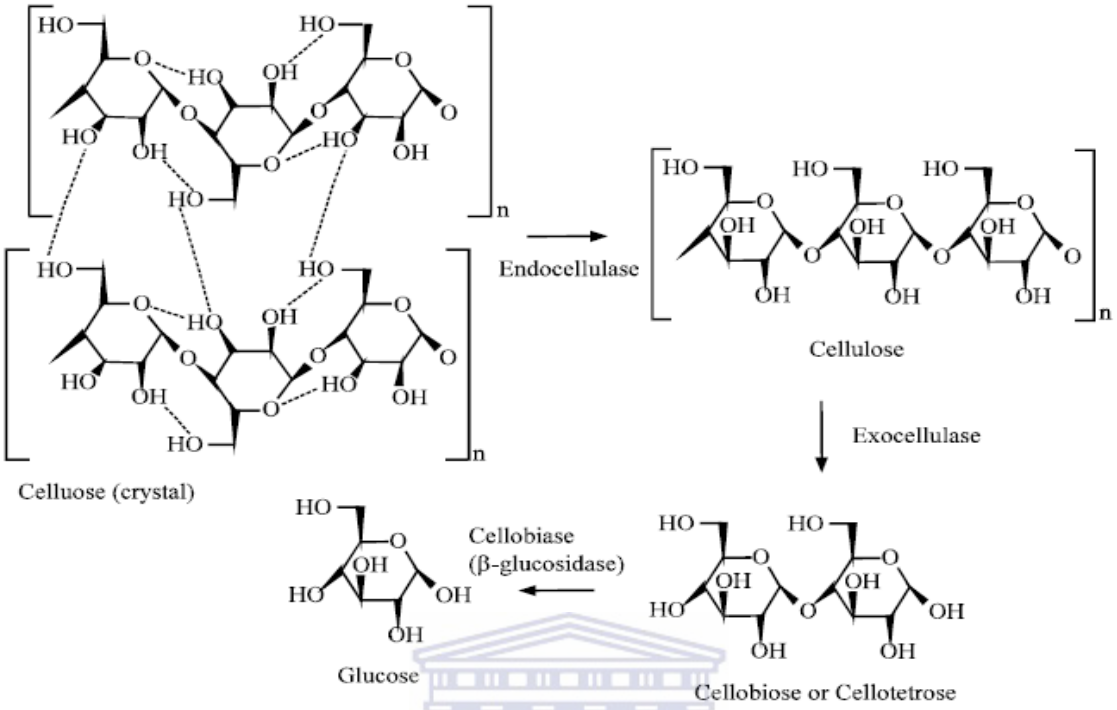


Figure 1.6 The catalytic activity of the cellulose group of enzymes: endo-glucanases, exo-glucanases and β-glucosidases (Taken from Karmakar and Ray, 2011).

1.4.2 Hemicellulases

As hemicellulose is highly variable in its structure, the complete degradation requires the synergistic action of numerous enzymes. These include primarily the endo-xylanases, exo-xylanases and β-xylosidases, as well as an assortment of accessory enzymes. Each of the xylanases exhibit its own mode of action on a hemicellulosic substrate similar to the mode of action of the cellulases (Shallom and Shoham, 2003). The endo-xylanases catalyse the oxidation/hydrolysis of the glycosidic bonds within the xylan molecular structure producing mainly β-D-xylopyranosyl oligomers. The β-xylosidases (also referred to as exo-xylanases)

have affinities for xylobiose or/and larger xylo-oligosaccharides, and hydrolyse small xylo-oligosaccharides and xylobiose to xylose (Figure 1.7) (Khan, 2010).

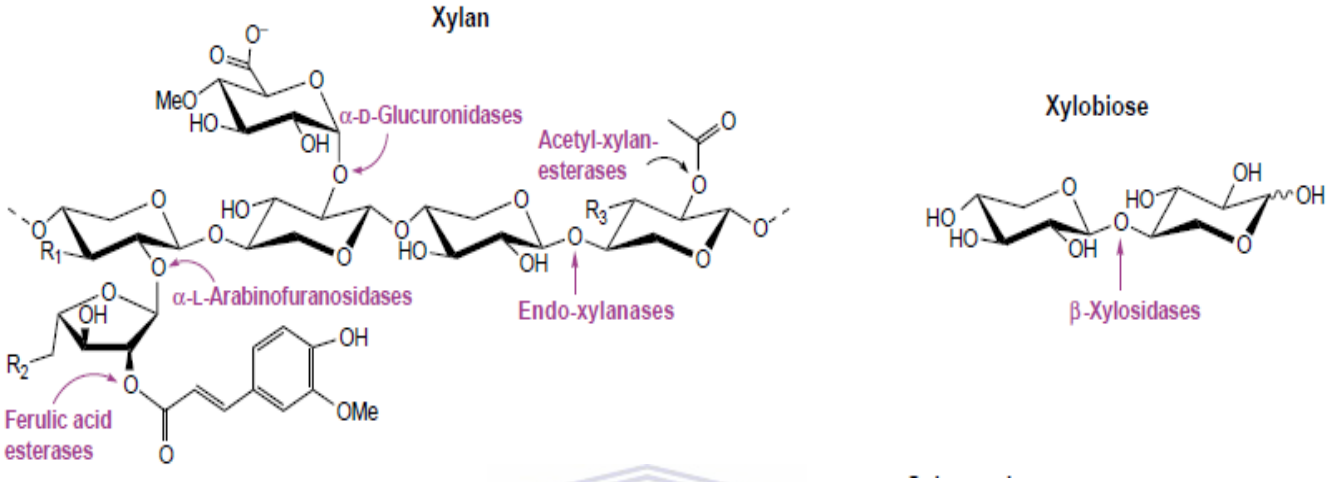


Figure 1.7 The glycosidic hyrdrolytic activity of the hemicellulose group of enzymes: endo-xylanases and β-xylosidases/exo-xylanases (Taken from Shallom and Shoham, 2003)

1.4.3 Accessory enzymes: α-L-arabinofuranosidases (AFases)

The accessory enzymes are required for the hydrolysis of various substituted xylans and this includes α-D-glucuronidases, acetyl xylan esterases, ferulic acid esterases, α-galactosidases, acetyl mannan esterases and α-L-arabinofuranosidases (Shallom and Shoham, 2003). When the side-chains are cleaved it allows steric access of the xylanases to liberate the xylose for subsequent fermentation (Shallom and Shoham, 2003). AFases catalyze the hydrolysis of the arabinofuranosyl bonds on the xylan backbone, specifically the α-1,2-; α-1,3- and α-1,5-arabinofuranosyl linked to the xylan, galactan or arabinan polymer (Koseki *et al.*, 2003). There have been recorded cases of the positive synergy effects that AFases impart when supplemented with endo- and exo-xylanases and other hemicellulosic accessory enzymes; resulting in improved yields of liberated pentose monomers (Alvira *et al.*, 2011). As such,

AFases are considered to impart important synergism in industrial scaled up production of biofuel processes (Numan and Bohsle, 2006; Antizar-Ladislao and Turrion-Gomez, 2008; Yeoman *et al.*, 2010). AFases are grouped into different classes of Glycoside Hydrolases according to the substrate preference (Naumoff, 2011a; Naumoff, 2011b).

1.4.3.1 Glycoside hydrolase families of AFases

Glycoside hydrolases (GH enzymes) are enzymes that utilize a water molecule to hydrolyze the glycosidic bond between two monosaccharides within a sugar polymer and traditionally GH enzymes would be grouped according to a substrate preference. However this approach to classification is complicated for GH enzymes with broad substrate specificity and dual activity capability (Krasikov *et al.*, 2001; Tymowska-lalanne and Kreis, 1998). Nevertheless AFases have been grouped into four major GH families according to amino acid similarities within the catalytic domain and these are GH 43, 51, 54 and 62 (Yeoman *et al.*, 2010; CAZy, http://www.cazy.org). Some AFases have been reported to possess dual functionality and these have been categorized into GH43 (Wagschal *et al.*, 2009). AFases of the GH51 family are predominantly more thermostable while GH 54 and GH 62 AFases are only derived from fungal species (Miyazaki, 2005; Wan *et al.*, 2007; Hashimoto *et al.*, 2011; Sakamoto *et al.*, 2011).

1.4.3.2 Substrates hydrolyzed by AFases

AFases are biochemically characterized based on the hydrolysis of synthetic and natural substrates. *p*-Nitrophenyl-glycosides (pNP-glycosides) are the typically used synthetic substrates to test for substrate specific hydrolysis. The two most common for determining

AFase activity are *p*-nitropheyl-α-L-arabinofuranoside and *p*-nitrophenyl-β-D-xylopyranoside. These represent the α-1,3-arabinofuranosyl bonds and β-1,4-xylopyranoside respectively found in hemicellulose (Saha, 2000; Taylor *et al.*, 2006; Wagschal *et al.*, 2009). Other substrates include L-arabinose containing polysaccharides namely; arabinoxylan, arabinan and debranched arabinan (Figure 1.8). Arabinoxylan consists of a xylan backbone with α-1,3- and α-1,2- linked arabinofuranosyl residues while arabinan consists of a α-1,5-linked arabinofuranosyl linked backbone with α-1,3- and α-1,2- linked arabinofuranosyl residues. Debranched arabinan consists only of the α-1,5- linked arabinofuranosyl linked backbone (Miller, 1959).

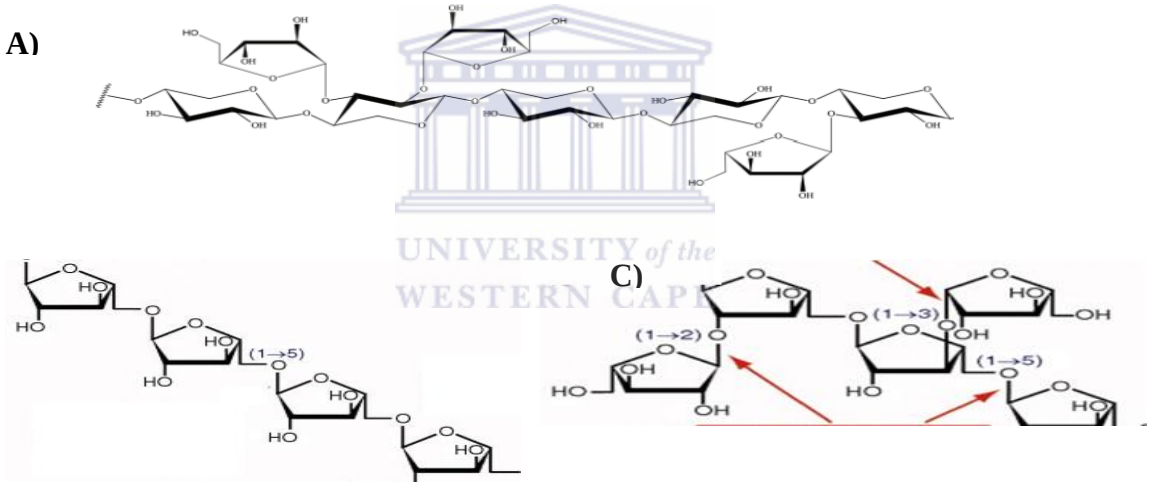


Figure 1.8 The diagramatic representation of the L-arabinose containing polysaccharides. A) Arabinoxylan, B) Debranched arabinan and C) Arabinan (Adapted from Yeoman *et al.*, 2010).

1.5 Bioethanol production from lignocellulosic biomass

1.5.1 Pretreatment

Bioethanol produced from plant biomass requires pretreatment steps and these pretreatment methods allow the recalcitrant nature of the lignocellulose biomass to be reduced (Chen and

Qiu, 2010). Reduction in the recalcitrance causes the crystalline structure to be oxidized which in-turn allows the polysaccharides to be liberated from bonds that inhibit enzymatic hydrolysis (Galbe and Zacchi, 2007). There are numerous chemical, physico-chemical and biological methods of lignocellulosic biomass pretreatment and the choice of method is dependent on the biochemical composition of the lignocellulosic biomass (Galbe and Zacchi, 2007). Chemical application methods have been in existence for several decades and have proven to be useful, however phenolic and other aromatic compounds released have been reported to increase inhibition of downstream fermentation of the liberated monosaccharides (Galbe and Zacchi, 2012). The most novel chemical pretreatment process to date is the application of ionic liquids to fractionate the crystalline structure (Galbe and Zacchi, 2012). The method uses large organic cations (1-butyl-3- methylimidazolium) and smaller inorganic anions which exist as liquid phase at room temperature (Heinz *et al.*, 2005). The function is to break the bonds of the crystalline lignocellulose causing the immediate solubility of the macromolecules without the formation of the fermentation inhibitory molecules (Heinz *et al.*, 2005). Steam explosion is the most common physico-chemical pretreatment and uses highly pressurized saturated steam to break the crystalline bonds followed by an immediate decompression (McMillan, 1994). The process uses temperatures between 160^{o}C-260^{o}C and pressure points of 0.69 – 4.83 MPa for up to several minutes before the biomass becomes exposed to atmospheric pressure (Sun and Cheng, 2002). There are biological means of pretreating biomass and it has proven to be the most cost effective, environmentally safe and less energy demanding application of lignocellulose pretreatment (Talebnia *et al.*, 2010). Fungi (such as white, brown and soft) contain lignin-degrading enzymes such as peroxidases and laccases (Okano *et al.*, 2005). Talebnia and co-authors (2010) have suggested that the lignin-degrading enzymes, used as a biological pretreatment step with *Aspergillus niger* and

Aspergillus awamori, have proven to yield higher liberated sugars during subsequent enzymatic hydrolysis.

1.5.2 Enzymatic hydrolysis

Subsequent to the pretreatment of lignocellulose an enzymatic sacchrification (generally referred to as enzymatic hydrolysis) process occurs. In this process, a suite of enzymes (Table 1.3) are required to hydrolyze the bonds of lignocellulose to decrystallize its structure completely (Alvira *et al.*, 2010). These lignocellulases can be divided into 3 categories: cellulases, hemicellulases and lignin-degrading enzymes (Alvira *et al.*, 2010). The hydrolysis process is complex because of the different pre-treatment methods and the diverse lignocellulose biomass types that exist. Therefore depending on the polymer constituents of the biomass, different pre-treatment methods will result in varied enzyme accessibility to the cellulose and hemicellulose polymers (Kumar *et al.*, 2009a; Kumar *et al*., 2009b). Enzymatic hydrolysis is dependent on the type of substrate used, the method and severity of the pretreatment as well as the type of enzyme and dosage (Kumar *et al*., 2009a; Kumar *et al*, 2009b). There are numerous limiting factors that affect the hydrolysis such as cellulose crystallinity, cellulose degree of polymerization, substrate´s available surface area, lignin barrier, hemicellulose content, and porosity of the pretreated lignocellulose biomass (Chang and Holtzapple, 2000, Kumar *et al*., 2009a, Zhang and Lynd, 2004; Chandel *et al*., 2007). The end-products of the hydrolysis can have an adverse effect on the enzymes' kinetics by decreasing its activity e.g. glucose can inhibit the β-glucosidase if reached at a particular concentration, therefore substrate concentration is a limiting factor to these enzymes (Galbe and Zacchi, 2012). All these factors affecting enzymatic sacchrification need to be addressed

such that these enzymes synergistically liberate the monosaccharides from lignocellulose biomass for subsequent fermentation to occur efficiently (Kuo and Lee, 2009).

Table 1.3 Enzymes involved in the hydrolysis of pretreated lignocellulose biomass. (EC numbers obtained from www.CAZy.org).

Cellulases	**EC No.**	**Hemicellulases**	**EC No.**
β-glucosidases	3.2.1.21	Endo-β-1,4-xylanases	3.2.1.8
Exoglucanases	3.2.1.91	β-xylosidases	3.2.1.37
Endoglucanases	3.2.1.4	Acetyl xylan esterases	3.1.1.72
		Ferulic acid esterases	3.1.1.73
		α-L-Arabinofuranosidases	3.2.1.55

1.5.3 Fermentation

After the pretreatmeant and enzymatic saccrification stages, microbial fermentation occurs on the liberated hexose or pentose sugar molecules (Gupta *et al.*, 2012). Conventionally, the hexose sugars are readily fermented by yeast such as *Saccharomyces cerevisiae* and bacteria such as *Zymomonas mobolis* (Gupta *et al.*, 2012; Delgenes *et al.*, 1996). Pentose sugars are readily fermented by yeasts *Pichia stipitis*, *Candida shehatae*, and *Pachysolen tannophilus* (Gupta *et al.*, 2012; Delgenes *et al.*, 1996). Currently, *Saccharomyces cerevisiae* and *Zymomonas mobolis* yield high amounts of ethanol of between 90-97% w/w but have inheritable low resistance to ethanol allowing concentrations of 100g of cells per litre yield of ethanol (10% w/v) (Claassen *et al.*, 1999). The xylose utilizing yeasts yield a low percentage of ethanol from fermenting xylose and have the inherit ability to constitutively catabolise the ethanol produced (Karakashev *et al.*, 2007). Furthermore, inhibitory molecules formed during

pretreatment and enzymatic sacchrification of lignocellulose biomass has an adverse effect on the fermentation integrity of the micro-organisms and causes the yield of ethanol to decrease significantly (Hahn-Hagerdal *et al.*, 2007). As a result it has been proposed that simultaneous fermentation of hexose and pentose sugars could improve ethanol yields, and engineered *Escherichia coli, Saccharomyces cerevisiae*, and *Zymomonas mobilis* strains have been developed which can utilize both xylose and glucose as electron acceptors during fermentation through genetic engineering (Hahn-Hagerdal *et al.* 2007; Sedlak and Ho 2004). These recombinant strains have the ability to simultaneously ferment xylose and glucose to ethanol. However, constraints to this idea have emerged. The transporters of *S. cerevisiae* and *Z. mobilis* have low affinity kinetics for xylose uptake in the presence of high glucose concentrations (Lee *et al.*, 2002). Other constraints associated with the conventional and non-thermal fermentation of lignocellulose (Zhao *et al.*, 2009a; Zhao *et al.*, 2009b) led to an interest in a thermogenic bioethanol process.

1.5.4 Conventional bioethanol production

Industrial lignocellulosic bioethanol production can occur in four strategies (Figure 1.9): Consolidated bioprocessing (CBP), separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF) and direct microbial conversion (DMC) (Talebnia *et al.*, 2010). Consolidated bioprocessing is a single-step alternative, an approach which combines enzyme production, biomass pretreatment and the fermentation of hexose and pentose sugars. CBP is advantageous because it bypasses the need for a dedicated enzyme generation step, thereby avoiding the associated costs involved with enzyme production (lynd *et al*., 2005). SHF is highly costly, inherits negative factors that inhibits high bioethanol yields due to the inhibitory molecules formed during the hydrolysis process which

has a direct impact on the fermentation integrity of the anaerobes used (Ohgren *et al.*, 2007). Glucose or xylose accumulation has an inhibitory effect on the fermentation if it exceeds a particular concentration or when these monosaccharides become metabolized (Kumar *et al.*, 2009b; Kumar *et al*., 2009c). SSF decreases capital costs and improves the overall process by eliminating the enzyme activity inhibition caused by glucose or xylose accumulation; however ethanol yield decreases because of intolerance of the fermenting mesophile to the end-product inhibition (Claassen *et al.*, 1999; Talebnia *et al.*, 2010).

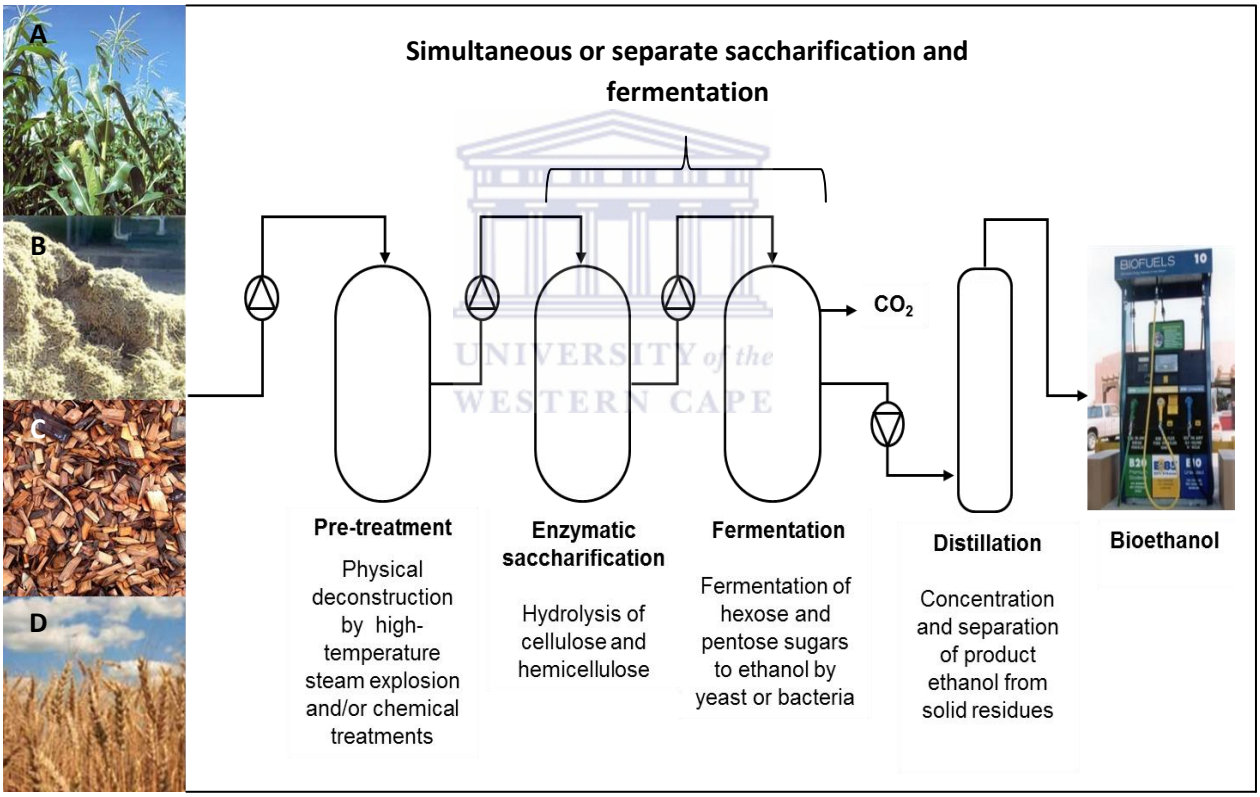


Figure 1.9 The graphical illustration of the main steps required for the hydrolytic liberation of fermentable sugars through enzymatic saccharification from pretreated lignocellulosic biomass feedstocks such as maize husks (A), sugar cane bagasse (B), (C) wood chips and wheat husks (D). A potentially efficient process would prefer the simultaneous saccharification and fermentation, from an appropriate thermophilic micro-organism, to yield bioethanol. (Adapted from Viikari *et al.*, 2007 and reproduced with permission from Dr R. Huddy).

1.5.5 Thermogenic bioethanol production

SSF bioethanol production (Figure 1.10) at high temperatures provides several advantages, making bioethanol from lignocellulose more feasible: At elevated temperatures increased solubility of the pretreated biomass allows an increase in surface area for lignocellulosic enzymes to hydrolyze the biomass. The rate of enzymatic bioconversion increases at high temperatures decreasing the time and the quantity of enzyme needed. Risk of microbial contamination in the fermentation decreases significantly and product recovery is more efficient because at high temperatures the volatility of ethanol is exploited by collecting it as a vapour (Georgieva and Ahring, 2007). Continuous removal of the ethanol vapour prevents ethanol toxicity or inhibition to the thermophilic fermenting micro-organism and therefore in comparison to a mesophilic process, thermophilic fermentation-based ethanol production increases the efficiency of the bioethanol process (Almarsdottir *et al*., 2011; Limayem and Ricke, 2012; Svetlitchnyi *et al*., 2013). Thermophilic enzymes can hydrolyze the pretreated biomass at the same high temperature at which the thermophilic anaerobes can ferment the hydrolyzates, improving process feasibility of bioethanol from lignocellulose as depicted in Figure 1.10 (Kadar *et al.*, 2004). Thermogenic DMC is the improved process of bioethanol production for it compresses the enzymatic hydrolysis and fermentation using a single thermophilic anaerobe that contains the genes for hydrolyzing pretreated lignocellulose biomass under thermophilic temperature ranges (Demirbas, 2005). *Fusarium oxysporum* F3 is a fungus that contains the genes for hydrolyzing pretreated lignocellulose, but has very low ethanol tolerance (Christakopoulos *et al.*, 1991). Therefore thermophilic ethanol tolerant fermenting micro-organisms such as *Thermoanaerobacter sp* and *Geobacillus thermoglucosidasius* can be engineered to express thermostable lignocellulosic hydrolyzing

genes to have the ability to ferment the pretreated hydrolyzates at the same high temperatures which will enhance cost reduction and feasibility at an industrial scale (Fong *et al.*, 2006; Zhao *et al.*, 2009a). Thermal processes for an industrial scale biofuel production have advantages over lower to moderate temperatures. It will require thermophilic micro-organisms combined with thermostable enzymes' biochemical properties to execute the production. This will also require a thermophilic consolidating bioprocessing on bioethanol production as depicted in Figure 1.10

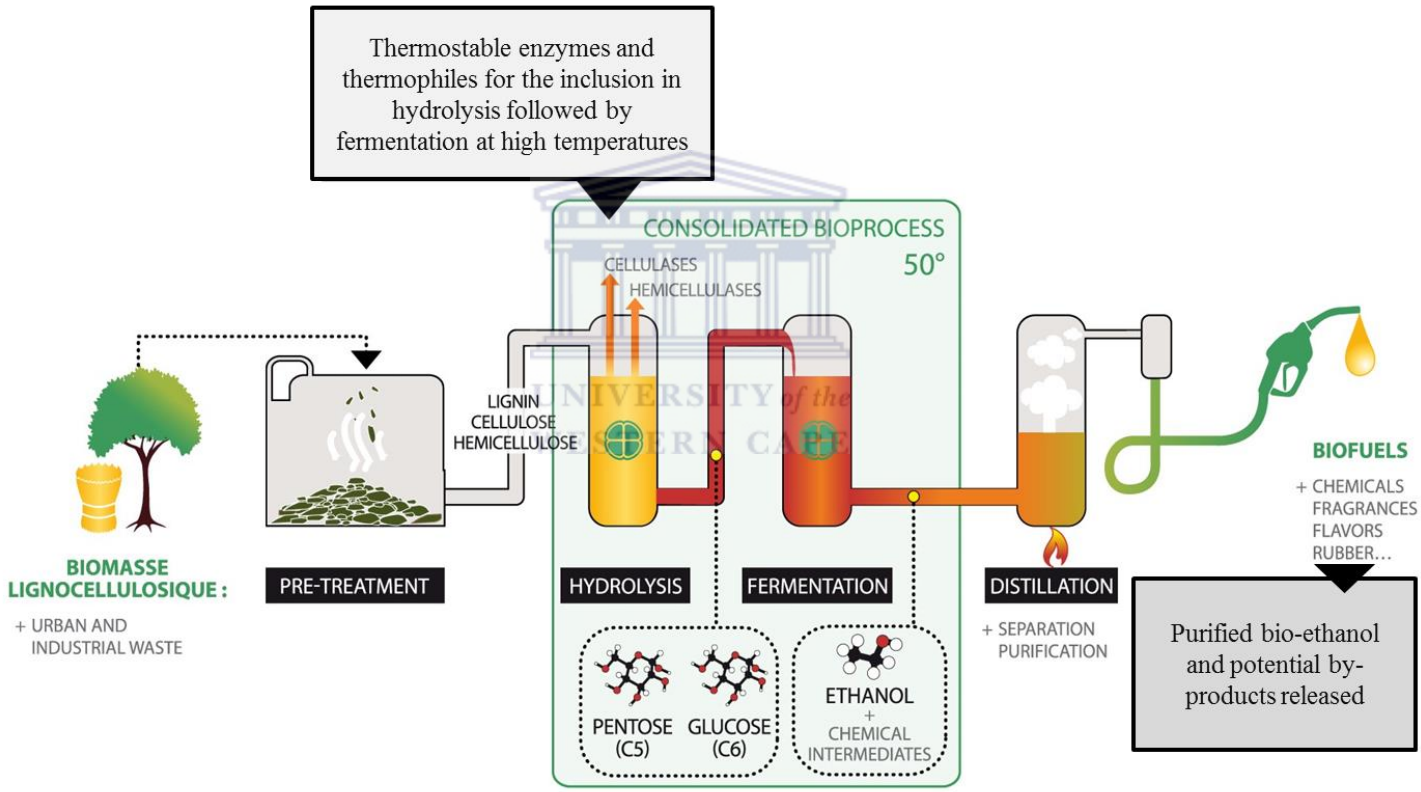


Figure 1.10 A diagrammatic illustration of a consolidating thermophilic bioethanol production process highlighting the potential simultaneous hydrolysis and fermentation within the same high thermal zone (Adapted from Hasunoma and Kondo, 2012).

1.5.5.1 Thermophiles

Micro-organisms that flourish at high temperature zones or in extreme heat environments are referred to as thermophiles. Thermophiles are classified in two distinct groups known as obligate and facultative. The former refers to micro-organisms that can only thrive within high temperatures while the latter refers to thermophiles that can equally thrive at moderate to high temperatures (Jones *et al*., 1984). Thermophilic micro-organisms are isolated from various thermal environments such as geothermal volcanic environments, hot springs, deep sea hydrothermal vents and decaying plat matter such as compost (Chang and Yao, 2011; Lynd *et al.*, 2008). Thermophiles are mostly prokaryotes and proliferate optimally between 55^{o}C-65^{o}C with a maxima growth at 90^{o}C whereas hyper-thermophiles proliferate between 90^{o}C – 113^{o}C (Wiley *et al.*, 2008; Bhalla *et al.*, 2013). Numerous hyper-thermophiles are Archaea and require elemental sulphur for growth as an electron acceptor during respiration (Kletzin *et al*., 2004). Thermophiles contain proteins/enzymes that are thermostable by default and therefore find many applications as industrial biocatalysts (Szijarto *et al.*, 2008). The thermophiles can be used at higher temperatures for fermentation of lignocellulosic material, or can be a source of enzymes for the hydrolysis stage (Viikari *et al.*, 2007).

1.5.5.2 Thermostable enzymes

Thermostable proteins possess unique biochemical characteristics, in comparison to non-thermostable proteins, that confer its ability to resist being denatured by high temperature (Taylor and Vaisman, 2010). The interior of thermostable enzymes have higher organized hydrophobic interactions amongst its amino acid residues and polar residues for increased ionic bonds, thus the structure of the protein is strengthened by larger amounts of hydrogen

and non-covalent bonds as well as disulphide bridges (Elcock, 1998; Mathews *et al.*, 2000; Kumar *et al.*, 2000; Pace *et al.*, 2000). Amino acids such as proline allow decrease in flexibilities of the different polypeptides which enables the proteins to be folded in a high temperature-stable conformation by special chaperone proteins (Mathews *et al.*, 2000). A growing interest in thermostable biocatalysts have emerged over the last decade owing to its flexibility adaptations to harsh industrialized conditions i.e. high temperatures. Several commercially available thermostable enzymes have been harnessed from both mesophiles and thermophiles, but *in vitro* the latter provided more robust thermostable enzymes (Nielsen *et al*., 2007). Exploiting thermostable enzymes directly from "nature" has proven to be inefficient, owing to the intracellular factors that govern the proteins ability to remain thermally stable (Kumar *et al*., 2000). Therefore protein engineering of "cell free" thermostable enzymes would potentially be required for adaptation in a commercialized industrial application (Kumar *et al*., 2000). Enzyme thermostability is determined by its primary structure and can be predicted by its tertiary and/or quaternary structure (Ebrahimi *et al*., 2011). A study performed by Lakizadeh *et al*. (2011) proposed a modelling discrimination between mesophilic and thermophilic enzymes based on crystal structures, codon usage, amino acid composition, salt bridges, aromatic interactions, hydrophobicity and the content of arginine, histidine, proline and tryptophan. The modelling discrimination forms the basis for engineering thermophilic enzymes suitable for harsh industrial applications. For a thermogenic bioethanol process lignocellulase thermostability is of particular importance in addition to catalytic efficiency at high process temperatures (Nigam, 2013). In 2012, Novozymes (Denmark) and Genencor (USA) commercially released an engineered improved version of cellulase cocktails known as Cellic CTec 3 and Accellerase Trio, respectively. These enzymatic cocktails have improved thermostability and claim to reduce the loading of

these enzymes by two fold, but this has not decreased the overall production costs of commercializing lignocellulosic ethanol (Skovgaard and Jorgensen, 2013). Reducing overall process costs significantly will occur in aid of thermostable enzymes. These enzymes have higher stability, specific activity and the rate of reactions are increased by several orders which leads to the less amount of enzymes needed for an extended period of hydrolysis and saccharification (Viikari *et al.*, 2007). These enzymes are compatible to the non-enzymatic designed processes to oxidize the crystal network of cellulose (Szijarto *et al.*, 2008). The enzymes are inactive at low temperatures and can be stored at room temperatures (Viikari *et al.*, 2007). A study performed by Skovgaard and Jorgensen (2013), showed the effects of high temperature and ethanol exposure to all classes of lignocellulosic enzymes. A comparison between non-thermostable and thermostable lignocellulosic enzymes activity after being exposed to high heat and ethanol indicated that mesophilic-based enzyme activity was decreased significantly while low effect was observed with the thermophilic counterparts. This emphasized the overall cost reduction effect by having thermostable enzymes potentially recycled after distillation for repetitive hydrolysis.

1.6 Metagenomic gene discovery

Micro-organisms possess diverse enzymes that can be exploited to suit industrialized lignocellulosic saccharification processes i.e. thermostable enzymes. However tapping into the genetic potential of intricate and dynamic microbial communities for novel biocatalysts can be a daunting process. Using the screening methods under conventional culture-dependent conditions restricts the ability to harness novel enzymes owing to inability to culture complex microbial communities under standard laboratory conditions. It is well understood that a vast diversity of micro-organisms (more than 99%) in the environment have

not been biocharacterized or classified due to their inability to be cultured (Amann *et al.*, 1995). Metagenomics solves the uncultivable issue by theoretically being able to recover all DNA from an environmental sample and thereby allowing full access to the total microbial diversity. In this way potentially novel biocatalysts and metabolic pathways could be harnessed for industrial applications for generating new or improved biocommodities (Figure 1.11) (Lee *et al.*, 2010). To achieve this successfully the extraction of metagenomic DNA, encoding novel enzymes, needs to be tailored to ensure capturing the full taxonomic diversity. Microbial DNA extraction includes a wide range of physical and chemical mechanisms such as freeze-thawing, beadmill homogenization, ultrasonication and chemical hydrolysis (SDS/NaOH); however these methods have limitations as it could shear the genomic DNA required to generate large insert metagenomic libraries in cosmids, fosmids and BACs (bacterial artificial chromosomes) (Niemi *et al.*, 2001; Simon *et al.*, 2009). While smaller insert libraries can be constructed using pUC19 and/or pBluescriptSK^{+} plasmid vectors, screening requires a 10 fold increase in the number of clones (Ranjan, 2005). Although smaller insert libraries are complex and tedious during screening, it becomes useful for the detection of frail expressed foreign genes (Daniel, 2005). These libraries are best preferred or suited for high copy amounts of DNA that will be utilized for sequence novelty instead of gene functional discovery (Venter et al., 2004). In contrast, large insert libraries are best suited for function based screening and this was substantiated by a study conducted by Gabor *et al*. (2004) demonstrating that obtaining a "hit" increases exponentially with increased insert sizes. Subsequent to ligating the fragments into vectors, the constructs are transformed/transfected into a heterologous host such as *Escherichia coli* (*E.coli*) and this is the universal bacterial species for library construction because it has non-fastidious growth requirements and has a fast life cycle at 37^{o}C (Wiley *et al.*, 2008). It's a useful host because it

lacks the genes for recombination (*recA* and *recBC*) and restriction (*mcrA* and *mcrBC*) of foreign DNA and numerous genetic manipulative tools are available that complements the use of this particular host (Uchiyama and Miyazaki, 2009). Alternative hosts, *Pseudomonas aeruginosa* and *Pseudomonas putida*, are used in an event of bias towards particular codon usage (Warren *et al*., 2008; Villegas and Kropinski, 2008). Metagenomic investigations that have used various vector-host library constructions from diverse environmental samples are illustrated in table 1.4. Following library construction, screening of the metagenomic libraries can occur in two ways; Functional-based and Sequence-based screening (Figure 1.11).

Table 1.4: The gene discovery studies through functional metagenomics approaches from 2005-2013.

Environment	Target gene	Host	Vector	Insert size (kb)	Assay technique	Reference
Compost	Esterase	*E.coli*	Plasmid	3.2	Agar Plate	Lammle *et al*., 2007
Compost (with poly-lactic acid disks)	Esterase	*E.coli*	Plasmid	2.5	Agar Plate	Mayumi *et al*., 2008
Soil	Xylanase	*E.coli*	Plasmid	8.5	Agar Plate	Hu *et al*., 2008
Loam Soil	Oxygenase	*E.coli*	Plasmid	5.5	Agar Plate	van Hellemond *et al*., 2007
Mining shaft, compost soil, sediments	Protease	*E.coli*	Plasmid	4	Agar Plate	Waschkowitz *et al*., 2009
Gypsy moth midgut	Production of *N*-acylhomoserine lactons	*E.coli*	Plasmid	3.3	Reporter	Guan *et al*., 2007
Rhizosphere soil from acid mine drainage	Nickel resistence	*E.coli*	Plasmid	2.5	Growth	Mirete *et al*.,2007
Glacial ice	DNA polymerase I	*E.coli*	Plasmid	4	Growth	Simon *et al*., 2009
Ground water from crude-oil storage	Aromatic hydrocarbon catabolic operon	*E.coli*	Plasmid	7	Reporter	Uchiyama *et al*., 2005
Intertidal flat sediment	Lipase	*E.coli*	Fosmid	N/A	Agar Plate	Kim *et al*., 2009
Deep-sea sediment	Lipase	*E.coli*	Fosmid	32.3	Agar Plate	Jeon *et al*., 2009
coke plant wastewater	Extradiol dioxygenase	*E.coli*	Fosmid	33	Cell lysate	Suenaga *et al*., 2007
Soil (pasture)	Degradation of *N*-acylhomoserine lactons	*E.coli*	Fosmid	35	Reporter	Riaz *et al*., 2008
Coke plant wastewater	Bleomycin resistance gene	*E.coli*	Fosmid	33	Growth	Mori *et al*.,2008
Forest rhizosphere soils	Fungal antagonism	*E.coli*	Fosmid	35	Growth Inhibition	Chung *et al*., 2008
Human fecal from healthy person and patients of Crohn's disease	Epithelial cell growth modulation	*E.coli*	Fosmid	43	Cell lysate	Gloux *et al*., 2007
Glacial ice	DNA polymerase I	*E.coli*	Fosmid	36	Growth	Simon *et al*., 2009
Surface seawater	Esterase	*E.coli*	BAC	70	Agar Plate	Chu *et al*., 2008

Pig fecal	Tetracycline-resistance gene	*E.coli*	BAC	15	Growth	Kazimierczak *et al.*,2009
Rabbit cecum	Cellulase	*E.coli*	Cosmid	35.1	Agar Plate	Feng *et al.*, 2007
Soil (wetland and sandbars)	4-Hydroxyphenylpyruvate dioxygenase	*E.coli*	Cosmid	40	Liquid-base	Lee *et al.*,2008
Cow rumen	Mannanase/glucanase/xylanase	*E.coli*	Phagemid	3	Agar Plate	Palackal *et al.*, 2007
Soil	Degradation of *N*-acylhomoserine lactons	*E.coli*	Phagemid	4.3	Reporter	Schipper *et al.*, 2009
Oil-contaminated so	Naphthalene dioxygenase	*P.putida*	Cosmid	25	Growth	Ono *et al.*, 2007
Rumen from Chinese yaks	GH43 β-D-xylosidase	*E.coli*	Cosmid	30 - 50	Cell lysate	Zhou *et al.*, 2012
German grassland soil metagenomes	GH30 α-L-arabinofuranosidase Cellulase	*E.coli*	Fosmid	N/A	Agar Plate	Nacke *et al.*, 2012
Bovine rumen-derived metagenomic library	xylanase	*E.coli*	BAC	50	Agar Plate	Gong *et al.*, 2013
elephant faeces	Cellulase	*E.coli*	Fosmid	23.5 -37.5	Agar Plate	Ilmberger *et al.*, 2012
Termite (*Pseudacanthotermes militaris*)	Hemicullases	*E.coli*	Fosmid	30 -74	Liquid-base	Bastien *et al.*, 2013
Compost	Esterase	*E.coli*	Fosmid	N/A	Agar Plate	Kang *et al.*, 2011

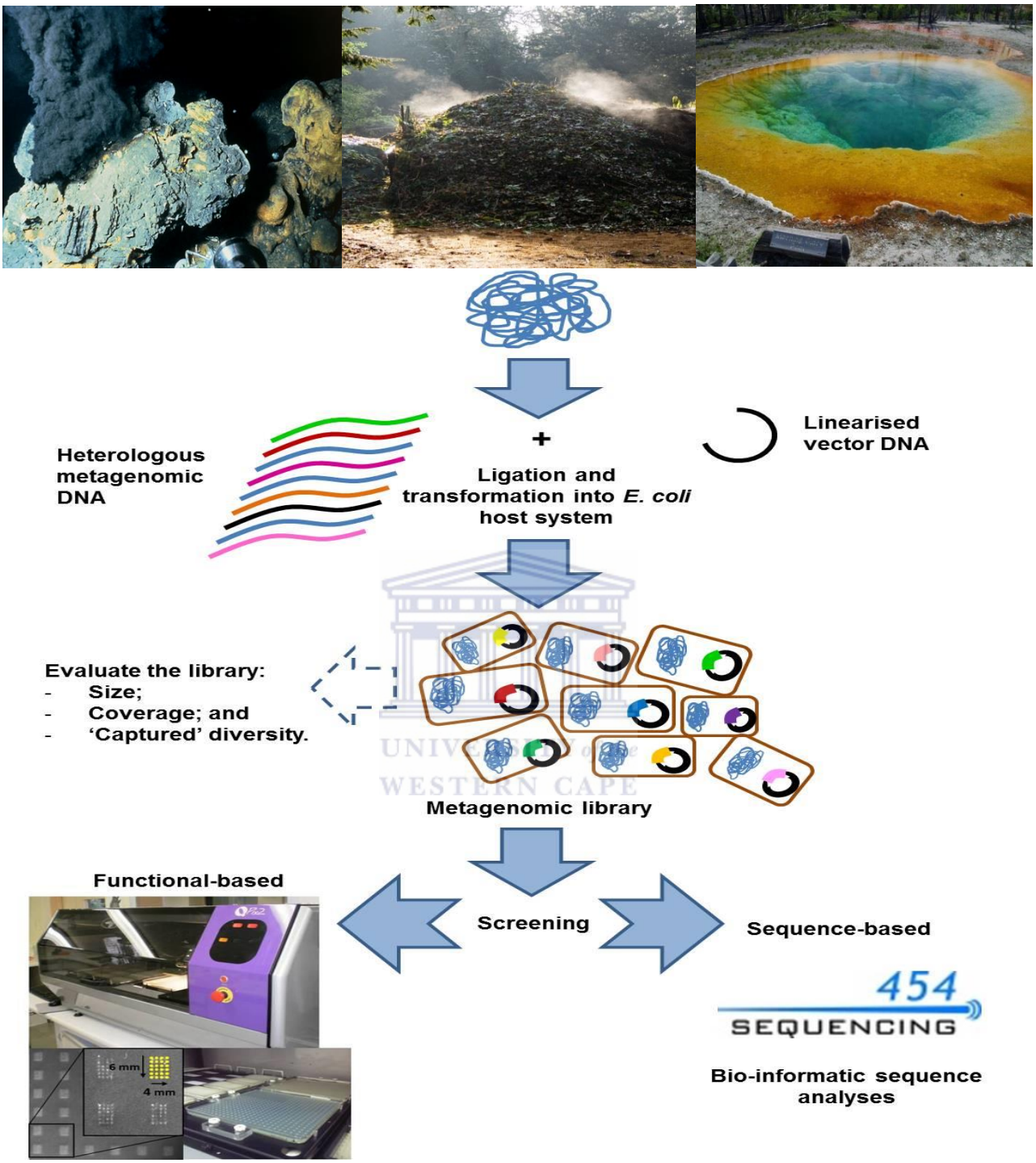


Figure 1.11 The diagrammatic overview of the metagenomic approach to isolating novel genes and gene products through extraction of DNA from an environmental source, construction of a metagenomic library followed by either functional base and/or sequence base screening of potential industrial biocatalysts (Handelsman, 2004).

1.6.1 Sequence-base screening

Sequence-based screening can allow for the detection of novel genes without the use of a gene expression system. Degenerate primers from the bioinformatically known conserved sequence is employed and followed by primer/gene walking to obtain the remainder of the gene (Morimoto and Fujii, 2009; Demaneche *et al*., 2009). Alternatively, primers can be designed for genetic elements flanking the open reading frame (ORF) provided that the genetic element is constant for the novel gene as well (Koenig *et al.*, 2009). Conserved sequences flanking the novel gene of the phylogenetically related gene are used to design primers that will amplify the flanking region and part of the gene which subsequently includes primer walking for the full sequence of the gene (Quaiser *et al.*, 2003). Sequence based screening can follow functional-based screening to obtain the nucleotide sequence of the novel gene as well as important flanking genetic elements such as the gene expression regulatory sequences. Transposon mutagenesis relies on the principle of random insertion of an antibiotic resistant gene into the novel gene and thereby resulting in a knock-out mutant. When conducted on a large insert clone expressing the activity of interest, sequencing out from the transposon enables one to determine the gene sequence conferring the activity of interest (Langridge *et al.*, 2009). The sequences of the different knock-out clones are aligned and the overlapping sequence forms the full nucleotide sequence of the novel gene. In this way sequence based screening is used to link gene product function to sequence (Steel *et al*., 2009; Simon and Daniel, 2009).

1.6.2 Functional-based screening

Functional-based screening involves detection of novel genes based on protein expression and functional activity on a substrate or inferred resistance to biochemical compounds that would inhibit proliferation of the host micro-organism (Riesenfeld *et al.*, 2004; Voget *et al.*, 2003). Gene-product-specific substrates can be supplemented in the solid growth medium and target clones can be detected by indication of a zone of hydrolysis, these include the use of tributyrin, carboxymethyl cellulose and RBB-xylan for the detection of lipase/esterase, cellulase and xylanolytic activity, respectively (Teather and Wood, 1982; Kouker and Jaeger, 1987; Hu *et al*., 2008; Mayumi and Akutsu-Shigeno, 2008). Chromogenic substrates can also be used. The *p*-nitrophenol based substrates are popular, where specific substrates can be conjugated to *p*-nitrophenol, and emission of the chromogenic *p*-nitrophenol indicates positive enzyme activity (Suenaga *et al*, 2007). Although the function-based metagenomic approach makes novel gene discovery easier than culture dependent or/and sequence base screening processes, it is not without limitations (Gabor *et al.*, 2004). These arise from issues relating to heterologous gene expression, due to the host cell's inadequate complementary translational and transcriptional factors to the metagenomic insert (Uchiyami and Miyazaki, 2009). These factors include bias to the start codon favouring the alternative codons besides AUG, decreased mRNA levels because of transcribing co-factors and mRNA becoming readily degraded, mRNA folding at the ribosomal binding sites and rare codon usage (Villegas and Kropinski, 2008; Kudla *et al.*, 2009; Wang *et al.*, 2006). In order to overcome these, library screening can be conducted in an alternative host (Wang *et al.*, 2006).

1.8 Concluding Remarks

Emerging potential biotechnological products such as thermostable enzymes are increasingly being recognized for inclusion in a thermophilic bioethanol process, using lignocellulose for complex liberation of fermentable sugars; as well as other emerging bio-industrial processes. Thermal environments provide a diverse source of thermophilic microbial communities containing genomes expressing potential biocatalyst which could be targeted for hydrolyzing the complex network of lignocellulose in an enzymatic sacchrification process. To enable the harnessing of glycoside hydrolases, and specifically AFases in this study, will require functional and sequence base metagenomic approaches to isolate and characterize thermostable biocatalysts. AFases possess the hydrolytic activity needed for synergistic liberation (together with other hemicellulosic and cellulosic enzymes) of fermentable sugars owing to the complex crystalline lignocellulose structure. Although both mesophilic and thermophilic AFases have been characterized, using the principles of metagenomics will extrapolate novel AFases containing hydrolytic functions with potential industrial applications other than their usage in the biofuel industry.

General introduction aims and objectives of the project

Thermal biofuel production has higher cost effective and process efficiencies in comparison to its mesophilic counterpart (Viikari *et al.*, 2007). Thermostable enzymes are compatible to the overall efficiency of the process for the complete sacchrification of crystalline hemicellulose and cellulose (Szijarto *et al.*, 2008). α-L-Arabinofuranosidases (AFases) are hemicellulosic accessory enzymes that have recently generated interest as a potential industrial application within liberating fermentable monomers in concert with other

lignocellulases for bioethanol production (dos Santos *et al.*, 2011). Non-thermostable AFases have been identified, through its mode of action, to have potential application in food, beverage and the pharmaceutical industries (Numan and Bhosle, 2006).

Composting is a process by which organic waste are disposed for transformation to manure (Sundberg *et al.*, 2004). Throughout the physico-chemical dynamic process, a succession of microbial communities occurs (Goyal *et al.*, 2004). The temperature of the process rises to 50^{o}C through the catabolic metabolism of the mesophilic microbial communities followed by a successive thermophilic microbial community (Tang *et al.*, 2004). It has also been demonstrated that thermotolerant micro-organisms are present as the compost mixture matures (Tiago *et al.*, 2004). Metagenomics permits novel genes to be isolated, sequenced and functionally characterized (Gabor *et al.*, 2004). Applying the principles of metagenomics have allowed access to AFase genes that encoded proteins with potentially differing thermostability states, that can be biochemically characterized for assessment as a potential industrial catalyst.

The present investigation formed an integral part of a broader study, in which researchers based at the Institute for Microbial Biotechnology and Metagenomics (IMBM; UWC), together with collaborators at the University of Stellenbosch, TMO (UK) and CSIR Biosciences were employing metagenomic technologies to identify novel lignocellulosic enzymes. This study was initiated by the construction of a fosmid metagenomic library from a compost sample (Ohlhoff, 2012) and through a high-throughput multiplex liquid screening (Smart, pers. comm), 12 AFase fosmid clones were identified. Selection of a thermolabile, moderate and highly thermostable AFase clones were obtained through the preliminary thermostability assay with the cell free crude protein extracts. These three AFase fosmid

clones (pFos_H4, pFos_E3 and pFos_D3) were sequenced and subjected to a functionally characterized study through the following objectives:

- Perform preliminary AFase thermostability testing on crude protein extracts from the recombinant fosmid clones.
- Mutagenized the fosmid clones through a kanamycin transposon mutagenesis and sequence the functionally inactivated AFase gene with the transposon specific primers.
- Bioinformatic analysis on the genes and flanking nucleotide sequences.
- Bioinformatic analysis on the fosmid-end nucleotide sequences.
- Subcloning into a pET21a expression vector.
- Purification of the AFase proteins by nickel-ion affinity liquid chromatography.
- Biochemical characterization on the purified AFase.
- Thermostability characterization on the purified AFases.
- Generate a putative sequence-structure-function relationship relative to the thermostability profiles

Chapter 2:

Materials and Methods

2.1 Culturing

2.1.1 Reagents and Media

All reagents used were analytical grade and obtained from Sigma-Aldrich Chemical Company, Kimix Chemical and Laboratory supplies or Merck Chemicals and Laboratory Supplies. Media components were acquired from either Oxoid Ltd or Biolabs. All media were autoclaved at 121ºC for 20 minutes prior to use, unless otherwise specified. Distilled water used for making solutions, media and diluting buffers was purified using a Lasec purite 15.2 MΩ (Millipore) water purification system. Restriction enzymes and the protein molecular weight marker used in this study were obtained from Fermentas (Fermentas, Germany).

2.1.2 Bacterial strains

E. coli strains (Table 2.1) were cultured in Luria-Bertani broth [LB; 1% (w/v) Tryptone, 0.5% (w/v) Yeast extract and 1% (w/v) NaCl] and on Luria-Bertani agar [LA; LB supplemented with 1.5 % (w/v) bacteriological agar]. Standard *E. coli* cultures were prepared by aseptically streaking a frozen glycerol stock onto LA. Briefly, glycerol stocks were prepared by the addition of 16% (v/v) sterile glycerol to an overnight bacterial culture in a sterile cryovial (SPL Life Sciences Co. Ltd), before being rapidly frozen and stored at -80^{o}C.

Table 2.1 *E. coli* strains used in this study.

Strains	Genotype	Reference
Epi300	F- mcrA D(mrr-hsdRMS-mcrBC) f80dlacZDM15 DlacX74recA1 endA1 araD139 (ara, leu) 7697 galK1-rspLnupGtrfAtonAdhfr	Invitrogen, USA
Genehog	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galUgalKrpsL (StrR) endA1 nupGfhuAIS2	Epicentre Biotechnology, USA
BL21 (DE3)	F- ompThsdSB(rB-mB-)gal dcm gal λ(DE3)	Invitrogen, USA

2.1.3 Bacterial plasmids/fosmids

E. coli clones harbouring pCC1Fos™ (Table 2.2) and recombinant pCC1Fos™ constructs (Table 2.2) were routinely cultured on LA, or in LB, supplemented with 12.5 µg mL^{-1} Chloramphenicol (Cam). However, recombinant pCC1Fos™ constructs (Table 2.2) that were mutagenized with the Kanamycin (Kan) transposon were cultured onto LA and LB containing 50 µg mL^{-1} Kan and 15 µg mL^{-1} Cam. Similarly, *E. coli* clones harbouring recombinant pJET1.2/blunt constructs, pET21a and recombinant pET21a constructs (Table 2.2) were routinely cultured on LA, or in LB, containing 100 µg ml^{-1} ampicillin (Amp).

Table 2.2 Plasmids used and generated in this study.

Plasmid	Genotype[a]	Reference
pCC1Fos™	L-Arabinose inducible promoter Copy Control; Cam^R, F-factor ori, *ori*V high copy ori, λcos site for λ packaging, Bacteriophage T7 RNA polymerase promoter	Epicentre, USA
pFos_H4	pCC1FOS containing 17.5 kb of cloned metagenomic DNA as an insert with AFase activity, Cam^R	IMBM
pFos_E3	pCC1FOS containing 20.7 kb of cloned metagenomic DNA as an insert with AFase activity, Cam^R	Dr C. Ohlhoff, IMBM, UWC, SA
pFos_D3	pCC1FOS containing 10. 7 kb of cloned metagenomic DNA as an insert with AFase activity, Cam^R	Dr C. Ohlhoff, IMBM, UWC, SA
pJET 1.2/blunt	Suicide cloning vector (eco47IR), blunt DNA ends for ligation with insert, T7 promoter, Amp^R	Fermentas, USA
pJET_H4	1467 bp AFase_H4 gene amplicon blunt-end ligated into pJet1.2	This study
pJET_E3	1547 bp AFase_E3 gene amplicon blunt-end ligated into pJet1.2	This study
pJET_D3	1482 bp AFase_D3 gene amplicon blunt-end ligated into pJet1.2	This study
pET21a(+)	Expression vector with a C- terminal His-tag, Amp^R, T7 promoter and terminator, MCS.	Novagen, USA
pET21a_H4	1467 bp *Nde*I-*Xho*I fragment from pJET_H4 cloned in pET21a.	This study
pET21a_E3	1547 bp *Nde*I- *Hin*dIII fragment from pJET_H4 cloned in pET21a.	This study
pET21a_D3	1482 bp *Nde*I-*Xho*I fragment from pJET_H4 cloned in pET21a.	This study

[a] Plasmid encoded Ampicillin and Chloramphenicol resistance is indicated as Amp^R and Cam^R, respectively. Ori is used as an abbreviation for origin of replication and MCS is the acronym for multiple cloning site. The Hexa-histidine tag is abbreviated to His-tag that was used to purify the AFase proteins of this study through nickel-ion affinity liquid chromatography

2.2 DNA manipulations

2.2.1 Competent cells and transformation

The following *E. coli* strains were prepared for electroporation as described by Seidman *et al*, (1997), but with modifications, detailed as follows. *E. coli* Epi300 and *E. coli* GeneHogs (Table 2.1) were aseptically streaked, from a frozen glycerol stock, onto a LA plate and incubated at 37^{o}C overnight. A single colony was subsequently inoculated into 5 mL LB and incubated at 37^{o}C with shaking (160 rpm). The overnight culture was diluted 1/200 into 200 mL of super optimal broth [SOB; 2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 0.05% (w/v) NaCl and 25 mM KCl] and incubated at 37°C until the optical density at 600 nm (OD_{600}) reached 0.35 (Thermo Scientific™ BioMate 3S UV-Visible spectrophotometer). The cells were harvested by centrifugation (4,000x *g* for 10 min at 4^{o}C), before being resuspended in one culture volume of ice-cold sterile distilled water (dH_2O). The cells were harvested, as before, and the cell pellet resuspended in an ice-cold sterile solution of 15% (v/v) glycerol and 2% (w/v) sorbitol, in a proportion of 2 mL resuspension solution per L of culture volume. The electrocompetent cells were aliquoted into 40 µL per eppendorf tube and stored at -80°C. A standard electroporation protocol (BioRad Gene Pulser) was used to transform electrocompetent *E. coli* cells with plasmid DNA. Briefly, a 40 µL aliquot of electrocompetent *E. coli* Epi300 cells was thawed on ice. To this 5-10 ng of plasmid DNA, in a maximum volume of 1 µL, was added and mixed, before being subjected to electroporation at 200 Ω, 25.0 uF and 1.8 kV (Bio-Rad Micropulser™) in an electroporation cuvette

[Micropulser^{TM} electroporation 0.1cm gap cuvettes (Bio-Rad)]. Cells were immediately recovered with the addition of 1 mL of Super Optimal Broth (SOB) and incubated of 37^{o}C with shaking (160 rpm) for 60 minutes, before being spread plated onto LA supplemented with the appropriate antibiotic(s).

Chemically competent *E. coli* BL21 cells were prepared, by $CaCl_2$ treatment, according to Dagert and Ehrlich (1979). Briefly, *E. coli* BL21 (Table 2.1) was streaked from a frozen glycerol stock onto LA agar and incubated at 37°C O/N. A single colony was inoculated into 10 mL of LB and incubated at 37°C with shaking (160 rpm). The overnight culture was diluted 1/20 in 50 mL Ψ-broth [2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 0.4% (w/v) $MgSO_4.7H_2O$ and 0.075% (w/v) KCl] and incubated at 37^{o}C with shaking (160 rpm) until the optical density at 600 nm (OD_{600}) reached 0.35-0.6. The cells were harvested from approximately 35 mL of the early log phase culture, by centrifugation (5,000x *g* for 5 min at 4^{o}C), and resuspended on ice in an equal volume of pre-chilled 0.1 M $MgCl_2$. Thereafter, the cells were harvested, as before and re-suspended in half the culture volume of pre-chilled 0.1 M $CaCl_2$ and incubated on ice for 2 hours. Cells were pelleted, as before, and resuspended in 1/10 culture volume of an ice-cold solution of 0. 1M $CaCl_2$ and 16% (v/v) glycerol. The competent cells were aliquoted into 100 µL per eppendorf tube and stored at -80°C.

A heat-shock transformation method was used to transform *E. coli* BL21 cells (Table 2.1) as described in by Dagert and Ehrlich (1979). Briefly, a frozen 50 µL aliquot of $CaCl_2$ chemical competent cells was thawed on ice, before approximately 10 ng of plasmid DNA, in a total volume of 10 µL, was added and incubated on ice for 20 minutes. The cells were heat shocked at 37^{o}C for 1 minute, followed by incubation on ice for 2 min. Thereafter, 800 µL of

Ψ-broth was added and incubated at 37°C for 30 min with shaking (250 rpm), before being spread plated onto LA supplemented with the appropriate antibiotic(s).

2.2.2 DNA isolation

2.2.2.1 Alkaline lysis plasmid DNA extraction protocol

Unless otherwise stated, the extraction of plasmid DNA was performed as described by Ish-Horowicz and Burke, 1981 from *E. coli* liquid cultures. Briefly, *E. coli* harbouring the plasmid of interest was inoculated from a freshly grown colony into 5 mL of LB containing the required antibiotic. One mL of the 5 mL O/N culture was inoculated into 9 mL of LB. The culture was incubated at 37°C for 5 hours before being harvested by centrifugation (13,000x *g* for 5min at 4°C). The cell pellet was resuspended in 100 µL of Solution 1 [50 mM Glucose, 25 mM Tris-HCl and 10 mM EDTA (pH 8.0)] and 10 µL of 100 ng/µL of DNAse-free RNase (Fermentas) followed by an incubation on ice for 15 min. Thereafter, approximately 200 µL of solution 2 [200 mM NaOH, 1% (w/v) SDS] was added and the contents mixed by gently inverting the microfuge tube several times, before being incubated at room temperature for 5 min. The cellular debris was precipitated by the addition of 150 µL of solution 3 [3 M Potassium Acetate (pH 5.5)] and 20 min incubation on ice, followed by centrifugation (13,000x *g* for 10 min at 4°C). The supernatant was removed to a new microfuge tube and the plasmid DNA precipitated by the addition of 1 mL absolute ethanol. The plasmid DNA was collected by centrifugation (13,000x *g* for 15 min at 4°C), washed with 70% (v/v) ethanol and air-dried. The extracted plasmid DNA was resuspended in approximately 20 µL of Tris-EDTA (TE) buffer [25 mM Tris-Cl, 10 mM EDTA, pH 8.0]. The concentration of the extracted plasmid DNA was determined spectrophotometrically (Nanodrop, ND-1000) and stored at 4°C.

2.2.2.2 Extraction of the Recombinant Fosmids

Fosmid clones were cultured from their respective glycerol stocks as described. Single colonies were aseptically inoculated into 10 mL LB containing the respective antibiotic (15 µg mL^{-1} Cam or 50 ug/ml Kan), and incubated O/N at 37^{o}C with shaking (160 rpm). The 10 ml culture was inoculated into 100 mL LB containing 0.01% (w/v) L-arabinose and antibiotic, and incubated O/N at 37^{o}C with shaking (160 rpm). Thereafter, the cells were harvested by centrifugation (6,000x *g* for 15 minutes at 4^{o}C) and the fosmids were isolated using the Qiagen Plasmid Midi-Kit, as per manufacturer's instructions. The DNA concentrations and purity of the extracted fosmids were determined spectrophotometrically using the Nanodrop (ND-2000) and stored at 4^{o}C.

2.2.3 Ligation

The three restricted and purified AFase gene insert fragments and two linearised pET21a(+) vectors (Figure 2.1), respectively, were ligated at a 1:4 (insert to vector) picomole ends ratio using T4 ligase (Fermentas), according to the manufacturer's instructions, at room temperature for 10 min. Approximately 5 µL of each ligation reaction were transformed by electroporation into *E. coli* GeneHog® cells, as described in section 2.3.1.

2.2.4 Restriction enzyme digestion

Fosmid or plasmid DNA was subjected to restriction enzyme analysis to determine the approximate size of the insert DNA. Approximately 1 µg was restricted with the respective restriction enzymes (Fermentas), according to the manufacturer's instructions. The resulting restriction fragments were separated, alongside a molecular weight marker (λ digested with

*Pst*I), on a 0.8% (w/v) TAE agarose gel and visualised under UV light using an Alpha-imager HP (Alpha-innotech©). The λ-*Pst*I molecular weight marker was prepared by the restriction endonuclease digestion of 72 µg of λ phage DNA (Fermentas) with *Pst*I (Fermentas), according to the manufacturer's instructions. The sizes of the DNA fragments following restriction digestion were determined relative to the mobility of the λ-*Pst*I molecular weight marker using DNAFRAG version 3.03, which is based on a least squares fit of fragment size to gel mobility (Schaffer and Sederoff, 1981).

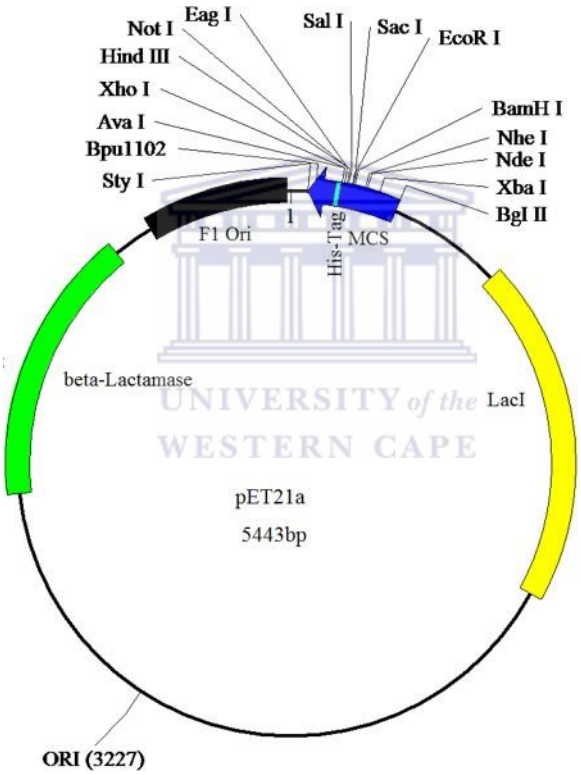


Figure 2.1 A plasmid map of the pET21a(+) protein expression vector (Novagen). The multiple cloning site (MCS) (▪) contains the phosphodiester cleavage sites for the indicated restriction endonucleases with the histidine tag (▪). An F1 origin (▪) of replication is present for production of a single-stranded vector under particular cellular conditions while the second origin of replication (ORI) is for the conventional replication of the vector. The Lac repressor gene (▪) is present with the ampicillin resistant gene, beta-Lactamase (▪).

2.2.5 Agarose electrophoresis

The integrity of the extracted plasmid DNA was determined by electrophoresis on a 0.8% (w/v) TAE agarose gel. Owing to the fosmids immediate usage in the subsequent experiments, the integrity of the fosmids was viewed on a 0.7% (w/v) TAE agarose gel. The resulting DNA fragments of the various gene and plasmid RE digests of this study were subsequently separated on a 0.8% (w/v) TAE agarose gel. Constructs observed to contain the correct sized insert were further analysed to ensure that the putative AFase gene(s) had been cloned in-frame with the C-terminal vector encoded histidine tag. Resulting PCR products were separated by electrophoresis on a 1.2% (w/v) TAE agarose gel. All gels were analysed with a λ*Pst*I molecular weight marker and visualised under UV light on an Alpha Imager system.

2.2.6 Transposon mutagenesis

Transposon mutagenesis was employed to identify the genes encoding functional AFase(s). Briefly, the extracted fosmid DNA of pFos_H4, pFos_E3 and pFos_D3 (section 2.2.2.2) was subjected to transposon mutagenesis using the HyperMu™ <KAN-1> Insertion Kit (EpiCentre Biotechnologies), according to the manufacturer's instructions (Summarised in Figure 2.2). Thereafter, the transposon-treated fosmids were transformed, by electroporation, into electrocompetent *E. coli* Epi300 cells (section 2.2.1) and cultured as described in section 2.1.2. The double antibiotic supplementation of Cam and Kan, within the growth media would select for those *E. coli* cells harbouring a recombinant fosmid that had under gone a transposon integration event. In addition, as a control, electrocompetent *E. coli* Epi300 cells were also transformed with untreated fosmids and spread plated onto LA supplemented with

15 µg mL^{-1} Cam and incubated at 37°C O/N. *E. coli* Epi300 fosmid clones capable of growth on both Cam and Kan were sub-cultured onto LA containing 15 µg mL^{-1} Cam and 50 µg mL^{-1} Kan and incubated O/N at 37°C, before being stored at 4°C as a back-up. The individual clones were subsequently inoculated into individual wells of sterile 96-well microtitre plates (Sterilin®) containing LB with 15 µg mL^{-1} Cam and 50 µg mL^{-1} Kan, and incubated O/N at 37°C with shaking (250 rpm). The master microtitre plates were manually duplicated, using a 96-pin colony picker, into a second set of sterile 96-well microtitre plates containing LB containing 15 µg mL^{-1} Cam, 50 µg mL^{-1} Kan and 0.01% (w/v) L-arabinose, and incubated at 37°C O/N with shaking (250 rpm). Sterile glycerol was added to each well of the master plates to achieve a final concentration of 15% (v/v), before they were covered with sterile foil covers and stored at -80°C.

The duplicate plates, for each of the AFase-positive fosmids, were subsequently screened for transposon-treated fosmid clones that had been functionally inactivated by the insertion of the transposable element into the gene(s) of interest. Following the O/N incubation step, the cells within the duplicate microtitre plates were lysed, and the soluble cell-free extracts released, by the addition of BugBuster™ protein extraction reagent (Novagen), according to the manufacturer's instructions. AFase enzyme activity was detected by addition of 1 mM *p*NP-Ara in 50 mM $NaPO_4$ buffer (pH 7) to each well of the MTPs containing the cell-free extracts and incubated at 37°C. The enzyme activity assays were incubated until a yellow colour, due to the release of *p*NP from the synthetic substrate, was visually observed for the three positive untreated fosmid controls, pFos_H4, pFos_E3 and pFos_D3, that had been treated as described above. Transposon-treated fosmid clones that did not develop the distinct yellow colour or were observed to produce significantly reduced levels of yellow colour relative to that of the untreated fosmid controls, were identified and chosen for further analysis. Loss of

activity or reduced activity was presumed to be as a result of the transposable-element having integrated into the coding region and/or regulatory element of the putative AFase gene(s). Those clones, that had lost the active AFase phenotype, were cross-referenced with the master microtitre plate(s) and cultured as previously described, before the functional screening was repeated to confirm the loss or reduction of activity. Following the secondary screening using the *p*NP-ara assay, the absorbance at 410 nm (OD_{410}; SPECTROstar Nano; BMG Labtech) was also determined for positive control and transposon-inactivated fosmids.

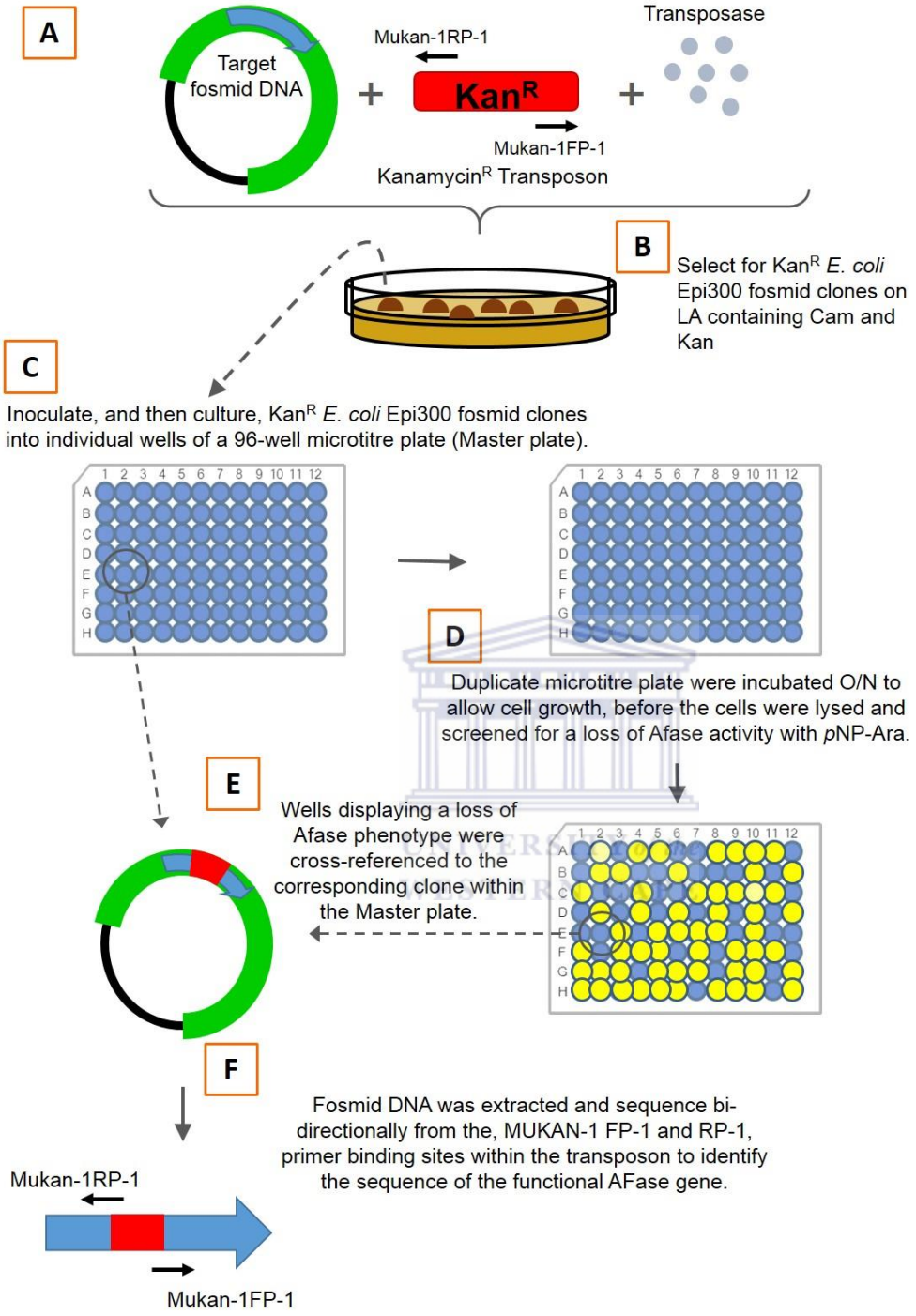


Figure 2.2 A diagrammatic representation of the approach used to identify the genes encoding α-L-arabinofuranosidase (AFase) activity. Transposon mutagenesis was performed using the HyperMu™ <Kan-1> Insertion Kit (A) obtained from EpiCentre Biotechnologies and subsequent mutagenized clones were identified using a functional screen similar to that outlined in Figure 2.2. Fosmid clones which were successfully mutagenized were identified by growth on double antibiotics,

chloramphenicol and kanamycin (B), before being inoculated and cultured in a master microtitre plate (C). Successfully mutagenized clones were identified from their lack of AFase activity and therefore no release of *p*-nitrophenol. Clones displaying reduced AFase activity was also selected as this may indicate the incorporation of the transposon into the regulatory region of the gene.

2.2.7 PCR (Polymerase Chain Reaction)

Oligonucleotide primers (Table 2.3.) were designed for the gene amplification and sub-cloning of AFase genes, namely AFase_H4, AFase_E3 and AFase_D3, into the pET21a(+) (Table 2.2 and Figure 2.1.) protein expression vector. Genes were amplified from 10 ng fosmid DNA (section 2.2.2.2) in a 25 µL polymerase chain reaction (PCR) containing 0.25 U of Phusion polymerase (Thermoscientific^{TM}), 0.25 mM dNTPs (Thermoscientific^{TM}) and 10 µM of each primer (Table 2.3, this study). The following cycling conditions were employed for the amplification of AFase_H4 and AFase_D3: initial denaturation at 98^{o}C for 30 seconds; cycling 35 times denaturation at 98^{o}C for 10 seconds, annealing at 70^{o}C for 30 seconds, elongation at 72^{o}C for 45 seconds and a final elongation at 72^{o}C for 10 minutes. For the amplification of the AFase_E3 gene a similar protocol was followed except that the annealing and elongation were performed at 72°C for 15 seconds at 72°C for 30 seconds, respectively.

2.2.8 Sequencing

Plasmid and fosmid sequencing was conducted by the University of Stellenbosch's Central Analytical Facility (CAF). The sequences were manually edited using Chromas version 2.01 (James, 2001) and DNAMAN version 4.13 (Lynnon Biosoft, 1999). Sequence identity and

similarity searches of DNA sequences were performed using the basic local alignment search tool (BLAST) programs (Altschul *et al.*, 1989; Altschul *et al.*, 1997; Marchler-Bauer *et al.*, 2005), as provided by the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/). Putative ORFs were identified within the consensus sequences using GeneMark for Prokaryotes (http://opal.biology.gatech.edu/genemark_prok_gms_plus.cgi), and Interproscan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) was used to identify conserved protein domains.

Table 2.3 Primers employed throughout this study.

Primer ID	5' - 3' sequence*	Reference
T7 Promoter forward	TAATACGACTCACTATAGGG	Epicentre Biotechnologies
pCC1Fos reverse	CTCGTATGTTGTGTGGAATTGTGAGC	Epicentre Biotechnology
MUKAN-1 FP-1	CTGGTCCACCTACAACAAAGG	EpiCentre Biotechnologies
MUKAN-1 RP-1	AGAGATTTTGAGACAGGATCCG	EpiCentre Biotechnologies
M13 Forward	CCCAGTCACGACGTTGTAAAACG	Inqaba biotec
M13 Reverse	AGCGGATAACAATTTCACACAGG	Inqaba biotec
H4_pET_Fwd	GTT**CATATG**AATCACATCAAGATTGATTTAGATCGTC	This Study
H4_pET_Rev	CG**CTCGAG**TAAGTCAAAGCTGAGC	This Study
E3_pET_Fwd	AT**CATATG**GACGGAGGCGCATGCG	This Study
E3_pET_Rev	C**AAGCTT**GGACGGTCGGCGG	This Study
D3_pET_Fwd	GAT**CATATG**AACAATGTCGTCATCAATGTGG	This Study
D3_pET_Rev	CTT**CTCGAG**ACCTAATCTTAGAATGCCGAC	This Study

* The restriction endonuclease sites incorporated into the PCR primers are indicated in coloured bold type. Red indicates *Xho*I, green indicates *Nde*I and blue indicates *Hin*dIII.

2.3 Protein expression and purification

2.3.1 Preparation of cell free extracts

E. coli BL21 (Table 2.1) transformants harbouring pET21a_H4, pET21a_E3 and pET21a_D3 (Table 2.2), respectively, were investigated in a comparative small-scale expression trial to determine the optimal induction conditions for expression of the recombinant AFase proteins. *E. coli* BL21 harbouring pET21a_H4, pET21a_E3 and pET21a_D3, respectively, were cultured from glycerol stocks as described in section 2.1.2 and incubated at 37^{o}C O/N. Single colonies were individually inoculated into 10 mL of LB with the same concentration of Amp as described in section 2.1.2 and incubated at 37°C with shaking (160 rpm) O/N. Approximately 2 mL of the liquid starter culture was inoculated into triplicate 500 mL Erlenmeyer flasks containing 50 mL of LB containing 100 µg mL^{-1} Amp and incubated at 37^{o}C with shaking (160 rpm) until they reached an OD_{600} of approximately 0.6-0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG; Fermentas) was added to two of the 50 mL liquid cultures to achieve final concentrations of 0.5 and 1 mM, respectively, while the third flask was not induced. Cultures were incubated for an additional 3 hours. One hundered microliters of each culture was aseptically removed to a fresh microfuge tube and the cells lysed using BugBuster protein extraction reagent (Novagen), according to the manufacturer's instructions. The cell debris was removed by centrifugation (13,000x *g* for 10 min at 4°C) and the cell-free extract removed to a fresh tube and combined with an equal volume of 4mM *p*NP-α-L-arabinofuranoside (*p*NP-Ara) was added to each tube and incubated at 37°C for 30 min. As a measure of the efficiency of heterologous expression, the OD_{410} was measured at 410nm using a spectrophotometer to determine the amount of liberated *p*-Nitrophenol.

2.3.2 Bradford's assay

The protein concentration of the recombinant constructs was determined using a modified Bradford assay (Bradford, 1976). Briefly, the assay was performed in triplicate by mixing 5 µl of the protein sample and 200 µl of undiluted Bradford's reagent (Sigma) in a well of a flat-bottom 96-well microtitre plate. The samples were gently mixed, by pipetting up and down several times, and the reaction incubated at room temperature for 30 min, before the absorbance was determined at 595 nm (OD_{595}) using a microplate reader (Spectrostar nano, BMG Labtech, Germany). Negative controls, or blanks, were prepared as above using 5 µL of the respective dialysis buffer. A standard curve was generated using a set concentration range (0.2 - 1.6 mg mL^{-1}) of Bovine Serine Albumin (BSA; Sigma) prepared in buffer, according the manufacturer's instructions. The concentration of the BSA standards was determined at 595 nm (OD_{595}) using the microplate reader. The protein concentration of the unknown samples was determined by comparison to the standard curve of the BSA standards.

2.3.3 SDS-PAGE

The cell-free extracts, prepared as described above, were analysed using SDS polyacrylamide gel electrophoresis (PAGE). Briefly, the soluble and insoluble proteins were analysed by SDS-PAGE according to Laemmli (1970). SDS-PAGE gels were cast using the Bio-Rad hand casting system (Bio-Rad), with a 10% (w/v) polyacrylamide separating gel [0.375 M Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 9.9% (w/v) acrylamide (Sigma), 0.264% (w/v) bis-acrylamide (Sigma), 0.05% (w/v) ammonium persulphate (APS) and 0.1% (v/v) tetramethylethylenediamine (TEMED; Sigma)] and 4% (w/v) polyacrylamide stacking gel [0.125 M Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 4% (w/v) acrylamide, 0.1% (w/v)

bisacrylamide, 0.03% (w/v) APS and 0.1% (v/v) TEMED]. Sample preparation consisted of mixing equal volumes of 2x loading dye [80 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 2% (v/v) SDS, 1% (v/v) β-mercaptoethanol and 0.02% (w/v) bromophenol blue] and the protein sample, and heating to 95°C for 5 min prior to loading onto the polymerized polyacrylamide gel. The protein samples were separated under denaturing conditions at a constant voltage of 100 volts until the dye front had migrated off the polyacrylamide gel. Following electrophoresis, the polyacrylamide gel was visualised by staining with Coomassie Brilliant Blue solution [40% (v/v) methanol, 10% (v/v) glacial acetic acid and 0.025% (w/v) Coomassie Brilliant Blue R-250] for 2 hrs and then destained with destaining solution [20% (v/v) methanol and 10% (v/v) glacial acetic acid] for 16 hrs (Sambrook and Russell, 2001). The molecular weight was determined by comparison to the PageRuler (Fermentas, Germany) molecular weight ladder separated on the same polyacrylamide gel.

2.3.4 Nickel-ion chromatography

Following IPTG induction of overexpression, as described in section 2.3.1, the cultures were transferred to sterile 500 mL centrifuge bottles (NalGene) and the cells harvested by centrifugation (4,000x g for 20 min at 4^{o}C). The cell pellets were resuspended in 10 mL of 1x binding buffer (Novagen; 250 mM NaCl, 20 mM Tris-HCl, 5 mM imidazole, pH 7.9) on ice. The cells were lysed by sonication (Bandelin Sono plus Ultrasonic Homogenizer, Germany) of 6 pulses for 20 seconds at 60% power setting followed by 30 seconds of no sonication. Throughout the sonication procedure the cell suspension was maintained on ice to prevent heat build-up within the sample. Following sonication the insoluble cell debris was removed by centrifugation (6,000x *g* for 20 min at 4°C). The supernatant, containing the crude cell-free protein extract was maintained on ice. The recombinantly expressed His-tag fused

proteins AFase_H4, AFase_E3 and AFase_D3 were purified by Ni-chelation chromatography with the His-Bind® Resin and Buffer kit (Novagen, USA), according to the manufacturer's instructions. Briefly, 10 mL disposable chromatography columns (Bio-Rad) were prepared according to the manufacturer's instructions with a 1.5 mL resin bed volume, before being equilibrated with 3 column volumes of distilled water, 5 column volumes of 1x charge buffer (Novagen; 250 mM $NiSO_4$) and 3 column volumes of 1x binding buffer. The cell-free extracts, prepared as described above, containing AFase_H4, AFase_E3 and AFase_D3, respectively, were filtered through 0.45 µm syringe filters (Millipore, USA), before being loaded onto the charged and equilibrated columns by gravity flow. Thereafter, the column was washed with 10 column volumes of binding buffer and 6 column volumes of wash buffer (Novagen; 250 mM NaCl, 60 mM imidazole, 20 mM Tris-HCl, pH 7.9). AFase_H4, AFase_E3 and AFase_D3 were eluted from the respective columns with 6 column volumes of elution buffer (Novagen; 250 mM imidazole, 125 mM NaCl, 10 mM Tris-HCl, pH 7.9). Fractions were collected at each step of the Ni-chelation chromatography process for SDS-PAGE analyses as described in section 2.3.3. Column purified proteins were dialysed in dialysis tubing with a 10,000 kDa molecular weight cut off (Thermo Fisher Scientific, USA) for 16 hrs against approximately 200 volumes of Sodium phosphate buffer [50 mM NaCl, 20 mM $NaPO_4$,(pH 7)] at 4°C. Purified AFase_H4, AFase_E3 and AFase_D3 were stored at 4°C following dialysis.

2.3.5 Gel filtration

Fast Protein Liquid Chromatography (FPLC), using a Superdex G200 column, was used to approximate the quaternary structures of AFase_H4, AFase_E3 and AFase_D3, respectively

as performed by Shi *et al.* (2013), dos Santos *et al.* (2011) and Maynard *et al.* (2005). Each of the concentrated of purified AFases (0.26, 1.43 and 3.04 mg of AFase_H4, AFase_E3 and AFase_D3 respectively) were resuspended in 50 mL of running buffer [50 mM NaCl, 20 mM Na_2PO_4 (pH 7)] and loaded onto the column according to the manufacturer's instructions. One mL of each protein marker was injected onto the injection loop with a 1 mL syringe and inserted into the ÄKTA FPLC (Amersham Biosciences) the retention times of the peak fractions were recorded from the following running conditions: flow rate was set at 500 µL/min of the buffer and target protein mixture through the column, the column pressure was set to a limit of 1.5 MPa, the average UV exposure time was set at 10 hours, the column was equilibrated at a rate of 0.1 ml with 50% of the running buffer of the which the AFase protein is stored, the protein sample was emptied out of the loop at 1.1 ml, the fractionation was set to 500 µL and the length of the elution was set at 1.5 min. The retention time of each AFase was used to calculate the approx. molecular weight of the protein from the standard curve. The peak fractions were collected and assayed for AFase activity as described below. The standard protein markers obtained from Sigma-Aldrich, namely β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa) and carbonic anhydrase (29 kDa), were used to generate a standard curve of molecular weight in kDa against retention time. Each of the standards was solubilized in the same running buffer and no protein concentrations were obtained for the standard markers owing to its use only as a marker for this process.

2.4 Enzyme assays

2.4.1 General AFase assay

The *p*-nitrophenyl (*p*NP) linked glycoside, *p*NP-α-L-arabinofuranoside (pNP-ara) (Carbosynth Ltd), was used as a synthetic chromogenic substrate for assaying α-L-arabinofuranosidase activity as performed (with minor modifications) by Amore *et al.* (2012) and Goncalvez *et al.* (2012). The assay was performed in 250 µl containing 2 mM *p*NP-ara, 50 mM buffer (citrate or phosphate) and the reaction started by the addition of 0.5 µg enzyme. 750 µL of 0.4 M Na_2CO_3 was added to yield a final concentration of 0.3 M to stop the reaction by raising the pH of the assay. A 250 µL aliquot of the terminated reaction was aliquoted into a single well of a 96-well microtiter plate (Sterilin®). The amount of *p*NP liberated was measured spectrophotometrically (SPECTROstar Nano; BMG Labtech) at 410 nm, by comparing it to a standard curve generated with *p*NP under assay conditions. One unit (U) of the enzyme is defined as the amount of enzyme that can liberate 1 µmol of *p*NP substrate per minute. All assays were performed in triplicate with the addition of the appropriate negative controls.

2.4.2 DNS assay

Assays performed with complex α-L-arabinose polysaccharide substrates were conducted with 0.5% (w/v) substrate in the optimum buffer and at the optimum temperature for each AFase. Hydolysis was allowed to continue for four hours and the reducing sugar released determined by performing a Dinitrosalicylic Acid (DNS) assay (Miller, 1959). Briefly, 150 µL of DNS reagent [20 mM DNS in 200 mL of NaOH] followed by heating of the solution

prior to yielding a separate 2 M Sodium Potassium Tartrate in 500 mL distilled water. The two solutions were added and volume was compensated to 1 L yielding a final solution of 4 mM DNS and 1 M Sodium Potassium Tartrate] was added to the assay mixture, 50 µL, and boiled for exactly 5 min before being rapidly cooled on ice. The volume was made up to 1000 µL with dH_2O and the absorbance determined at 510 nm (OD_{510}; SPECTROstar Nano; BMG Labtech). In order to determine the amount of L-arabinose hydrolysed from these given complex substrates, an L-arabinose standard curve was generated under the same reaction conditions.

2.5 Biochemical characterisations

2.5.1 Preliminary themostability testing

An initial assessment of the thermostability of the three fosmid encoded AFases was conducted in order to determine any variations in their respective thermostability profiles. The fosmid clones, pFos_H4, pFos_E3, and pFos_D3, were inoculated from their respective glycerol stocks onto LA containing 15 µg mL^{-1} Cam and incubated at 37°C overnight (O/N). A single colony of each was inoculated individually into 10 mL LB containing 0.01% (w/v) L-arabinose and 15 µg ml^{-1} Cam and incubated at 37^{o}C O/N with shaking (160 rpm). Approximately 200 µL of each overnight culture was pipetted into sterile microfuge tubes and the cells lysed by the addition of 10 µL Bugbuster™ protein extraction reagent. The resulting cell-free extracts were incubated, in triplicate, at 25, 40, 50, 60, 70, 80 or 90°C for 60 minutes. A sample kept on ice for the duration of the experiment was used as a control. Following thermal incubation, the reaction tubes were incubated on ice for 30 min, before any insoluble protein and cellular debris were removed by centrifugation (13,000x *g* for 5 minutes at 4°C). Approximately 100 µL of the treated extracts were aliquoted into a flat

bottom 96-well microtiter plate (Sterilin®) and 2 mM *p*NP-ara [100 mM $NaPO_4$ (pH 7)] was added to each well. The microtitre plate was covered with an adhesive foil cover, to prevent evaporation and cross-well contamination, and incubated for 60 min at 37°C. The release of *p*NP, indicating AFase enzyme activity, was determined as described by section 2.4.1. Residual AFase activity in the heat-treated samples was determined as a percentage of the untreated control sample incubated on ice for the duration of the experimental period.

2.5.2 Thermostability

A total of 0.05 mg mL^{-1} of AFase_H4, AFase_E3 and AFase_D3 protein was incubated on either ice or at 60, 70, 80 or 90^{o}C for 5, 10, 15, 30, 60 or 90 min, respectively. Longer incubations were also performed at the following temperatures: 25, 40, 50, 60, 70 and 80^{o}C for 2, 4, 6, 8, 10, 12 and 24 hours. Samples were removed at 2 hourly intervals for 12 hours and after 24 hours, and the AFase activity determined, as described above, relative to a sample maintained on ice for the duration of the 24 hour period. Following incubation at the various temperatures, a *p*NP-ara assay was performed as described in section 2.6.1 to determine the residual AFase activity. The activity of the samples that had been incubated on ice (no thermal treatment) was considered 100% enzyme activity.

2.5.3 pH Optima

Initially, all assays were performed at 40^{o}C as the optimum temperatures for these AFases were unknown. The citrate and phosphate buffers, with pH values between 2 and 8, at a final concentration of 50 mM was used to determine the pH optima of the AFase enzymes. Assays were performed as described in 2.4.1 and a negative control containing no enzyme was

included for each pH buffer tested. Specific enzyme activity in each of the pH buffers was determined, and the pH buffer in which the highest enzyme activity was observed was used in further assays. For AFase_H4, AFase_E3 and AFase_D3 this was pH 5.0, pH 4.0 and pH 4.5, respectively.

2.5.4 Temperature optima

The temperature at which each AFase displays optimum activity was determined by performing *p*NP-ara assays at 6, 20.5, 25, 40, 50, 60, 70, 80 and 90^{o}C. The temperature at which each AFase displayed the highest activity was used as incubation temperature in further assays. For AFase_H4 and AFase_E3 this was 60°C, and forAFase_D3 25^{o}C.

2.5.5 Substrate specificity

The substrate specificity for each of the three purified AFases was determined against various *p*NP-linked glycosides following the method described in section 2.4.1. with the same concentration of the following substrates: *p*-Nitrophenyl-β-D-cellobioside (Sigma), *p*-Nitrophenyl-α-D-mannopyranoside (Sigma), *p*-Nitrophenyl-β-D-fucopyranoside (Sigma) and *p*-Nitrophenyl-β-D-glucuronide (Sigma), all were solubilized according to the manufacturer's instructions in methanol at 65°C. Similarily, *p*-Nitrophenyl-α-L-arabinofuranoside (Carbosynth Ltd), *p*-Nitrophenyl-β-D-xylopyranoside (Carbosynth Ltd), *p*-Methylumbelliferyl-β-D-xylopyranoside (Carbosynth Ltd), *p*-Nitrophenyl-α-D-glucopyranoside (Sigma), *p*-Nitrophenyl-α-L-arabinopyranoside (Sigma), *p*-Nitrophenyl- β -D-glucopyranoside (Sigma), *p*-Nitrophenyl-β-D-mannopyranoside (Sigma) and *p*-

Nitrophenyl-β-L-arabinopyranoside (Sigma) were solubilized, according to the manufacturer's instructions, in methanol at room temperature.

The hydrolytic efficiencies of AFase_H4, AFase_E3 and AFase_D3 were also tested on more complex substrates. The polysaccharide substrates included arabinoxylan, arabinan and linear branched arabinan, with differing internal bonds between arabinose subunits or side chains (Table 2.4; Megazyme, Ireland). Each substrate was prepared according to their respective method of dissolution, according to the manufacturer's instructions. The enzyme activity was determined using the DNS assay as described in 2.4.2.

Table 2.4. Complex α-L-arabinose polysaccharide substrates used to determine L-arabinose liberation.

α-L-arabinose polysaccharide	**α-L-arabinofuranosyl bonds**	**Source**
Arabinoxylan	1,3 and 1,2 on xylan backbone	Rye
Arabinan	1,3 and 1,2 on 1,5 arabinan backbone	Sugar Beet
linear Arabinan	1,5 arabinan backbone	Synthesized

2.5.6 Kinetic characterization

The kinetic parameters including K_m, V_{max} and K_{cat} of AFase_H4, AFase_E3 and AFase_D3 were determined by performing *p*NP-ara assays with increasing amounts of the synthetic *p*NP substrates (section 2.5.5). Assays were performed as described in section 2.4.1., except that the assay time was reduced to 1 min. Change in absorbance over the 1 min was determined

and a Michaelis-Menten plot was constructed using GraphPad Prism 4 (GraphPad Software, Inc). This graph was used to calculate the kinetic parameters of interest.

Chapter 3:

Results and Discussion

3.1 Preliminary analyses

3.1.1 Identification of metagenome-derived α-L-arabinofuranosidases

Metagenomic libraries constructed from compost may contain genetic information encoding potentially novel enzymes from otherwise unculturable microorganisms (Handelsman, 2004; Handelsman, 2007; Pang *et al*, 2009). One of the advantages of the metagenomic approach is that the DNA sequence(s) of the genes encoding these possibly novel biocatalysts may have very low sequence identity to previously characterised proteins with similar functional characteristics (Reisenfeld *et al*, 2004). Therefore functional screening, as compared to bioinformatic analysis and screening of these libraries for genes of interest may result in the isolation of potentially novel biocatalysts (Nacke *et al*, 2012). Furthermore, tailoring the functional screening process may increase the likelihood of identifying enzymes with desired industrial properties (Appendix, Figure A2). Preceding this study, a high temperature compost metagenomic library was constructed in order to identify thermostable lignocellulosic enzymes for use in the enzymatic saccharification of agricultural waste (Ohlhoff, 2012). The metagenomic library consisted of more than 150,000 recombinant *E. coli* Epi300 fosmid clones with an average insert size of approximately 31 kb, which represents a coverage of approximately 1,300 microbial genomes. Approximately 51,300 library clones were screened for α-L-arabinofuranosidase (AFase) activity and 25 putative AFase positive clones were identified and subjected to further analysis (Drs Smart and Huddy). Three fosmid clones, namely pFos_H4, pFos_E3 and pFos_D3, expressing recombinant AFase(s) with initial different thermostability profiles, were selected for analysis in this project (Appendix, Figure A2).

3.1.2 Thermostability of fosmid-encoded AFases

The thermostability profiles of putative AFases encoded by the recombinant fosmids, pFos_D3, pFos_H4 and pFos_E3, expressed in *E. coli* Epi300 cells, were confirmed by conducting thermal inactivation assays of the cell free extracts. All three of the AFases displayed distinct thermostability differences following incubation at various temperatures for 60 min (Figure 3.1). The putative AFase expressed by pFos_D3 was the least thermostable, when compared to the thermostability of the AFases encoded by pFos_H4 and pFos_E3. The AFase activity of the cell free extract prepared from pFos_D3 retained close to 100% activity when incubated at 25 and 40°C, however, at 50°C activity dropped to 90% and 40% at 60°C. At incubation temperatures of 70°C and higher, negligible AFase activity was detected from cell free extract of *E. coli* cells harbouring the pFos_D3 fosmid (Figure 3.1). The relative AFase activity encoded by pFos_H4 increased following pre-incubation at temperatures of 25, 40, 50 and 60°C with activities of greater than 100% measured. Thereafter, the relative activity was seen to decrease following pre-incubation at temperatures greater than 60°C. In contrast, the AFase activity encoded by pFos_E3 displayed an increase in relative activity following pre-incubation at temperatures greater than 40°C, with the highest activity observed following 60 min incubation at 60-70°C. This recombinant AFase was shown to maintain more than 100% of its activity, relative to an untreated control maintained on ice for the duration of the experiment, up to a pre-incubation temperature of 90°C (Figure 3.1). Therefore, on the basis of the thermostability profiles of the crude protein extracts of clones pFos_D3, pFos_H4 and pFos_E3, it was anticipated that a thermolabile, moderately thermostable and a thermostable AFase, respectively had been isolated and these were selected for further experimental analysis.

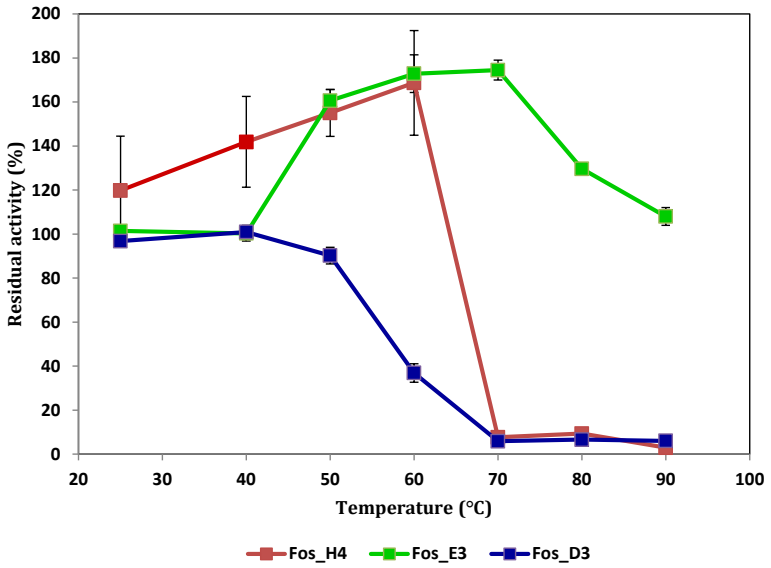


Figure 3.1 Thermostability of cell-free extracts prepared from E. coli Epi300 cells harbouring the recombinant fosmids pFos_H4 (■), pFos_E3 (■) and pFos_D3 (■). The cell-free extracts were incubated at 25, 40, 50, 60, 70, 80 or 90°C for 60 min, before being assayed for AFase activity. The percentage residual activity was calculated compared to the enzymatic activity of extracts maintained on ice for the duration of the experiment and assayed as indicated. Data points represent the average of three replicates ± standard error (n=3).

3.1.3 Restriction analysis and end sequencing of fosmids

The three recombinant fosmids, pFos_D3, pFos_H4 and pFos_E3, were analysed with *Eco*RI and *Hin*dIII to confirm that they contained different metagenomic DNA inserts. The restriction endonuclease 'fingerprints' of the three recombinant fosmids were significantly different (Figure 3.2), which suggested that they encoded different inserts. The copy control fosmid backbone, of approximately 8.2 kb, was present in the restriction patterns generated for all three recombinant fosmids. The size of insert fragments of cloned metagenomic DNA were calculated to be approximately 17.5, 20.7 and 10.7 kb for pFos_H4, pFos_E3 and pFos_D3, respectively. These are unusually small inserts since fosmids, being phage-based

vectors, are generally expected to package DNA genomes of approximately 40 kb (Treangen and Salzberg, 2012). Larger insert sizes have been reported ranging from 19-49 kb (Hu, 2011; Ohlhoff, 2012), however, smaller insert sizes have also been reported (Ghai *et al.*, 2010).

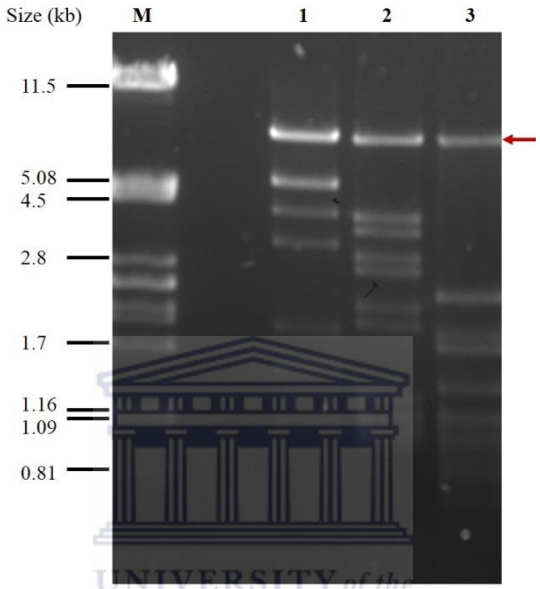


Figure 3.2 The three recombinant fosmids pFos_H4 (lane 1), pFos_E3 (lane 2) and pFos_D3 (lane 3) were restricted with EcoRI and HindIII, and the resulting DNA fragments separated on a 0.8% TAE agarose gel. The fosmid backbone of approximately 8.2 kb is indicated by the red arrow. Lane M contains a λ*Pst*I DNA ladder.

End-sequencing of the three recombinant fosmid constructs confirmed the initial conclusion, based on the restriction enzyme fingerprinting analysis, that these three fosmids contained unique inserts (Appendix Figure 3.8). The end-sequences have sequence identity to different microorganisms. In general, the end sequences to all the fosmid inserts displayed very low sequence similarity to a range of organisms which can be expected to be associated with compost (Table 3.1). However, due to any significant and consistent identities, it is not possible to deduce from which organism the cloned inserts from the 3 fosmids could have

originated. However, this does highlight the significant level to which novel sequences are continuously discovered through functional metagenomic screens.

Table 3.1 The putative domains from the BLASTx alignment of the fosmid forward and reverse end sequences.

FOSMID ID	Putative Domains	Micro-organism	% identity	Accession No.
FOS_H4_pCC1	ABC-ATPase Transporter	*Streptococcus agalactiae*	37	WP001185986.1
FOS_H4_T7	Hypothetical Protein	*Thermobaculum terrenum*	41	YP003322965.1
FOS_E3_pCC1	Glutamate synthase	*Bacillus alcalophilus*	50	WP00332326.1
FOS_E3_T7	Metallo-dependant hydrolases	*Paenibacillus sp*. HW 567	53	WP019910497.1
FOS_D3_pCC1	α-L-Fucosidase	*Paenibacillus sp*. PAMC 26794	82	WP017688670.1
FOS_D3_T7	HSP90-ATPase	*Flavobacterium rivuli*	83	WP020214552.1

3.2 Sequence Identification and bioinformatic analysis

3.2.1 Transposon mutagenesis

Metagenomic libraries are constructed using large-insert accepting vectors such as fosmids, cosmids or BACs to reduce the number of clones required for a significant genome coverage of the metagenomic library (Nacke *et al*., 2012; Handelsman, 2007; Pang *et al*., 2009; Handelsman, 2004). The metagenomic DNA library from which the putative AFases were identified, was constructed using a fosmid system kit (Epicentre Biotechnologies). As the fosmids contain large metagenomic DNA inserts, the subsequent identification and sequencing of the gene(s) of interest is challenging. However, various options are available to obtain these sequences. Next-generation sequencing (NGS) or conventional Sanger sequencing by primer walking may be used to obtain the DNA sequence of the entire

fragment of cloned metagenomic DNA. Alternatively, the fosmid may be subjected to transposon mutagenesis to identify the gene encoding the desired enzyme activity by the loss of activity following the insertion of the transposon into the gene sequence. The gene is then sequenced by Sanger sequencing with primers reading out from the transposon. The latter is an ideal approach to use for the identification of potentially novel genes when a simple agar-based screening assay is available. The former two options carry significant financial cost implications and could require in-depth bioinformatic knowledge to analyse NGS data. Furthermore, both NGS and primer walking have the previously highlighted drawback of the need to be able to identify the potentially novel gene of interest by subsequent bioinformatic searches of sequence databases. If the gene sequences are entirely novel, sequence similarity to previously identified genes may not be achieved using general databases. Transposon mutagenesis has been successfully used in a number of studies to identify and sequence genes of interest from clones harbouring large fragments of genomic DNA (Reisenfeld *et al*., 2004; Gao, 2012) Therefore, it was deemed suitable for the identification of the genes harbouring AFase activity on the pFos_H4, pFos_E3 and pFos_D3 fosmids (Appendix, Figure 3.9).

Transposon mutagenesis of fosmid pFos_H4 resulted in 9 constructs for which minimal AFase activity was detected and 1 mutant (H4AC) displaying significantly reduced AFase activity, when activities were compared to the activity obtained from the non-mutated pFos_H4 control (Figure 3.3 A). Where reduced AFase activity is observed the transposon may have inserted within a non-coding region, potentially containing a transcriptional regulatory sequence (Ruiz *et al.*, 2013). Transposon mutagenesis of pFos_E3 and pFos_D3, generated 10 and 7 mutants displaying minimal AFase activities relative to their untreated fosmid controls, respectively (Figure 3.3 B and C).

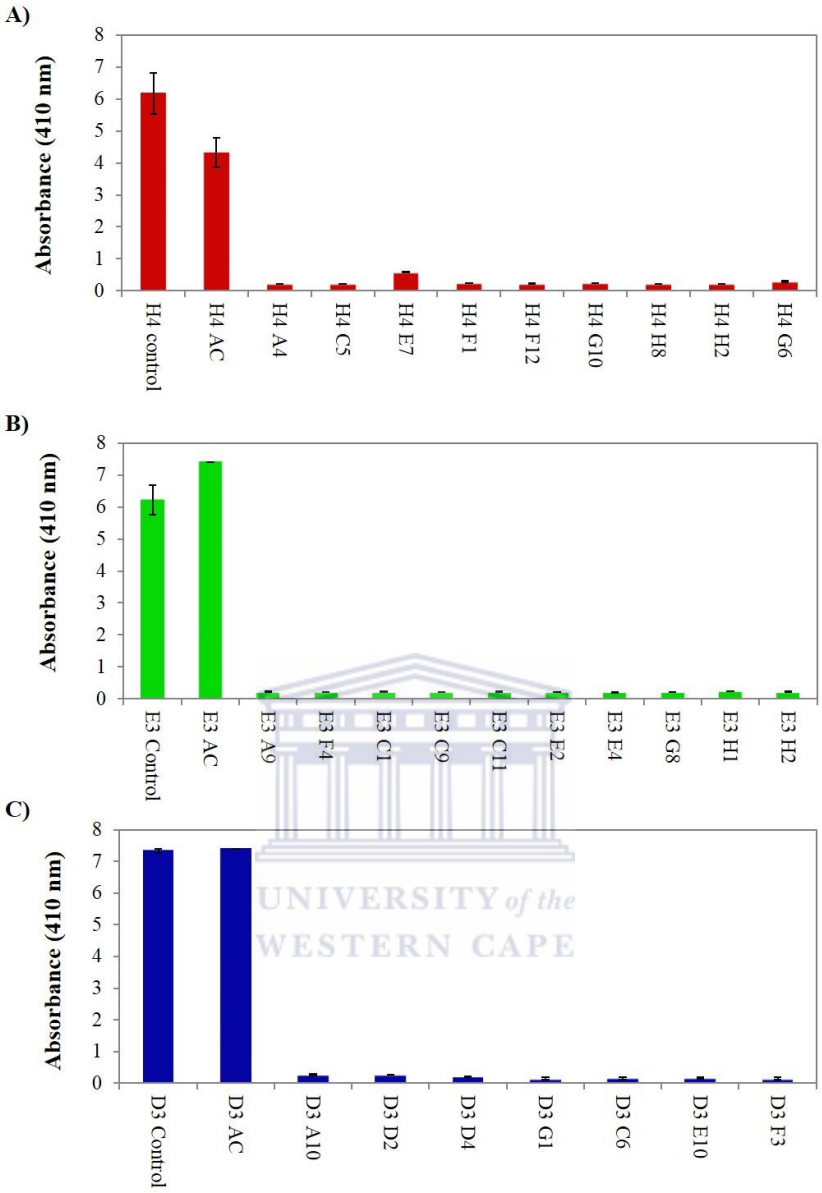


Figure 3.3 Transposon mutagenesis of pFos_H4 (A; ■), pFos_E3 (B; ■) and pFos_D3 (C; ■). H4, E3 and D3 control samples represent the fosmids that had not been treated with the transposon and were included as a positive control for the AFase enzymatic assay. The H4 AC fosmid clone sample retained AFase activity, however was significantly reduced compared to the untreated control. Data represents the average of three replicates ± standard error (*n*=3).

Five, four and six mutated fosmid clones displaying minimal AFase activity were chosen for sequencing from the pFos_H4, pFos_E3 and pFos_D3 clones, respectively. From these, 2693, 2406 and 2505 bp of sequence information was obtained for the fosmids pFos_H4, pFos_E3 and pFos_D3, respectively (Figure 3.4 A-C).

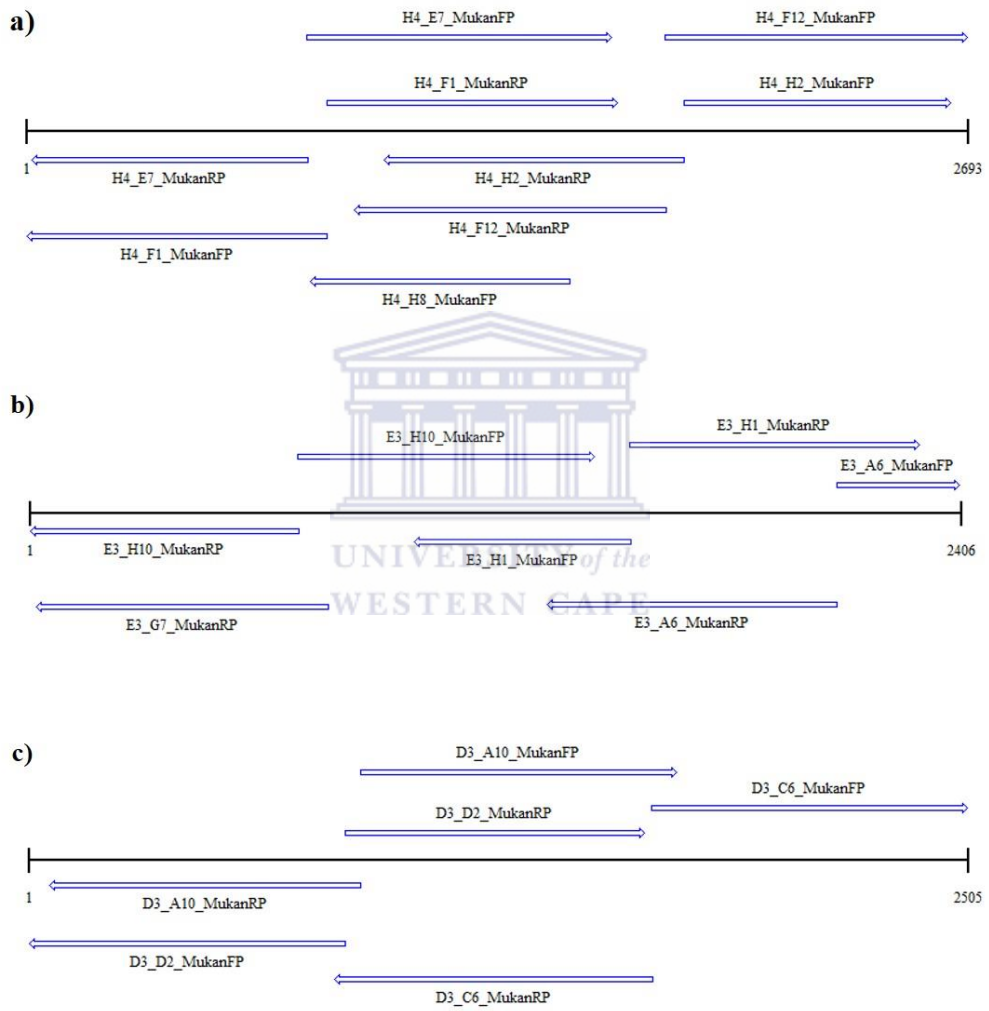


Figure 3.4 Diagramatic representation of the consensus sequences assembled assembling forward (MuKanFP) and reverse (MuKanRP) sequences of several transposon-treated, AFase-inactivated pFos_H4 (A), pFos_E3 (B) and pFos_D3 (C) fosmids. Letters and numbers after the fosmid name represent the microwell of the microtitre plate from which the transposon mutant phenotype was located and corresponds with the naming depicted in Figure 3.3.

3.2.2.1 ORF Annotation from pFos_H4

Three putative open reading frames (ORF's) were identified within the 2693 bp sequence of pFos_H4 (Figure 3.5.1). The putative ORFs had 54 %, 61 % and a 51 % sequence similarity to the following proteins; respectively: an extracellular-arabinose-binding protein (*Clostridium sp*.) an AFase protein (*Truepera radiovictrix)* and a transcriptional regulator implicated in the expression of genes involved in purine metabolism (PurR) (*Thermobaculum terrenum).* Extracellular-arabinose-binding proteins (EAB) are ABC-type (ATPase binding cassette) transporters and often form part of operons involved in arabinose catabolism (Desai and Rao, 2010). The transporter is involved in the transport of arabinose into the cell and this could require the adjacent *AFase_H4* gene to execute the hydrolysis in the extracellular environment. However, no signal peptides were detected upstream of the *AFase_H4* ORF or on reported prokaryotic AFase amino acid sequences. PurR transcriptional regulators have not been shown to be involved in arabinose metabolism and contain ligand-binding and helix-turn-helix (HTH) DNA binding domains similar to those from the *Lac*I transcriptional regulator family (Jendresen *et al.*, 2012). Lactose binds to the ligand-binding domain which decreases the binding affinity for the operator sequence of this purine metabolism operon (Jendresen *et al.*, 2012). Purine biosynthesis involves the generation of ribose-5-phosphate from the pentose phosphate pathway during the oxidation of galactose and glucose from which lactose is hydrolysed (Hove-Jensen *et al.*, 2003).

The translated sequence of *AFase_H4* was subjected to BLASTp analysis and the top 10 hits are given in Table 3.2. Taking all the sequence data generated from pFos_H4 (Tables 3.1 and 3.2) together, it seems feasible that this DNA might have originated from a thermophilic

organism. Analysis of the sequence for the presence of domains revealed that AFase_H4 contained a putative catalytic domain, AFase C-terminal domain and a putative cellulase-like domain (Figure 3.5.1).

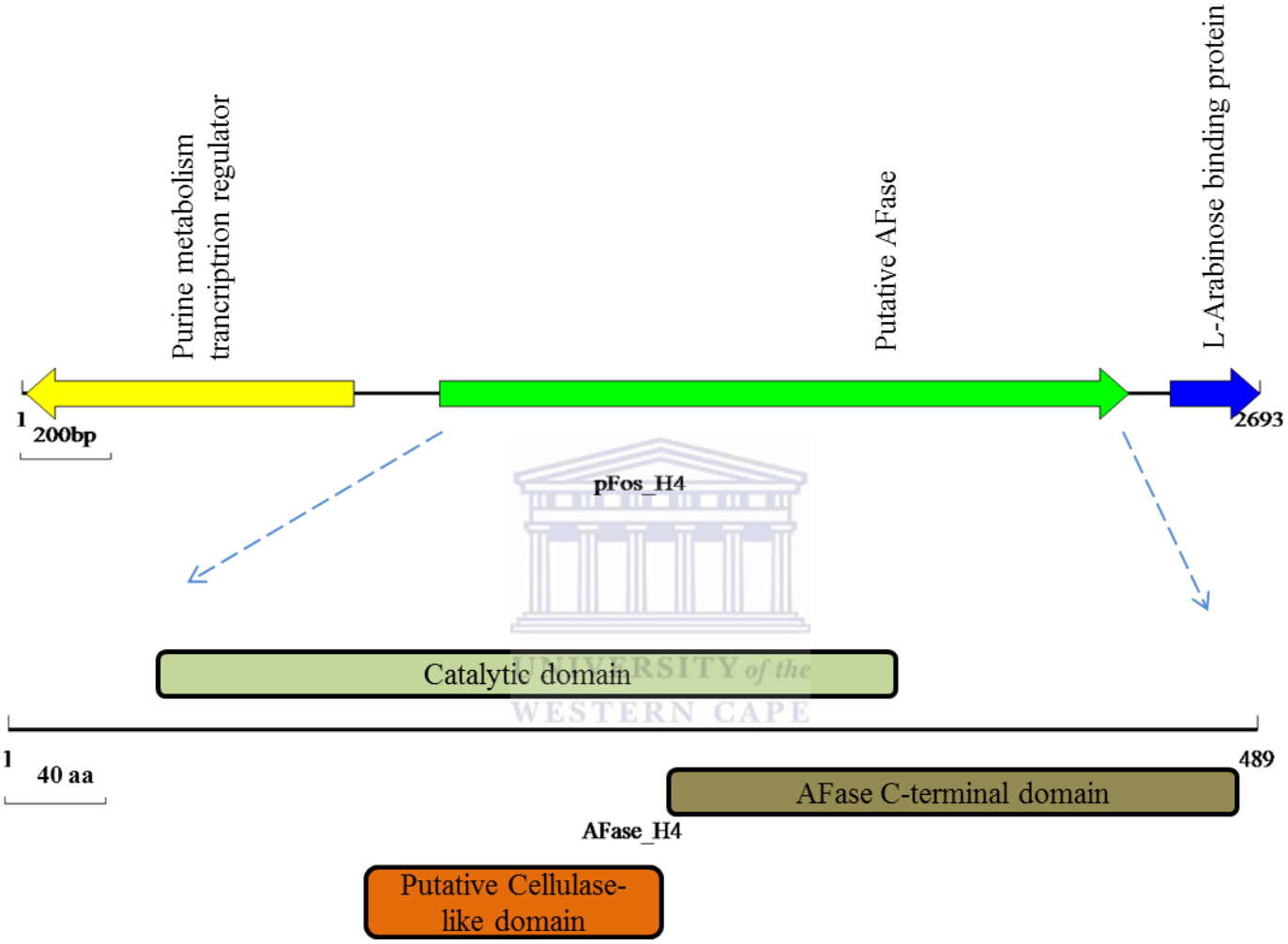


Figure 3.5.1 The identification of the AFase_H4 open reading frame (ORF's) from the consensus sequences was obtained using DNAMAN (Ref) and the functional domains within the AFase ORF was identified using Interproscan.

Table 3.2 BLASTp analysis results using the in silico translated AFase_H4 sequence as the template (NCBI Database accessed 28/02/2014).

Hit No.	Micro-organism	Protein	Accession no.	% Sequence coverage	% Sequence identity	E-value
1	*Truepera radiovictrix* DSM 17093	AFase	YP_003703846.1	99	61	0
2	*Anaerophaga thermophila*	AFase	WP_016778898.1	98	48	6e-159
3	*Alistipes onderdonkii*	AFase	WP_018696806.1	98	48	6e-157
4	*Alistipes sp* CAG:29	AFase	WP_022333570.1	98	48	3e-156
5	*Anaerophaga thermophila*	AFase	WP_010420942.1	98	48	1e-155
6	*Ktenobacter racemifer*	AFase	WP_007915147.1	99	47	2e-155
7	*Candidus Solibacter usitatus*	AFase	YP_824901.1	99	48	8e-155
8	*Marinilabilia salmonicolor*	AFase	WP_010665024.1	98	46	1e-152
9	*Rhodothermus marinus*	AFase	YP_004824259.1	98	47	1e-150
10	*Acidobacteriaceae bacterium* KBS	AFase	WP_020720356.1	99	44	8e-143

3.2.2.2 ORF annotations from pFos_E3

Two putative ORF's encoding an AFase and a flanking ABC transporter were identified within the consensus sequence of pFos_E3 following BLASTp analysis. The ABC transporter ORF indicated a sequence query coverage of 7% with 57% sequence similarity to a similar ORF from *Desmospora sp* 8437. Unlike the ABC transporter identified on pFos_H4, no extracellular-arabinose-binding domain was present within the ABC transporter of pFos_E3. ABC transporters are highly abundant and have therefore been thoroughly studied as a protein superfamily. They display a diverse range of functions such as the translocation of ions, lipids, peptides, sugars and other hydrophilic substrates into the prokaryotic cell (Bendsten *et al.*, 2004). BLASTp analysis of the *in silico* translated *AFase_E3* sequence showed 68% sequence identity to an AFase from *Paenibacillus sp*. PAMC 26794. Certain

Paenibacillus spp are thermophiles that have been shown to proliferate in compost at temperatures between 50 and 70^{o}C and are involved in the degradation of humic organic compounds (Park and Kim, 2013; Charbonneau *et al.*, 2012). *Desmospora spp* are thermoactinomycetes that have been isolated from thermal compost and have indicated optimum proliferation between 30^{o}C-60^{o}C which suggests involvement in both the mesophilic and thermophilic stages of composting (Chang *et al.*, 2009; Yassin *et al.*, 2009). *Thermobacillus composti*, a moderately thermophile, was isolated from fed-batch composting reactors exceeding 50^{o}C (Watanabe *et al.*, 2007). *AFase_E3* displays sequence similarity to a number of *Geobacillus sp* (Table 3.3.). These are soil dwelling genera that have been isolated from both mesophilic and thermophilic soil environments. Specifically *Geobacillus thermodenitrificans* have been isolated from manure compost that reaches temperatures of 50-60^{o}C and have been shown to be important in compost maturation (Charbonneau *et al.*, 2012). The *in silico* translated *AFase_E3* sequence was subjected to domain identification analysis through interproscan. Figure 3.5.2 demonstrates the putative catalytic domain and an AFase C-terminal domain.

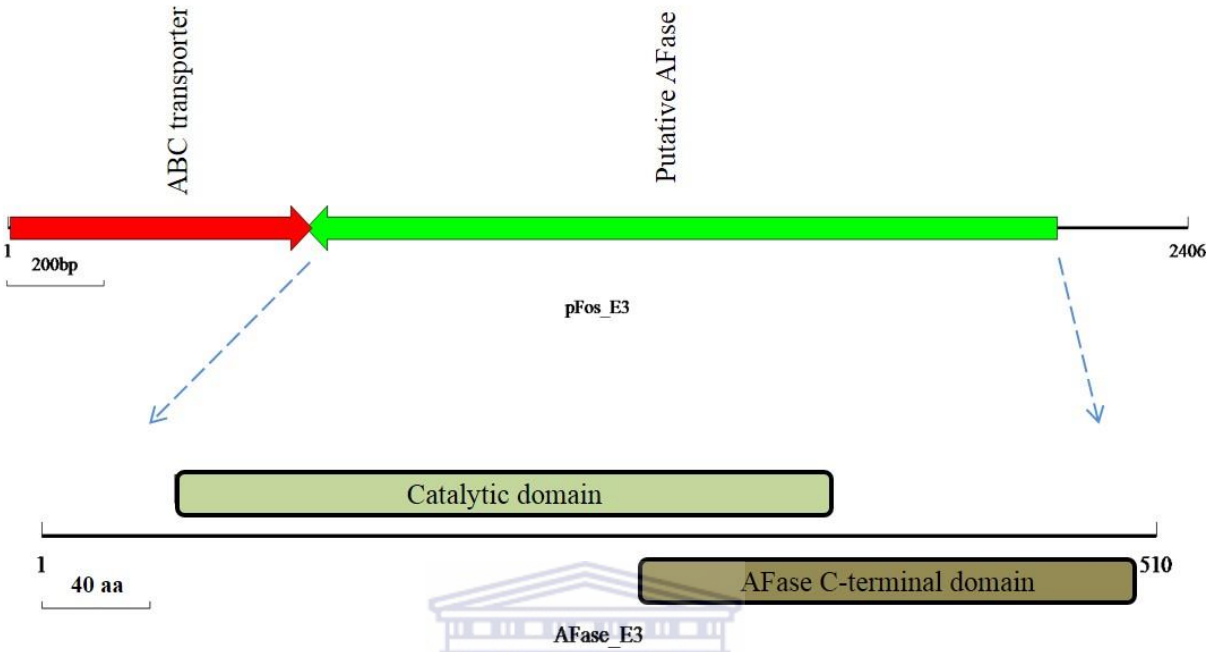


Figure 3.5.2 The identification of the AFase_E3 open reading frame (ORF's) from the consensus sequences was obtained through DNAMAN and the functional domains within the AFase ORF was identified through Interproscan.

Table 3.3: BLASTp analysis results using the in silico translated AFase_E3 sequence as the template (NCBI Database accessed 28/02/2014).

Hit No.	Micro-organism	Protein	Accession no.	% Sequence coverage	% Sequence identity	E-value
1	*Paenibacillus sp*. PAMC 26794	AFase	WP_017690226.1	97	68	0
2	*Desmospora sp*. 8437	AFase	WP_009711777.1	99	68	0
3	*Thermobacillus composti* KWC4	AFase	YP_007211736.1	97	69	0
4	*Geobacillus sp* G11MC16	AFase	WP_008880048.1	97	67	0
5	*Geobacillus sp* GHH01	AFase	YP_007402223.1	97	67	0
6	*Geobacillus sp* JF8	AFase	YP_008417789.1	97	67	0
7	*Geobacillus thermodenitrificans* NG80-2	AFase	YP_001125896.1	97	66	0
8	*Geobacillus sp* A8	AFase	WP_021321564.1	97	66	0
9	*Geobacillus sp* C56-T3	AFase	YP_003671207.1	98	66	0
10	*Geobacillus sp* MAS1	AFase	WP_023633860.1	98	66	0

3.2.2.3 ORF annotations from pFos_D3

Only a single ORF was identified within the consensus sequence of pFos_D3 after the sequencing of three transposon mutated clones. The ORF was confirmed to be a putative AFase nucleotide sequence. BLASTp results (Table 3.4.) revealed that *AFase_D3* shared 69% sequence similarity to an AFase from *Geobacillus sp*. Y412MC61, as well as to other *Geobacillus* and *Paenibacillus* species (Table 3.4). Some *Paenibacillus spp* are thermophilic bacteria that have shown to grow within compost (as previously mentioned from the pFos_E3 BLASTp result) and within geothermal heat soils (Tin *et al*, 2011). The *AFase_D3* ORF was subjected to domain identification analysis through interproscan (Figure 3.5.3). The domain analysis of AFase_D3 did not result in the identification of a putative catalytic domain as was identified for *AFase_H4* and *AFase_E3,* however, an AFase C-terminal domain was identified. Instead the AFase_D3 aa sequence presented a superfamily glycoside hydrolase domain that is ubiquitously present in all AFases proteins. The location of the catalytic domain is within the superfamily glycoside hydrolase domain similar to AFase_H4 and AFase_E3 (not shown in their respective Figures).

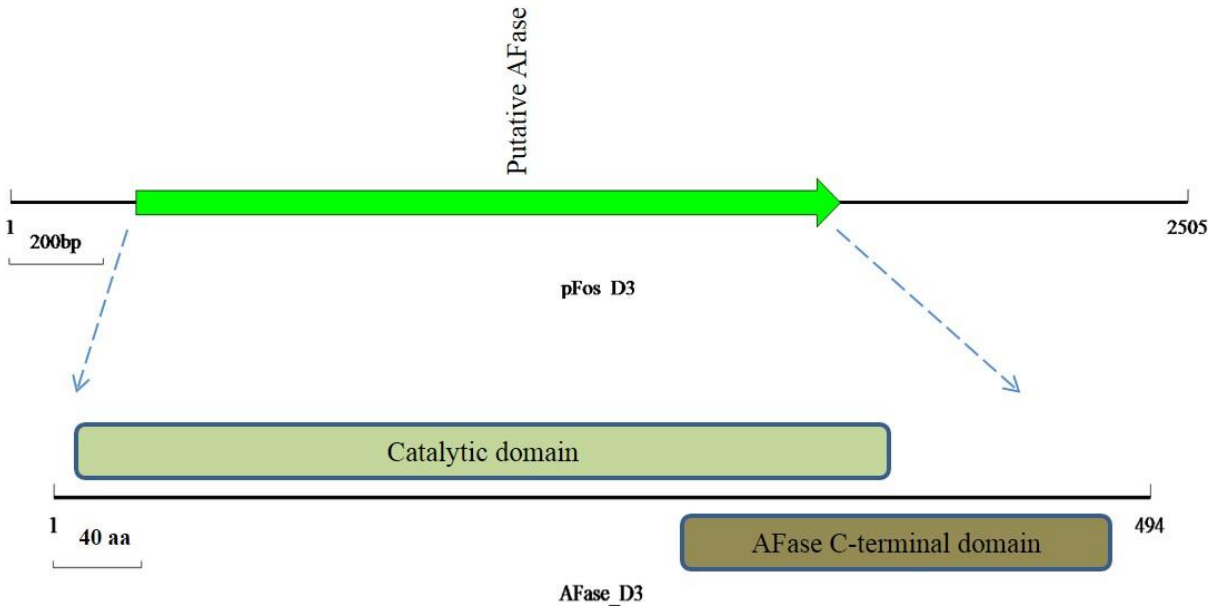


Figure 3.5.3 The identification of the AFase open reading frame (ORF's) from the consensus sequences was obtained through DNAMAN and the functional domains within the AFase ORF was identified through Interproscan.

Table 3.4: BLASTp analysis results using the in silico translated AFase_D3 sequence as the template (NCBI Database accessed 28/02/2014).

Hit no.	Micro-organism	Acession no.	% Sequence Coverage	% Sequence identity	E-value
1	*Geobacillus sp*. Y412MC61	YP_003253791.1	99	69	0
2	*Geobacillus stearothermophilus*	ACE73681.1	99	69	0
3	*Paenibacillus ginsengihumi*	WP_019536202.1	99	70	0
4	*Geobacillus sp*. WSUCF1	WP_020755790.1	99	69	0
5	*Paenibacillus mucilaginosus* KNP414	YP_004644582.1	99	70	0
6	*Paenibacillus mucilaginosus* 3016	YP_005315634.1	99	70	0
7	*Paenibacillus sp*. HW567	WP_019912317.1	98	67	0
8	*Paenibacillus fonticola*	WP_019636755.1	99	68	0
9	*Paenibacillus barengtlzii*	WP_016313143.1	98	68	0
10	*Paenibacillus sp.* oral taxon 786	WP_009226440.1	98	67	0

3.2.3 Phylogenetic analysis of AFase_H4, AFase_E3 and AFase_D3

AFase sequences were retrieved from the GH family 43, 51, 54 and 62, and the phylogenetic relationship was analysed. Figure 3.6 indicated that all three AFases cluster within the GH 51 AFases. Furthermore the sequences of the putatively thermostable AFase_H4 and AFase_E3 were more related in comparison to the less thermostable AFase_D3. AFase_E3 clustered with AFase sequences derived from thermophiles such as *Geobacillus spp* and *Thermotoga spp*. AFase_D3's sequence is more closely related to the AFase derived from mesophilic micro-organisms such as *Bacillus subtilis*. Although AFase_H4 clustered more closely with the thermophile derived AFases, it is distinctly and deeply branched. AFase_E3 shares a sequence similarity to a *Bacillus subtilis* derived AFase. *Bacillus subtilis* contain mesophilic and thermophilic AFases, and AFase_E3 and D3 clustered with the mesophilic and thermophilic *Bacillus subtilis*_AFase, respectively (Inacio *et al*., 2008; Hoffman *et al*., 2013). All other functionally characterized AFases within the GH 51 family are declared to be thermostable (indicated in Figure 3.6). AFases from *Streptosporangium roseum* (EEP18193.1), *Kribella Flavida* (EEJ18245.1), *Catenulispora acidiphila* (EEN29837.1), *Cytophaga xylanolytica* (O68278), *Cellvibrio japonicas* (ACE86344.1), *Cytophaga xylanolytica* (O68279) and *Bacteroida ovatus* (Q59219.1) Therefore the phylogenetic relationship of AFases within the GH 51 family cannot be resolved on the basis of thermostability.

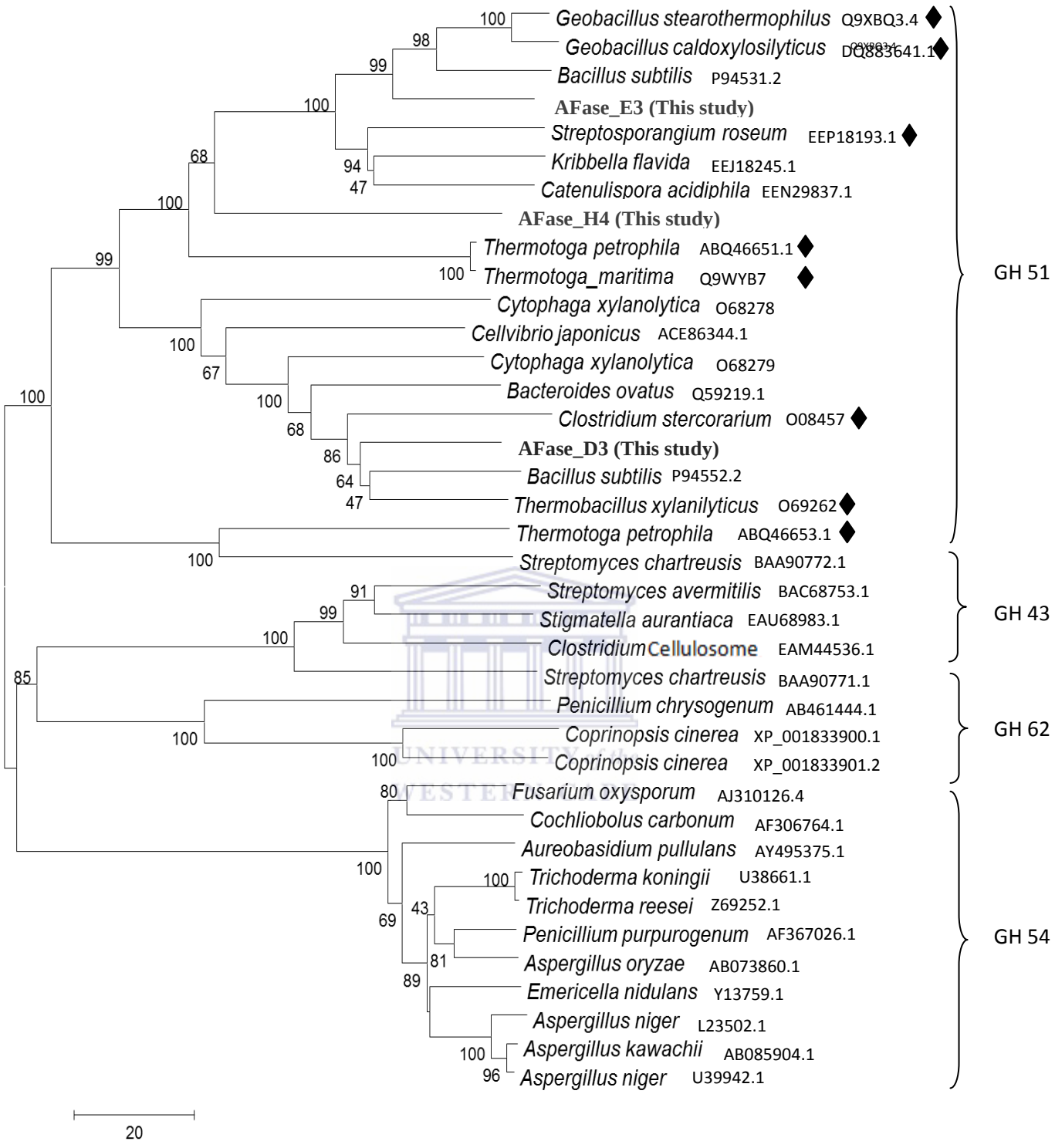


Figure 3.6 The phylogenetic relationship of AFase_H4, AFase_E3 and AFase_D3 to characterised AFases from within the GH family 43, 51, 54 and 62 (as indicated). The accession numbers for the characterised AFases are indicated in brackets. The functionally characterized, as thermophilic, AFases are indicated with the black diamond symbol (◆). The scale bar indicates the number of amino acid substitutions.

3.2.4 Nucleotide sequence analysis and classification

3.2.4.1 AFase_H4 nucleotide sequence

An open reading frame (ORF) of 1467 nucleotides within pFos_H4 was predicted to begin at the putative methionine start codon (ATG) at position 1 (M^1) and continue to the in-frame TGA stop codon at position 490 marked with the asterix (*) (Figure 3.7.1). This ORF encodes a 489 amino acid protein with a predicted molecular weight of approximately 54.9 kDa. Putative transcriptional regulatory sequences (-10 and -35 promoter elements) were identified in regions upstream of the start codon (Figure 3.7.1). Conserved catalytic and AFase C-terminal domains were identified within the predicted amino acid sequence (Figure 3.7.1), and no evidence of a signal peptide within the predicted translated amino acid sequence was detected. This could indicate that a "non-classical protein secretion" mechanism is used to export the AFase_H4 protein (Bendtsen *et al*, 2004, 2005). The catalytic domain overlaps the AFase C-terminal domain and is positioned from P^{41}-A^{346} and E^{261}-V^{482}, respectively (Figure 3.7.1). A cellulase-like domain is located within the catalytic domain from V^{161}-W^{240}. The catalytic domain has sequence identity to that of Glycoside Hydrolase (GH) family 51 (GH51) as depicted from Figure 3.6. Therefore, this ORF was likely to be responsible for the AFase activity encoded by pFos_H4 and was designated AFase_H4.

```
-236      CTCTCTTGTTTAATGTACAAGTATTTGACAGGTTTGGCAAAACGGCATATATATAGATTG
          -35                      -10
-177      TGAAAGCGCTCGACTGAACGCACTTATGCTTAATTTAGCATTCTAACCTTGCAGGCGATT

-118      TCGGGAATTCGGCGCTGAATGGGCCGTATAGATTTAGTAAACACGCCGGCCAAGAATCAG

-59       CTTTTTAGGGCGTATTTCGCCTGAAACTGGACGTTGTTTTCATGAATCACATCAAGATTG

1         ATGAATCACATCAAGATTGATTTAGATCGTCAATCCGGCTGTATCCACCCTCATATTTTT
1          M  N  H  I  K  I  D  L  D  R  Q  S  G  C  I  H  P  H  I  F

61        GGCGGCTTCGTCGAACATCTGGGACGCTGCGTGTATGGCGGCATCTACGAGCCCGGTTCG
21         G  G  F  V  E  H  L  G  R  C  V  Y  G  G  I  Y  E  P  G  S

121       CCATTGGCCGACAAACAGGGTTTTCGCCGGGATGTTCTGGAAGCCCTGCAGCGACTGAAC
41         P  L  A  D  K  Q  G  F  R  R  D  V  L  E  A  L  Q  R  L  N

181       CTGTCTATTGTACGCTATCCCGGGGGCAACTTCGTATCGGGCTATCGCTGGATGGATGGT
61         L  S  I  V  R  Y  P  G  G  N  F  V  S  G  Y  R  W  M  D  G

241       ATCGGTCCGGTCGAAGATCGCCCCACCCGGCCTGACCTCGCCTGGGGAGCCATCGAAACG
81         I  G  P  V  E  D  R  P  T  R  P  D  L  A  W  G  A  I  E  T

301       AATCACTTTGGAACGAACGAATTCATCCAGTTCTGCCGCACGATCAACGCTGAACCCTAT
101        N  H  F  G  T  N  E  F  I  Q  F  C  R  T  I  N  A  E  P  Y

361       CTGGTGGTCAATTGTGGCGATGGGGACATGCGCGAAGCGCGGGACTGGGTGGAATACTGC
121        L  V  V  N  C  G  D  G  D  M  R  E  A  R  D  W  V  E  Y  C

421       AACGGCACTCAGAACACAGCCCTCGCCAACCTGCGCCGCCAGCACGGCTTCGACGCGCCT
141        N  G  T  Q  N  T  A  L  A  N  L  R  R  Q  H  G  F  D  A  P

481       CATAACGTGAAATATTGGGGCATTGGCAATGAGGTCGATGGCGACTGGCAGATCGGCTAT
161        H  N  V  K  Y  W  G  I  G  N  E  V  D  G  D  W  Q  I  G  Y

541       AAGACCGCGCAGGAATACGCTCGCGCCTATAAAGAATTCGCTAAAGTCATGCGCTGGGTC
181        K  T  A  Q  E  Y  A  R  A  Y  K  E  F  A  K  V  M  R  W  V

601       GACCCATCGATCAGCCTGCTGGCTTCAGCGGTATCCAGTTGGCGAACGGACTTTGTCGAA
201        D  P  S  I  S  L  L  A  S  A  V  S  S  W  R  T  D  F  V  E

661       CGTATCCAACTTCTCCTCGACCACGCGCCAGACCTGATCGACTATCTCGCCATCCACTGG
221        R  I  Q  L  L  L  D  H  A  P  D  L  I  D  Y  L  A  I  H  W

721       TATGTCGGCAATCCAGAGGGCGATTTCGAAAAGTATATGGCCGTATCCGAGCTGATCGAG
241        Y  V  G  N  P  E  G  D  F  E  K  Y  M  A  V  S  E  L  I  E

781       GAGCGGATCAGCGCTATCGAAGGCCTGATCAGGGTGATGAAATTGCAGCGCAACATTCAG
261        E  R  I  S  A  I  E  G  L  I  R  V  M  K  L  Q  R  N  I  Q
```

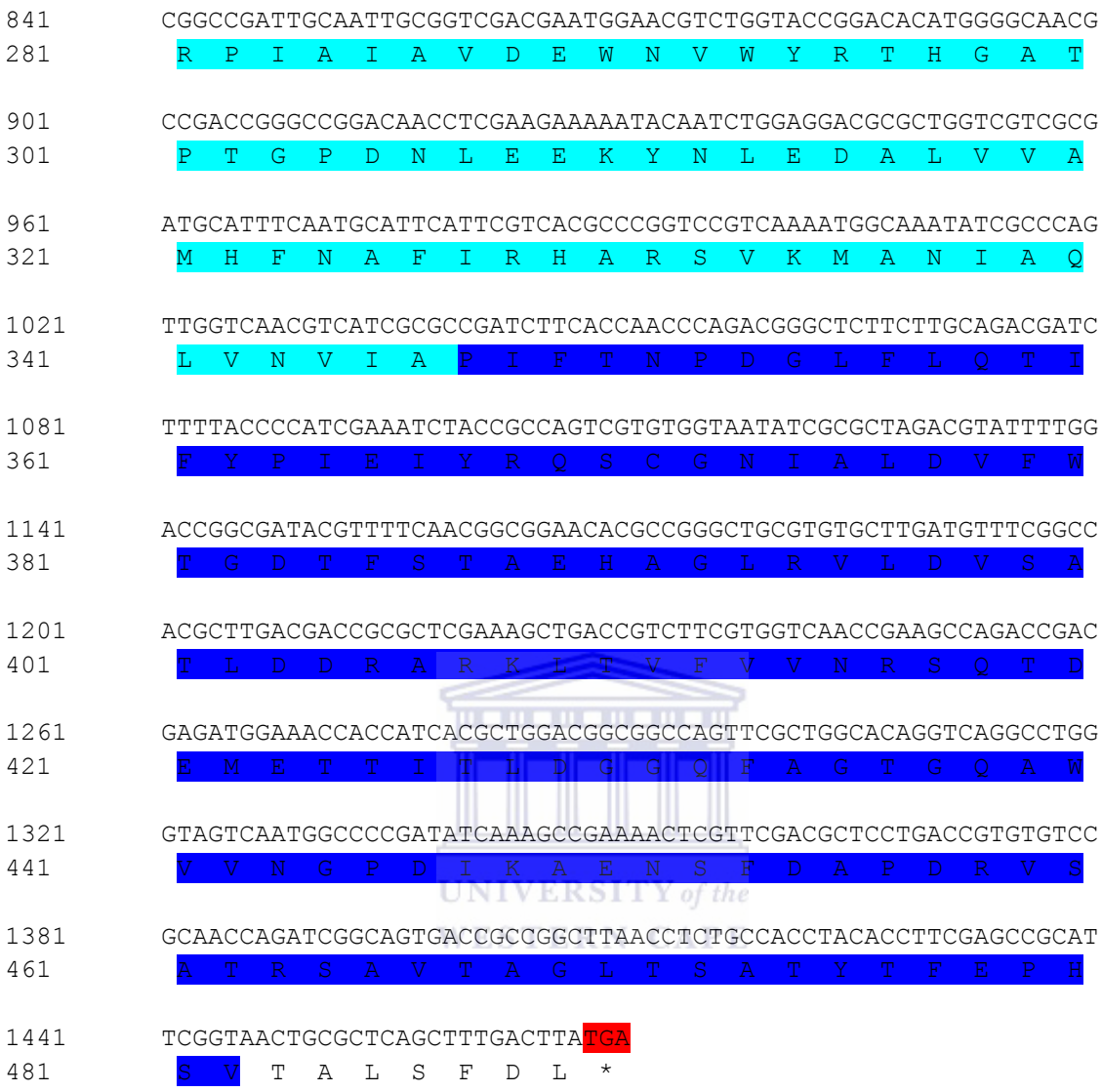


Figure 3.7.1 The nucleotide sequence of AFase_H4 and the deduced amino acid sequence of the AFase protein it encodes, AFase_H4. The putative -10 and -35 promoter regions are underlined and labelled. The ATG start codon is highlighted in green, while an asterix (*) and red highlighting indicates the TGA stop codon. The conserved catalytic domain is shaded in yellow, while the AFase C-terminal domain is shaded in dark blue and the overlapping portion of the two domains is shaded in turquoise. The cellulase-like domain is indicated in bold and italics.

3.2.4.2 AFase_E3 nucleotide sequence

An open reading frame (ORF) of 1530 nucleotides within pFos_E3 was predicted to begin at the putative methionine start codon (ATG) at position 1 (M^1) and continue to the in-frame TGA stop codon at position 511 marked with the asterix (*) (Figure 3.7.2). This ORF encodes a 510 amino acid protein with a predicted molecular weight of approximately 57.5 kDa. Putative transcriptional regulatory sequences were identified in the vicinity of the -10 and -35 regions upstream of the translational start codon (Figure 3.7.2). There was no evidence of a signal peptide within the predicted translated amino acid sequence, however, conserved glycoside hydrolase (GH) family 51 (as depicted from Figure 3.6) catalytic and AFase C-terminal domains were identified (Figure 3.7.2). Therefore, this ORF was likely to be responsible for the AFase activity encoded by pFos_E3 and was designated AFase_E3. The conserved GH51 catalytic domain, identified within AFase_E3, overlaps the AFase C-terminal domain and is positioned from P^{47}- C^{362} and D^{281} - W^{500}, respectively.


```
-118    GCGCGGTCCCGATCGTCGGGATCGCGCCGTTTTGCTGTTCTGAATTGACGGACCGCCGCC
                  -35                  -10
-59     GCTGTGGTTTAATATGTATGGTTATGTTTTTATTATTTAATTCATATTTATATTAATCAA

1       ATGGACGGAGGCGCATGCGTCATGGCATCGGCCAAACTTCGCGTCGATCGCGCGTATACG
1        M  D  G  G  A  C  V  M  A  S  A  K  L  R  V  D  R  A  Y  T
61      ATCGGCGATACGGATCCGCGGTTGTTCGGGGCGTTCGTTGAACATCTCGGCCGGGCGGTG
21       I  G  D  T  D  P  R  L  F  G  A  F  V  E  H  L  G  R  A  V
121     TACGGCGGGATTTACGAGCCGGGGCATCCGGAAGCTGACGAACAAGGATTCCGGCGGGAC
41       Y  G  G  I  Y  E  P  G  H  P  E  A  D  E  Q  G  F  R  R  D
181     GTGCTGGAGCTCGTCCGCGAGTTGGGCGTCCCGATCGTTCGTTACCCGGGCGGAAATTTC
61       V  L  E  L  V  R  E  L  G  V  P  I  V  R  Y  P  G  G  N  F
241     GTGTCGGGATATAACTGGGAAGACGGCGTCGGCCCGAAGTCCGAGCGCAAACGCCGTCTT
81       V  S  G  Y  N  W  E  D  G  V  G  P  K  S  E  R  K  R  R  L
301     GAGCTCGCCTGGCGATCGATCGAGACCAACGAATTCGGGACGAACGAATTCGTCGATTGG
101      E  L  A  W  R  S  I  E  T  N  E  F  G  T  N  E  F  V  D  W
361     TGCAAAAAAGCAGGGGCCGAACCGATGCTCGCAGTCAACCTCGGCACGCGCGGCATCGAC
121      C  K  K  A  G  A  E  P  M  L  A  V  N  L  G  T  R  G  I  D
421     GAGGCCCGCAATCTGGTGGAATATTGCAATTACCCCTCAGGCACGTATTGGAGCGACTTG
```

```
141          E  A  R  N  L  V  E  Y  C  N  Y  P  S  G  T  Y  W  S  D  L
481        CGGGTGCGGCACGGCTACCGCGAACCGCATCGCGTCAAAGTCTGGTGTCTCGGCAATGAA
161          R  V  R  H  G  Y  R  E  P  H  R  V  K  V  W  C  L  G  N  E
541        ATGGATGGATCCTGGCAAATCGGCAGCAAAACGGCCGACGAATACGGACGGTTGGCCTGC
181          M  D  G  S  W  Q  I  G  S  K  T  A  D  E  Y  G  R  L  A  C
601        GAAACTGCCAAGGCGATGAAGTGGGTCGACCCGTCGATCGAACTCGTCGCCTGCGGCAGC
201          E  T  A  K  A  M  K  W  V  D  P  S  I  E  L  V  A  C  G  S
661        TCGAGTCCTTCCATGCCGACGTTTCCGGAATGGGAAAGGATCGTTCTCGAACATACGTAC
221          S  S  P  S  M  P  T  F  P  E  W  E  R  I  V  L  E  H  T  Y
721        GATCACGTCGACTACGTGTCGCTGCACATCTATTTCGGCAATCCGGAAAACGACACGGCA
241          D  H  V  D  Y  V  S  L  H  I  Y  F  G  N  P  E  N  D  T  A
781        AACTTTCTCGCCAAGCCGCTGGATATGGAAAGGTTTATCCGGATCGTCAAAGCGACGTGC
261          N  F  L  A  K  P  L  D  M  E  R  F  I  R  I  V  K  A  T  C
841        GATTACGTCAAAGCGAAGAAGCGCGGTAAAAAGGATCTTTACATTTCGTTCGATGAGTGG
281          D  Y  V  K  A  K  K  R  G  K  K  D  L  Y  I  S  F  D  E  W
901        AACGTCTGGTATCATTCCCATGAAGCGGATAAACAACAAAAACCGTGGCAAATTGCGCCC
301          N  V  W  Y  H  S  H  E  A  D  K  Q  Q  K  P  W  Q  I  A  P
961        CCGCTTCTCGAAGACGTCTACAATCTGGAAGATGCTCTTGTCGCCGGTTGCATGTTGATC
321          P  L  L  E  D  V  Y  N  L  E  D  A  L  V  A  G  C  M  L  I
1021       ACGTTGCTCCGTCACGCCGACCGTGTCAAGATCGGCTGTCTTGCGCAACTCGTCAACGTG
341          T  L  L  R  H  A  D  R  V  K  I  G  C  L  A  Q  L  V  N  V
1081       ATCGCCCCGATCATGACGAAGACGGGCGGCCCGGTCTGGAGACAGACGATTTTTTATCCG
361          I  A  P  I  M  T  K  T  G  G  P  V  W  R  Q  T  I  F  Y  P
1141       TTCCTCCACGCATCCCGCTACGGACGCGGCACCTCGCTCGTCACACTCGTCGACGGGCCG
381          F  L  H  A  S  R  Y  G  R  G  T  S  L  V  T  L  V  D  G  P
1201       AAGTACGACAGCCGCGACTACACGGACGTGCCGTATGTCGAAGCGGCGGCAGTGTACCGA
401          K  Y  D  S  R  D  Y  T  D  V  P  Y  V  E  A  A  A  V  Y  R
1261       GAAGACGTCGGTGAGTTGACCGTGTTCGCGGTCAACCGGCATTTGCACGAGCCGATCGTG
421          E  D  V  G  E  L  T  V  F  A  V  N  R  H  L  H  E  P  I  V
1321       CTGGAAGGAAAATTGGACGGTTTTGAAAATGTTCGAGTGGTCGAGCATCTTGTGCTCGAC
441          L  E  G  K  L  D  G  F  E  N  V  R  V  V  E  H  L  V  L  D
1381       CATCCCGATCTGAAAGCGGCCAACACCGTCGACGCCCCGCATCGCGTTGTGCCTAGGCGA
461          H  P  D  L  K  A  A  N  T  V  D  A  P  H  R  V  V  P  R  R
1441       CTATCGGACGCCGCCGCGGAAAACGGTCTTTTGACGGCGCAGTTGCCCGCAGCGTCCTGG
481          L  S  D  A  A  A  E  N  G  L  L  T  A  Q  L  P  A  A  S  W
1501       AACGTCATCCGCCTTGCCCGCCGACCGTCCTGA
501          N  V  I  R  L  A  R  R  P  S  *
```

Figure 3.7.2 The nucleotide sequence of AFase_E3 and the deduced amino acid sequence of the AFase protein it encodes, namely AFase_E3. The putative -10 and -35 regions of a putative promoter region are underlined and labelled. The ATG start codon is highlighted in green, while an asterix (*) and red highlighting indicates the TGA stop codon. The conserved catalytic domain is shaded in yellow, while the AFase C-terminal domain is shaded in dark blue and the overlapping portion of the two domains is shaded in turquoise.

3.2.4.3 AFase_D3 nucleotide sequence

An open reading frame (ORF), of 1482 nucleotides, within pFos_D3 was predicted to begin at the putative methionine start codon (ATG) at position 1 (M^1) and continue to the in-frame TGA stop codon at position 495 marked with the asterix (*) (Figure 3.7.1). This ORF encodes a 494 amino acid protein with a predicted molecular weight of approximately 55.791 kDa. Putative transcriptional regulatory sequences were identified in the vicinity of the -10 and -35 regions upstream of the translational start codon (Figure 3.7.3). Conserved catalytic and AFase C-terminal domains were identified within the predicted amino acid sequence (Figure 3.7.3) however; there was no evidence of a conserved signal peptide within the predicted translated amino acid sequence. Interestingly, unlike for AFase_H4 and AFase_E3, no catalytic domain was detected for AFase_D3 within the interproscan server. However the superfamily Glycoside Hydrolase domain was present and theoretically contains the Glycoside Hydrolase. This domain was still categorized as Glycoside Hydrolase family 51 and overlaps the AFase c-terminal domain. The AFase c-terminal domain is positioned at P^{285}- D^{475}.


```
                                                -35                    -10
-118      CCCAGATGGCATAAAAAATTTGCAGTTGTTTTATTCTGCCATTTGAGATTAAATAGTATT

-59       ATCTATTACGTTATTTCATTAATCAATTAATATTATTATTAATCCAAAGGAGAAGATGAA

1         ATGAACAATGTCGTCATCAATGTGGATCATTCCAGAGGAAGCATTAATCGCAACATCTAC
1          M  N  N  V  V  I  N  V  D  H  S  R  G  S  I  N  R  N  I  Y
61        GGTCACTTTGCAGAGCACCTCGGCAGATGTATATATGAAGGTATTTGGGTCGGGGAGGAT
21         G  H  F  A  E  H  L  G  R  C  I  Y  E  G  I  W  V  G  E  D
121       TCACCAATTCCGAATATACAAGGTATCCGCACTGATGTGTTAGCGGCGTTACGTCAATTG
41         S  P  I  P  N  I  Q  G  I  R  T  D  V  L  A  A  L  R  Q  L
181       AAAATTCCCGTGTTGCGCTGGCCGGGGGGCTGTTTCGCCGATGAATACCATTGGATGGAT
61         K  I  P  V  L  R  W  P  G  G  C  F  A  D  E  Y  H  W  M  D
241       GGGATTGGACCTCGTGAAACGCGAAAACGAATGGTGAATACCCATTGGGGCGGTGTAGTG
81         G  I  G  P  R  E  T  R  K  R  M  V  N  T  H  W  G  G  V  V
```

```
301         GAGAATAACCATTTCGGAACTCATGAATTTCTGCTGCTCTGTGAGTTGTTAGGCTGTGAG
101          E  N  N  H  F  G  T  H  E  F  L  L  L  C  E  L  L  G  C  E
361         CCTTATATTAACGGTAACGTTGGCAGCGGAACGGTGAAAGAGATGCAGGAGTGGATTGAA
121          P  Y  I  N  G  N  V  G  S  G  T  V  K  E  M  Q  E  W  I  E
421         TATATGACCTTCGATGGTGATTCGCCTATGGCGAATTTGCGCAGAGCCAATGGTCGCGAG
141          Y  M  T  F  D  G  D  S  P  M  A  N  L  R  R  A  N  G  R  E
481         AAGCCATGGAAGGTGAGCTACTTCGGAGTAGGTAATGAGAATTGGGGCTGCGGCGGCAAC
161          K  P  W  K  V  S  Y  F  G  V  G  N  E  N  W  G  C  G  G  N
541         ATGCGTCCAGAATACTATGCCGATCTGTACCGGCGTTATCAGACATACGTGCGAAGCTAC
181          M  R  P  E  Y  Y  A  D  L  Y  R  R  Y  Q  T  Y  V  R  S  Y
601         GGCGACAATCGTATCGCGAGAATTGCTTGCGGTGCCAGCGATTACGACACGAATTGGACT
201          G  D  N  R  I  A  R  I  A  C  G  A  S  D  Y  D  T  N  W  T
661         GAAGTGCTAATGCGTGAATCCGCAAGGTATATGGATGCACTCACCCTTCATTACTATACC
221          E  V  L  M  R  E  S  A  R  Y  M  D  A  L  T  L  H  Y  Y  T
721         ATTCCAAATACGTGGAAAGAGAAAGGCTCGGCAACACAATTCGGTGAACGGGACTGGTTC
241          I  P  N  T  W  K  E  K  G  S  A  T  Q  F  G  E  R  D  W  F
781         ATCACGCTTCGCAAGGCAATTGCCATGGACGAGCTGTTGAATAAGCATGAGACGATCATG
261          I  T  L  R  K  A  I  A  M  D  E  L  L  N  K  H  E  T  I  M
841         AATAAATACGATCCAGATAAGCGGGTAGCGCTAATCGTTGACGAATGGGGTACTTGGTTC
281          N  K  Y  D  P  D  K  R  V  A  L  I  V  D  E  W  G  T  W  F
901         GATGTCGAGCCAGGGACGAACCCAGGCTTTCTGTATCAGCAAAGTACGATGCGTGATGCC
301          D  V  E  P  G  T  N  P  G  F  L  Y  Q  Q  S  T  M  R  D  A
961         TTGGTTGCCGCTTTAACGCTGCATATTTTCCACGGTCATTGCGACCGCGTACAGATGGCC
321          L  V  A  A  L  T  L  H  I  F  H  G  H  C  D  R  V  Q  M  A
1021        AACATCGCGCAGACCGTTAACGTGCTCCAATCCGTTGTATTGACGGAAGGCGCAGCAATG
341          N  I  A  Q  T  V  N  V  L  Q  S  V  V  L  T  E  G  A  A  M
1081        ACACTTACCCCGACCTATCATGTGTTCGATATGTTTAAGGTACATCAGGATGCGGAAGCG
361          T  L  T  P  T  Y  H  V  F  D  M  F  K  V  H  Q  D  A  E  A
1141        CTCGATGTATTCACCCAGTCTGAATGCTACGAGATGGATGGTGAGTCGATTCCCCAGCTA
381          L  D  V  F  T  Q  S  E  C  Y  E  M  D  G  E  S  I  P  Q  I
1201        AGTGTCTCGGCTTCCCGCAATTCAGATGGGATCATCCATATCAGCCTGTGTAACGTTCAT
401          S  V  S  A  S  R  N  S  D  G  I  I  H  I  S  L  C  N  V  H
1261        CATGACCGCGAGGCTGTATTGACAATTGATCTAAGAGGCACTGATGGAAGCGATCAAGTC
421          H  D  R  E  A  V  L  T  I  D  L  R  G  T  D  G  S  D  Q  V
1321        ATAGAGGGAACCGTTCTTCGTGCTTCGGAGCTCAATGCTCACAATACCTTCACTCAACCG
441          I  E  G  T  V  L  R  A  S  E  L  N  A  H  N  T  F  T  Q  P
1381        AATACCGTGCAGCCCACTAGTCTCACCAGCATCGTTCGCAAGGATGACGGTACGATCACG
461          N  T  V  Q  P  T  S  L  T  S  I  V  R  K  D  D  G  T  I  T
1441        GTTACGATGCCTCCGGCTTCAGTCGGCATTCTAAGATTAGGTTGA
481          V  T  M  P  P  A  S  V  G  I  L  R  L  G  *
```

Figure 3.7.3 The nucleotide sequence of AFase_D3 and the deduced amino acid sequence of the AFase protein it encodes, namely AFase_D3. The putative -10 and -35 regions of a putative promoter region are underlined and labelled. The ATG start codon is highlighted in green, while an asterix (*) and red highlighting indicates the TGA stop codon. The conserved catalytic domain is shaded in

yellow, while the AFase C-terminal domain is shaded in dark blue and the overlapping portion of the two domains is shaded in turquoise.

3.3 Gene expression and protein purification

3.3.1 Cloning of AFase genes for recombinant protein expression

The ORFs of the three AFase proteins were amplified from the respective fosmid clones by PCR (Figure 3.10). The oligonucleotide primers were designed to contain restriction enzyme recognition sequences for the subsequent in-frame cloning of these genes into the pET21a(+) expression vector. Following PCR amplification, *AFase_H4*, *AFase_E3* and *AFase_D3* were successfully sub-cloned into the pJET1.2/blunt cloning vector to create the recombinant constructs pJET1.2*_AFase_H4*, pJET1.2*_AFase_D4* and pJET1.2*_AFase_E3*. The cloned inserts were sequenced to confirm that PCR errors were not introduced. The cloned genes were subsequently excised from the three recombinant pJET*_AFase* constructs (Figure 3.11) and directionally sub-cloned into pET21a(+) vector that had been restricted with the same restriction enzymes. This resulted in the successful construction of the three recombinant constructs, pET21a(+)*_AFase_H4*, pET21a(+)*_AFase_E3* and pET21a(+)*_AFase_D3*. These constructs were employed, as described in the following sections, for heterologous expression and purification of AFase_H4, AFase_E3 and AFase_D3.

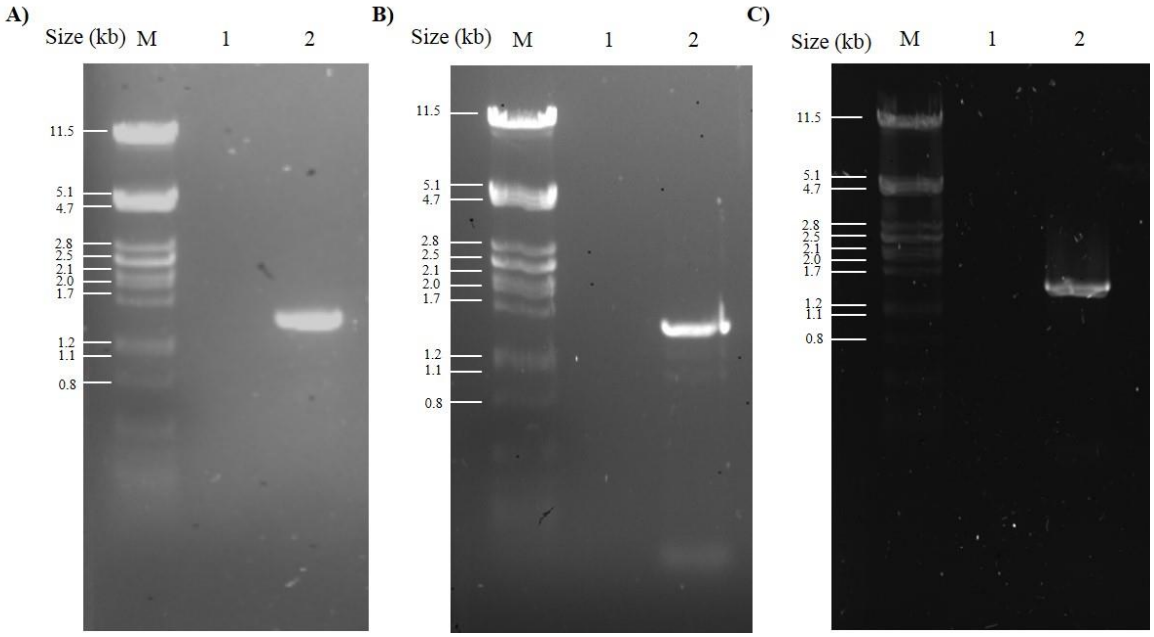


Figure 3.10 PCR amplification of AFase_H4 (A), AFase_E3 (B) and AFase_D3 (C). Amplified single PCR products were separated by electrophoresis in a 0.8% (w/v) TAE agarose gel. Lane M, λ*Pst*I molecular weight marker; lane 1, No template control; and Lane 2, Fosmid amplified AFase genes. The sizes of the DNA fragments of the molecular weight marker are shown. The expected PCR amplicon sizes for AFase_H4, AFase_E3 and AFase_D3 were approximately 1.467, 1.547 and 1.482 kb, respectively.

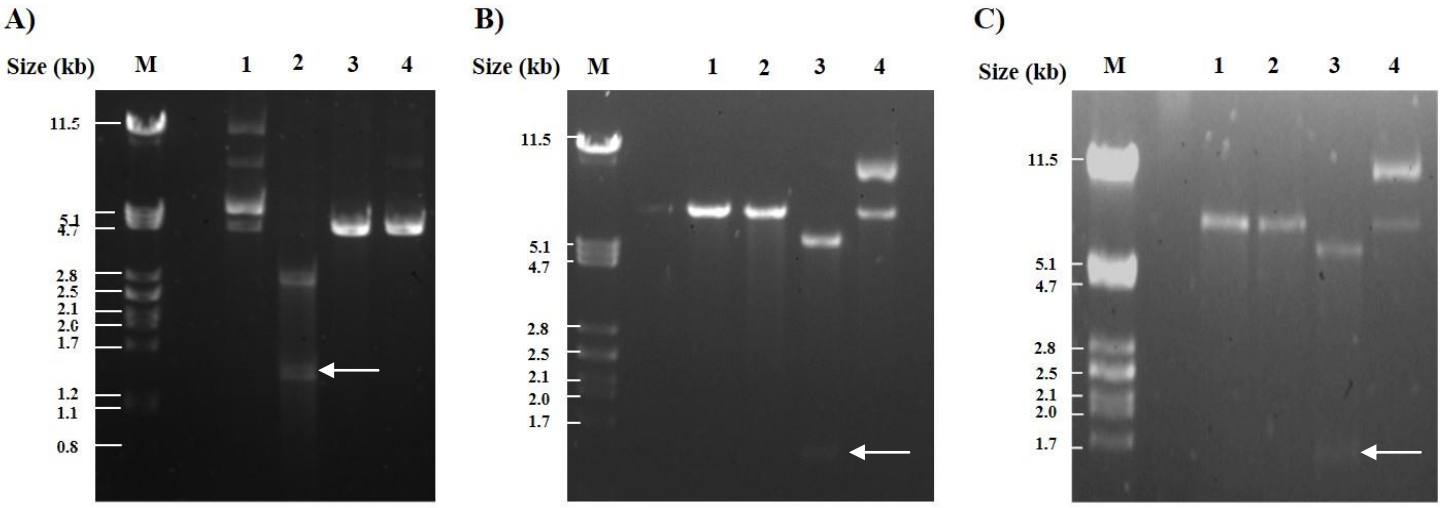


Figure 3.11 Agarose gel electrophoresis of pJET1.2/blunt and recombinant pJET1.2/blunt constructs, pJET1.2_*AFase_H4* (A), pJET1.2_*AFase_D4* (B) and pJET1.2_*AFase_E3* (C), digested with restriction enzymes to release the respective AFase gene fragments indicated by the arrow. A) Lane M, λPstI molecular weight marker; lane 1, undigested pJET1.2_*AFase_D3*; lane 2, pJET1.2_*AFase_D3* digested with *Xho*I and *Nde*I; lane 3, pJET1.2_*AFase_D3* digested with *Xho*I; and lane 4, pJET1.2_*AFase_D3* digested with *Nde*I. B) Lane M, λPstI molecular weight marker; lane 1, pJET1.2_*AFase_H4* digested with *Nde*I; lane 2, pJET1.2_*AFase_H4* digested with *Xho*I; lane 3, pJET1.2_*AFase_H4* digested with *Xho*I and *Nde*I; and lane 4, undigested pJET1.2_*AFase_H4*. C) Lane M, λPstI molecular weight marker; lane 1, pJET1.2_*AFase_E3* digested with *Nde*I; lane 2, pJET1.2_*AFase_E3* digested with *Xho*I; lane 3, pJET1.2_*AFase_E3* digested with *Xho*I and *Nde*I; and lane 4, undigested pJET1.2_*AFase_E3*. The relative sizes, in kb, of the λ*Pst*I MW markers are indicated.

3.3.2 Expression of AFase genes in E. coli, purification of proteins and FPLC analyses

All three AFases were successfully expressed in *E. coli* as enzyme activity, confirmed by the release of *p*NP from the pNP-α-L-arabinofuranoside (*p*NP-ara substrate), was detected in cell free extracts (Figure 3.12).

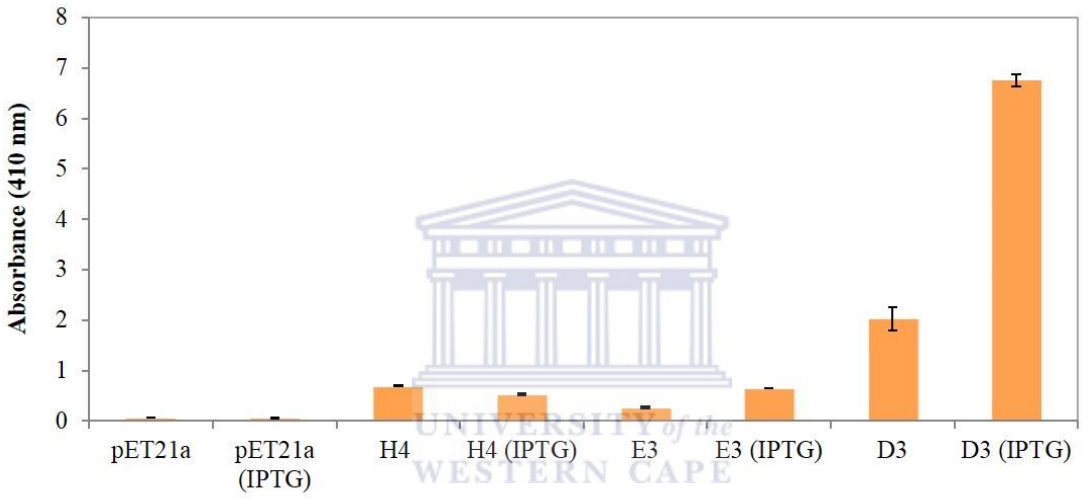


Figure 3.12 The effect of induction with 0.5 mM IPTG on AFase activity of E. coli BL21 (DE3) harbouring pET21a(+), pET21a_AFase_H4 (H4), pET21a_AFase_E3 (E3) and pET21a_AFase_D3 (D3). The cultures were either uninduced or induced (IPTG) and the cell-free extracts subsequently assayed for AFase activity using pNP-α-L-arabinofuranoside as the substrate at pH 7.0 and incubated at 37°C. The liberation of pNP was monitored by spectrophotometric measurements at 410 nm. Data represents the average of three replicates ± Standard error (*n*=3).

Optimisation of IPTG concentrations to ensure sufficient gene expression and subsequent protein production was performed. Induction with 0.5 or 1 mM Isopropyl-β-D-1-thiogalactopyranoside (IPTG) resulted in the production of a visible quantity of AFase_D3 protein as shown on Figure 3.12 and Figure 3.13. Protein levels achieved for AFase_H4 and

AFase_E3 were lower than those observed for AFase_D3 in the soluble crude fractions. Proteins, AFase_H4, AFase_E3 and AFase_D3, were purified from these cells using Ni^{2+} affinity chromatography.

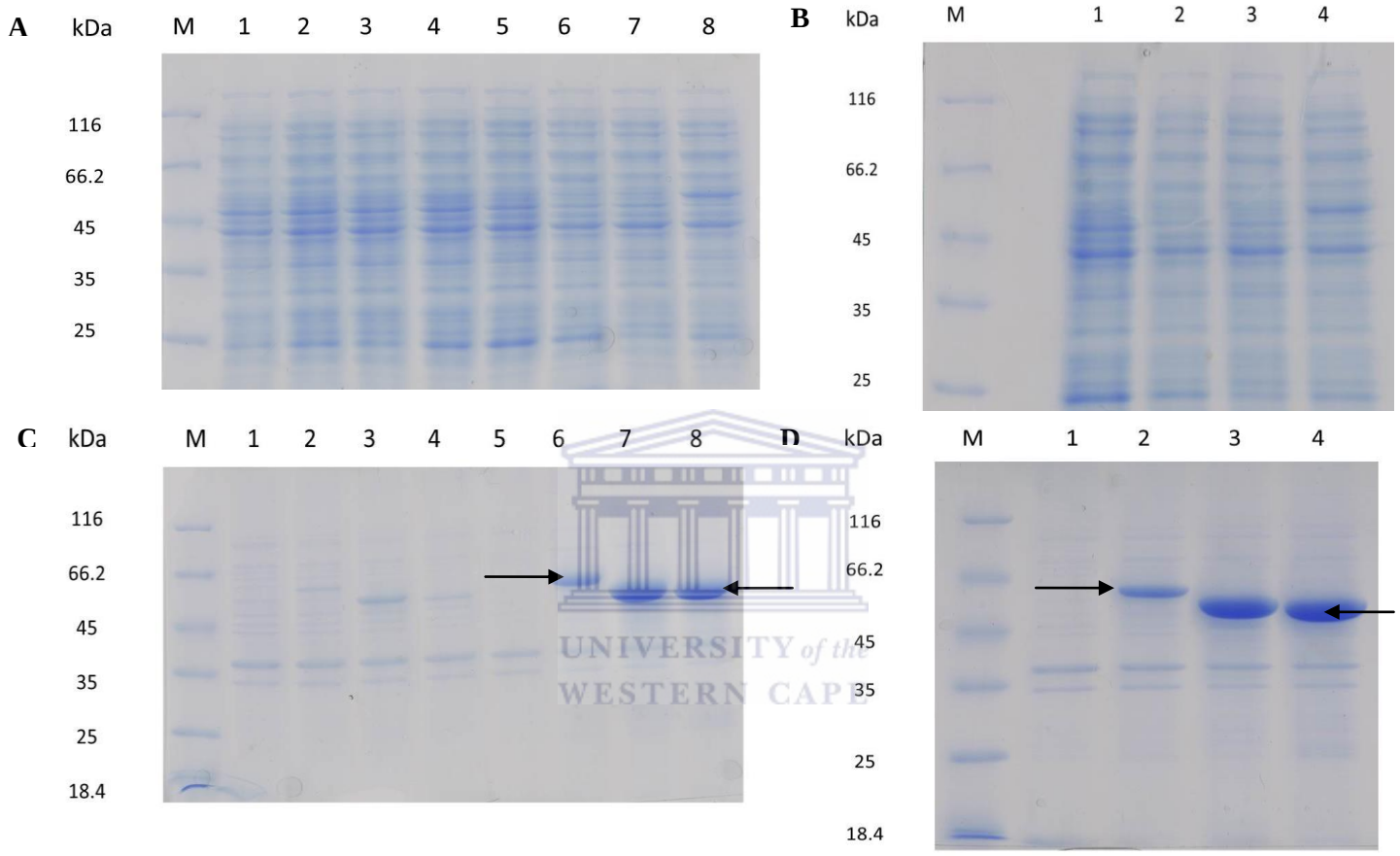
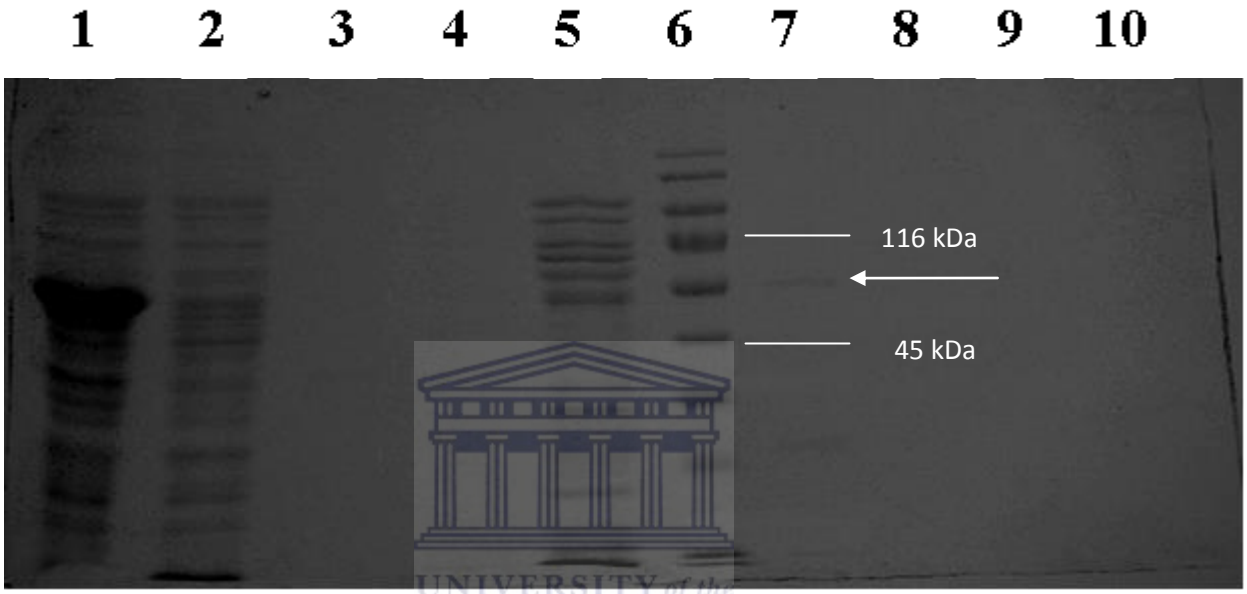


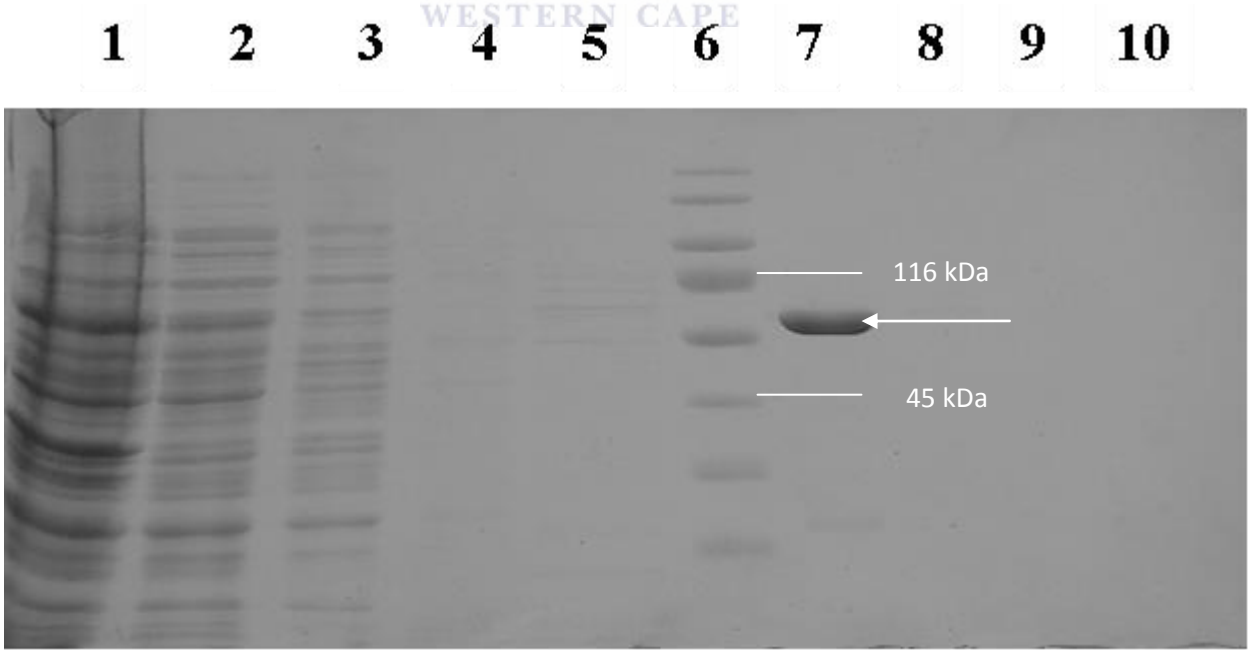
Figure 3.13 Analysis of proteins in the soluble and insoluble cell fractions of E. coli BL21 cells harbouring pET21a_AFase_H4, pET21a_AFase_E3 and pET21a_AFase_D3 constructs by 12% (w/v) SDS-PAGE. A) The soluble proteins obtained following induction of the recombinant *E. coli* BL21 cells with either 0.5 mM (Lanes 1 to 4) or 1 mM (Lanes 5 to 8) IPTG. Lanes contain cellular extracts prepared from *E. coli* BL21 cells harbouring: pET21a(+) vector only (Lanes 1and 5); pET21a_*AFase_E3* (Lanes 2 and 6), pET21a_*AFase_H4* (Lane 3 and 7), pET21a_*AFase_D3* (Lane 4 and 8). B) The soluble proteins from the *E.coli* BL21 cells Lane 1-4: pET21a-1mM IPTG, pET21a_*AFase_E3*-1mMIPTG,pET21a_*AFase_H4*-1mMIPTG, pET21a_*AFase_D3*-1mMIPTG. C) and D) are presented in the same orientation with the insoluble protein fractions. The position of the AFase proteins is located between the arrows.

AFase_H4, AFase_E3 and AFase_D3 proteins were purified to near homogeneity using nickel-ion affinity chromatography. Representative samples were taken when different buffers were added to the His affinity resin to track the removal of *E. coli* proteins and ensure the continued binding of the proteins of interest (Figure 4.13)

A)



B)



C)

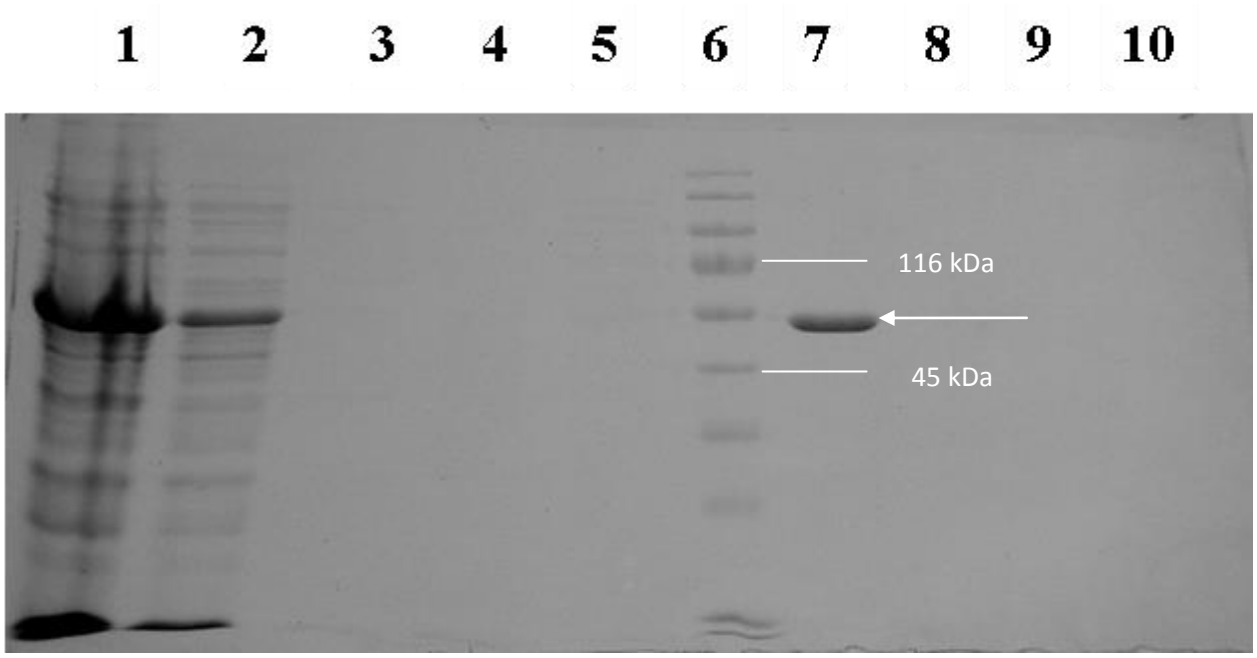


Figure 3.14 12% polyacrylamide matrix with the filtering and purification steps as well as the 250 mM imidazole eluted purified protein to apparent homogeneity. (A-C): Lane 1: Insoluble protein fractions, Lane 2: Soluble protein fractions, Lane 3: Soluble protein column flow through, Lane 4: binding buffer effluent, lane 5: Wash buffer effluent, lane 6: Stained protein marker with indicate sizes of 55 kDa and 70 kDA, lane 7: AFase_H4 – 54 kDa, AFase_E3 – 57.5 kDa and AFase_D3 – 55.7 kDa; respectively. Lane 8-9: 500 mM and 1 M imidazole elution. Lane 10: Strip buffer elute. All proteins have been quantified using the Bradford reagent method (Bradford, 1976).

These results suggest that the proteins of interest, AFase_H4, AFase_E3 and AFase_D3, remained bound to the His-affinity resin throughout the washing steps and that 250 mM imidazole was sufficient to elute the proteins from the column. Purified proteins contained only a small proportion of contaminating *E. coli* proteins as can be seen from the presence of faint bands additional to the proteins of interest.

Purified protein extracts were subjected to gel filtration analysis to determine the quaternary structure of the native AFase_H4, AFase_E3 and AFase_D3 proteins. Chromatograms are

shown in the Appendices (Figure. AA, AB and AC; Table A1.). The respective retention times of the three purified AFases, relative to that of molecular weight standards suggest that all three proteins are either homo-hexamers or dimers of trimers. Hexameric structures for AFase proteins have been reported, specifically for the GH 51 family of AFases (Hoffman *et al.*, 2013), however other quaternary forms have been reported such as tetramers and octamers from the same family of AFases. Homotetrameric structures have been reported for AFases purified from a *Streptomyces spp*, *Geobacillus caldoxylolyticus* TK4 and *Anoxybacillus kestanbolensis* AC26Sari (Shi *et al.*, 2010; Canakci *et al.*, 2007; Canakci *et al.*, 2008). A homo-octameric structure has also been reported for an AFase from *Caldicellulosiruptor saccharolyticus* (Lim *et al*., 2010). The hexameric structures of the three purified AFases, AFase_H4, AFase_E3 and AFase_D3, are consistent with crystalized AFases from *Thermotoga maritime* (**PDB ID**: 2VRQ), *Thermotoga petrophila* (**PDB ID**: 3S2C), *Geobacillus stearothermophilus* (**PDB ID**: 1PZ3), *Thermobacillus xylanilyticus* (**PDB ID**: 2VRK) and *Clostridium thermocellum* (**PDB ID**: 2C7F). Each monomer of the homo-hexamer consists of a $(\alpha/\beta)_8$ barrel catalytic domain and a 12 stranded β-sandwich domain (Im *et al.*, 2012; Dumbrepatil *et al.*, 2012; Hovel *et al*., 2003; Paes *et al.*, 2008; Souza *et al.*, 2011; Taylor *et al.*, 2006).

3.4 Biochemical characterization of AFases

3.4.1 pH and temperature optima of AFases

All three AFases displayed optimum activity within the acidic pH range of 4.0-5.0. Figure 3.15 displays the pH optima activity for each purified AFase .These results are similar to other biochemically characterized AFases belonging to the GH 51 family and isolated from

Paenibacillus sp, *Aureobasidium pullulans*, *Aspergillus oryzae* and *Streptomyces coelicolor* (Shi *et al.*, 2013; Ohta *et al.*, 2013). Specifically AFase_E3 and AFase_D3 function over a broad pH optimum range from pH 4.0 – pH 5.5 and pH 4.5 – pH 6.0, respectively. Similar results were observed by AFases characterized in a study performed by Yan *et al*, (2012) and Shi *et al*, (2013) in which the pH optima for AFases isolated from *Chaetomium sp* and *Paenibacillus sp*, displayed wider pH optima compared to AFase_E3 and AFase_D3.

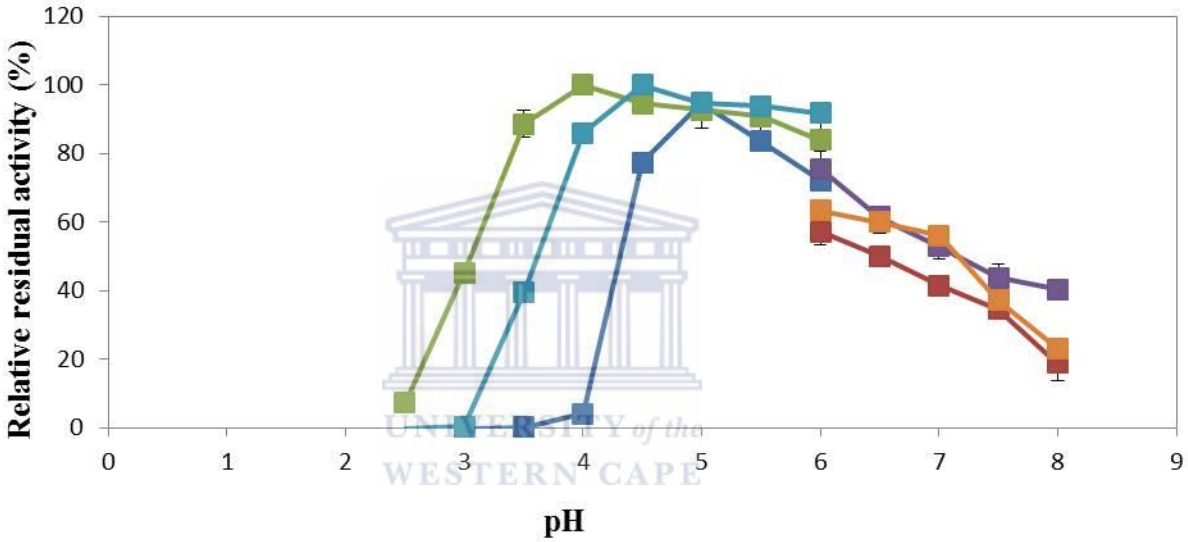


Figure 3.15 pH optima of AFase_H4 ,AFase_E3 and AFase_D3 were determined by performing *p*NP-α-L-arabinofuranoside assays in phosphate and citrate buffers between pH 2.5 and 6.0 and pH 6.0 and 8.0, respectively. The assays were performed at 37°C for 2 minutes. The liberated *p*NP was measured spectrophotometrically at 410 nm (Spectrostar nano,BMG Labtech). Data represents the mean ± standard error (*n*=3).

The temperature optimum for the purified AFases is indicated by Figure 3.16. AFase_E3 and AFase_H4 displayed optimum activity at 60°C. AFases that have been characterized in this similar protocol, as depicted in Figure 3.16, have been confirmed to be thermophilic AFases (Nurcholis *et al.*, 2012; Damasio *et al.*, 2012). It has been shown that thermophilic AFases' amino acid sequence with its catalytic domain categorizes them generally into the GH51

family only (Fritz *et al.*, 2008; Dumbrepatil *et al.*, 2012) but AFase_D3 displayed a temperature optimum at 25°C, however remained active between 80 and 90°C retaining approx. 50% of its maximum activity (Figure 3.16). To the best of our knowledge, this is the first reported mesophilic AFase with this temperature optimum activity characteristic and that can be considered thermotolerant, a phenomenon that can occur in enzymes as demonstrated by Kota *et al*. (2010).

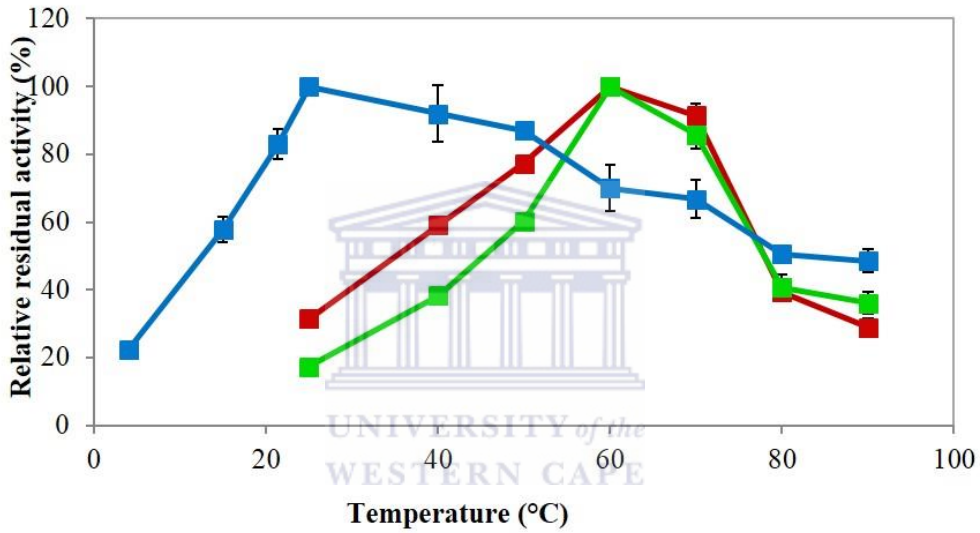


Figure 3.16 The temperature optima of AFase_H4 (■), AFase_E3 (■) and AFase_D3 (■) using the substrate p-nitrophenyl-α-L-arabinofuranoside at their respective pH optima 5.0, 4.0 and 4.5. The liberated p-Nitrophenol was measured at an OD_{410nm} (Spectrostarnano, BMG Labtech). Data represents the mean ± standard error (*n*=3).

3.4.2. Thermostability of AFase_H4, AFase_E3 and AFase_D3

The thermostability of the AFase proteins were investigated and in comparison to the results obtained from the initial thermostability tests on crude extracts from the fosmid clones (Figure 3.1), these three purified AFases displayed differing thermostability characteristics

(Figure 3.17). AFase_D3 maintained 100% residual activity following an hour incubation at 40°C and maintained up to 50% of its residual activity until incubation until 50°C. AFase_H4 was more thermostable than AFase_D3 and maintained 100% residual activity after 12 hour incubation at 50°C and inhibition of its residual activity occurred when incubated at 60°C. AFase_E3 displayed the greatest thermostability and maintained nearly 60% of its residual hydrolytic activity after 24 hour incubation at 70°C. AFase_H4 and AFase_D3 showed lower thermostability characteristics during the 24 hour temperature incubations (Figure 3.17). Similar thermolabile characteristics, such as that observed for AFase_D3, have been reported from AFases isolated from a *Paenabacillus sp* (Wagschal *et al.*, 2008) and *Geobacillus thermoleovorans* (Wagschal *et al.*, 2009). However, these AFases belong to the GH43 family. Currently there is no characterized GH51 family AFase that displays a similar thermolabile profile as AFase_D3. AFase_H4 and AFase_E3 thermostability profiles are both similar to the functionally characterized AFases cloned and purified from *Streptomyces sp*.S9 (Shi *et al.*, 2010) and *Geobacillus caldoxylolyticus* TK4 (Canakci *et al.*, 2007). The different thermostable characteristics of AFase_H4, AFase_E3 and AFase_D3 can be correlated according to their amino acid composition located within their respective catalytic domains.

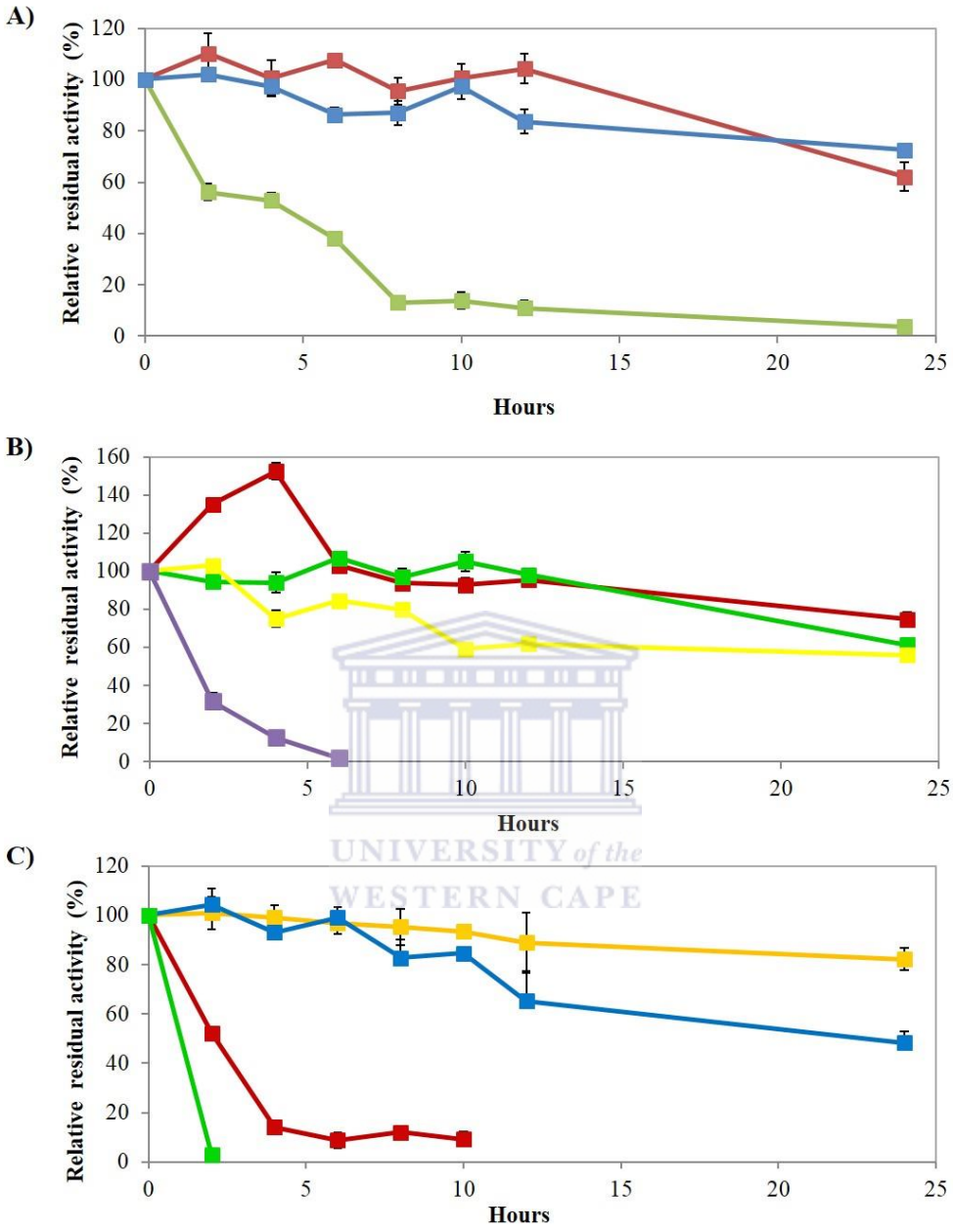


Figure 3.17 The thermostablity profiles for the purified AFase, AFase_H4 (A), AFase_E3 (B) and AFase_D3 (C). AFase activity was detected using 2 mM *p*-nitrophenyl-α-L-arabinofuranoside and the release of *p*-nitrophenol, by enzymatic hydrolysis, measured spectrophotometrically at 410 nm following incubation at 25 (■), 40 (■), 50 (■), 60 (■), 70 (■) and 80°C (■) for 1 hour. The percentage residual activity was calculated from a triplicate protein sample that had been incubated on ice for the duration of the experiment. Data represents the average of three replicates ± Standard error (*n*=3).

Several studies have correlated thermostability to specific amino acids (Hovel *et al*., 2003; Taylor *et al.*, 2006; Paes *et al.*, 2008; Souza *et al*., 2011), and specifically for the hexameric GH51 AFases (Table 3.5; Souza *et al* 2011). These amino acid residues are involved with thermostabilizing each tertiary monomer of the globular hexameric protein structure. These amino acid positions and their prevalence are deduced in the amino acid sequences of AFase_H4, AFase_E3 and AFase_D3 are given in Table 3.5.

Table 3.5. The presence of amino acid residues within the AFase_H4, AFase_E3 and AFase_D3 protein previously reported to be involved in thermostabilizing the hexameric structure of GH51 AFases (Souza *et al.*, 2011).

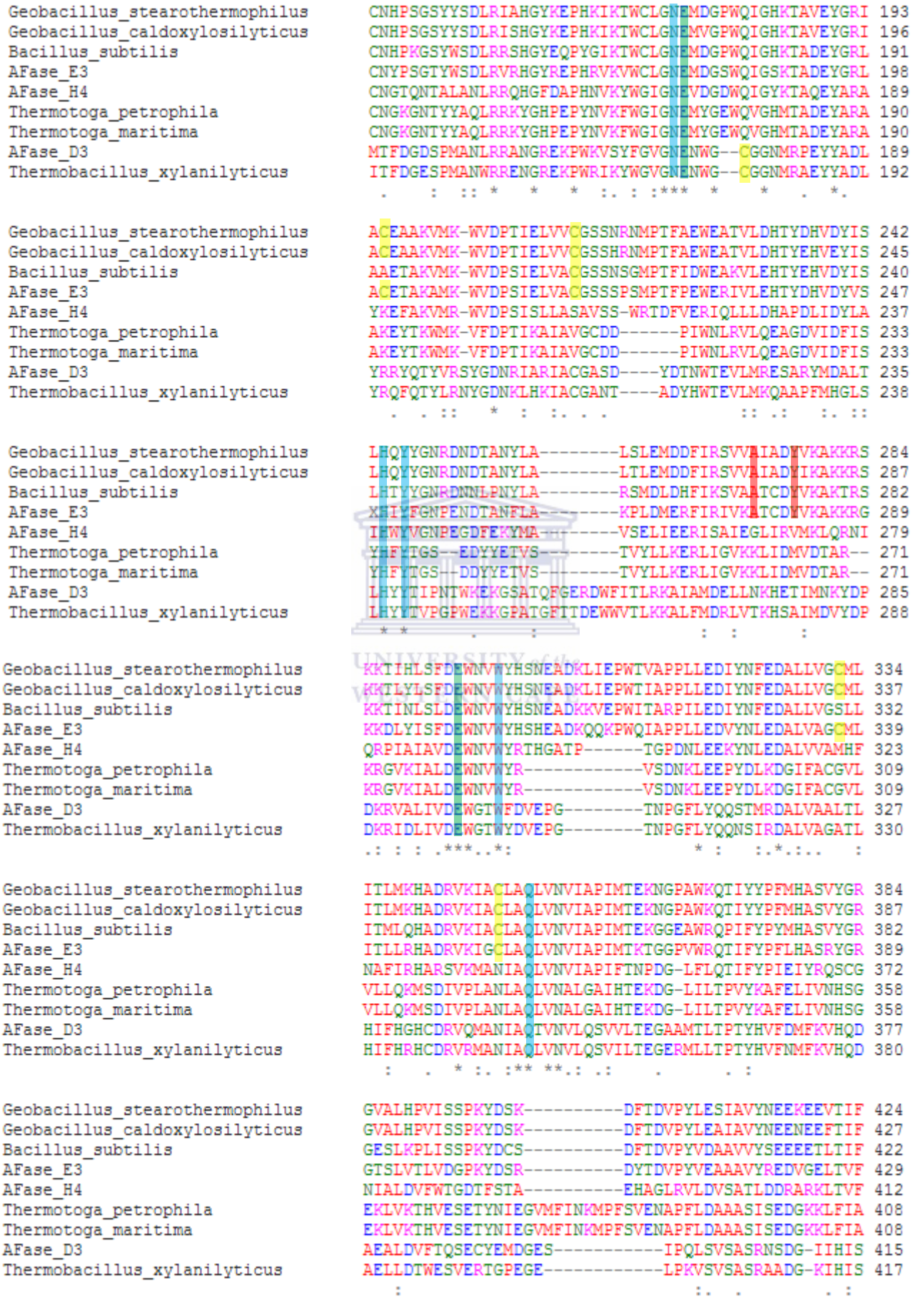
AFase_H4	AFase_E3	AFase_D3	Amino acid
		✓	H77
✓	✓	✓	E99
	✓		D132
✓			N146
	✓		Y148
			E176
			H181
✓	✓		E186
✓	✓		R189
✓			K192
✓	✓		E193
✓	✓		K196
✓			E255
	✓		K261
		✓	K262
			K363

The total number of thermostabilizing amino acids present in each of the AFases appears to correlate with the degree of thermostability measured experimentally. AFase_D3 is the least thermostable and only contains three of the 16 previously hypothesised amino acids, whereas, AFase_H4 and AFase_E3 contain eight each. AFase_E3 was the most thermally stable

enzyme reported in this study, and although it contains the same number of thermostabilizing amino acids as AFase_H4, the amino acids are at different positions within the peptide. However, there are other factors such as hydrophobic core, increased rigidity, increased hydrogen bonding, salt bridges and polar surface area that affect thermostability of an enzyme or protein (Taylor and Viasmen, 2010) and these factors may also be important in conveying thermostability to the AFases isolated in this study.

AFase_E3 is the most thermostable amongst the three AFases of this study and this possible observation can be supported based on sequence evidence which compliments the thermostabilizing amino acids. Figure 3.18 indicated the Cysteine residues that are present in AFase_E3 that have been correlated with increasing thermostability within certain AFases of the GH51 family, where these specific cysteine residues formed disulphide bonds with alternate cysteine residues within the catalytic domain (Nurcholis *et al.*, 2012). These cysteine residues (highlighted in yellow in Figure 3.18), that are not present within AFase_H4, have been experimentally validated to increase thermostability by 10 to 20^{o}C higher (Nurcholis *et al.*, 2012; Zhang *et al.*, 2011). Through site-directed mutagenesis (alanine for cysteine substitutions) a decrease in thermostability occurred within the AFase cloned and purified from *Geobacillus cladoxylolyticus, Geobacillus stearothermophilus, Thermobacillus xylanilyticus and Bacillus subtilis* (Figure 3.18) (Canakci *et al.*, 2007; Paes *et al.*, 2008; Nurcholis *et al.*, 2012). It is possible that AFase_E3's higher thermostability compared to AFase_H4 and AFase_D3 could be caused by these specific cysteine residues that forms disulphide bonds with alternate cysteine residues within the catalytic domain. A study performed by Hoffman *et al.* (2013) indicated that the residues (within the AFase sequence from *Bacillus subtilis, Geobacillus stearothermphilus, Geobacillus*

cladoxylolyticus, Thermotoga petrophila and *Thermotoga maritime,* highlighted in red in Figure 3.18) could be involved in oligomeric interactions of the individual monomers of the enzyme, thereby stabilizing thethe enzyme under higher temperatures. These residues are not encoded within the sequence of the AFase_H4 and AFase_D3, but are present within the sequence of AFase_E3 as depicted in Figure 3.18. AFase_D3 possesses cysteine residues that have been mutated to alanine within an AFase purified from *Thermobacillus xylanilyticus* and this site-directed mutagenesis has shown to decrease the enzymes' thermostability (Paes *et al.*, 2008). It can also be noted that AFase_D3 contains a higher quantity of cysteine residues than AFase_H4 but AFase_D3 is omitted of the polar residues involved in thermostabilizing its hexameric structure which therefore could account for its lower thermostability (Taylor and Viasmen, 2010).


Geobacillus_stearothermophilus CNHPSGSYYSDLRIAHGYKEPHKIKTWCLGNEMDGPWQIGHKTAVEYGRI 193
Geobacillus_caldoxylosilyticus CNHPSGSYYSDLRISHGYKEPHKIKTWCLGNEMVGPWQIGHKTAVEYGRI 196
Bacillus_subtilis CNHPKGSYWSDLRRSHGYEQPYGIKTWCLGNEMDGPWQIGHKTADEYGRL 191
AFase_E3 CNYPSGTYWSDLRVRHGYREPHRVKVWCLGNEMDGSWQIGSKTADEYGRL 198
AFase_H4 CNGTQNTALANLRRQHGFDAPHNVKYWGIGNEVDGDWQIGYKTAQEYARA 189
Thermotoga_petrophila CNGKGNTYYAQLRRKYGHPEPYNVKFWGIGNEMYGEWQVGHMTADEYARA 190
Thermotoga_maritima CNGKGNTYYAQLRRKYGHPEPYNVKFWGIGNEMYGEWQVGHMTADEYARA 190
AFase_D3 MTFDGDSPMANLRRANGREKPWKVSYFGVGNENWG--CGGNMRPEYYADL 189
Thermobacillus_xylanilyticus ITFDGESPMANWRRENGREKPWRIKYWGVGNENWG--CGGNMRAEYYADL 192
Geobacillus_stearothermophilus ACEAAKVMK-WVDPTIELVVCGSSNRNMPTFAEWEATVLDHTYDHVDYIS 242
Geobacillus_caldoxylosilyticus ACEAAKVMK-WVDPTIELVVCGSSHRNMPTFAEWEATVLDHTYEHVEYIS 245
Bacillus_subtilis AAETAKVMK-WVDPSIELVACGSSNSGMPTFIDWEAKVLEHTYEHVDYIS 240
AFase_E3 ACETAKAMK-WVDPSIELVACGSSSPSMPTFPEWERIVLEHTYDHVDYVS 247
AFase_H4 YKEFAKVMR-WVDPSISLLASAVSS-WRTDFVERIQLLLDHAPDLIDYLA 237
Thermotoga_petrophila AKEYTKWMK-VFDPTIKAIAVGCDD------PIWNLRVLQEAGDVIDFIS 233
Thermotoga_maritima AKEYTKWMK-VFDPTIKAIAVGCDD------PIWNLRVLQEAGDVIDFIS 233
AFase_D3 YRRYQTYVRSYGDNRIARIACGASD----YDTNWTEVLMRESARYMDALT 235
Thermobacillus_xylanilyticus YRQFQTYLRNYGDNKLHKIACGANT----ADYHWTEVLMKQAAPFMHGLS 238
Geobacillus_stearothermophilus LHQYYGNRDNDTANYLA--------LSLEMDDFIRSVVAIADYVKAKKRS 284
Geobacillus_caldoxylosilyticus LHQYYGNRDNDTANYLA--------LTLEMDDFIRSVVAIADYIKAKKRS 287
Bacillus_subtilis LHTYYGNRDNNLPNYLA--------RSMDLDHFIKSVAATCDYVKAKTRS 282
AFase_E3 XHIYFGNPENDTANFLA--------KPLDMERFIRIVKATCDYVKAKKRG 289
AFase_H4 IEWYVGNPEGDFEKYMA---------VSELIEERISAIEGLIRVMKLQRNI 279
Thermotoga_petrophila YHFYTGS--EDYYETVS--------TVYLLKERLIGVKKLIDMVDTAR-- 271
Thermotoga_maritima YHFYTGS--DDYYETVS--------TVYLLKERLIGVKKLIDMVDTAR-- 271
AFase_D3 LHYYTIPNTWKEKGSATQFGERDWFITLRKAIAMDELLNKHETIMNKYDP 285
Thermobacillus_xylanilyticus LHYYTVPGPWEKKGPATGFTTDEWWVTLKKALFMDRLVTKHSAIMDVYDP 288
Geobacillus_stearothermophilus KKTIHLSFDEWNVWYHSNEADKLIEPWTVAPPLLEDIYNFEDALLVGCML 334
Geobacillus_caldoxylosilyticus KKTLYLSFDEWNVWYHSNEADKLIEPWTIAPPLLEDIYNFEDALLVGCML 337
Bacillus_subtilis KKTINLSLDEWNVWYHSNEADKKVEPWITARPILEDIYNFEDALLVGSLL 332
AFase_E3 KKDLYISFDEWNVWYHSHEADKQQKPWQIAPPLLEDVYNLEDALVAGCML 339
AFase_H4 QRPIAIAVDEWNVWYRTHGATP------TGPDNLEEKYNLEDALVVAMHF 323
Thermotoga_petrophila KRGVKIALDEWNVWYR----------VSDNKLEEPYDLKDGIFACGVL 309
Thermotoga_maritima KRGVKIALDEWNVWYR----------VSDNKLEEPYDLKDGIFACGVL 309
AFase_D3 DKRVALIVDEWGTWFDVEPG--------TNPGFLYQQSTMRDALVAALTL 327
Thermobacillus_xylanilyticus DKRIDLIVDEWGTWYDVEPG--------TNPGFLYQQNSIRDALVAGATL 330
Geobacillus_stearothermophilus ITLMKHADRVKIACLAQLVNVIAPIMTEKNGPAWKQTIYYPFMHASVYGR 384
Geobacillus_caldoxylosilyticus ITLMKHADRVKIACLAQLVNVIAPIMTEKNGPAWKQTIYYPFMHASVYGR 387
Bacillus_subtilis ITMLQHADRVKIACLAQLVNVIAPIMTEKGGEAWRQPIFYPYMHASVYGR 382
AFase_E3 ITLLRHADRVKIGCLAQLVNVIAPIMTKTGGPVWRQTIFYPFLHASRYGR 389
AFase_H4 NAFIRHARSVKMANIAQLVNVIAPIFTNPDG-LFLQTIFYPIEIYRQSCG 372
Thermotoga_petrophila VLLQKMSDIVPLANLAQLVNALGAIHTEKDG-LILTPVYKAFELIVNHSG 358
Thermotoga_maritima VLLQKMSDIVPLANLAQLVNALGAIHTEKDG-LILTPVYKAFELIVNHSG 358
AFase_D3 HIFHGHCDRVQMANIAQTVNVLQSVVLTEGAAMTLTPTYHVFDMFKVHQD 377
Thermobacillus_xylanilyticus HIFHRHCDRVRMANIAQLVNVLQSVILTEGERMLLTPTYHVFNMFKVHQD 380
Geobacillus_stearothermophilus GVALHPVISSPKYDSK----------DFTDVPYLESIAVYNEEKEEVTIF 424
Geobacillus_caldoxylosilyticus GVALHPVISSPKYDSK----------DFTDVPYLEAIAVYNEENEEFTIF 427
Bacillus_subtilis GESLKPLISSPKYDCS----------DFTDVPYVDAAVVYSEEEETLTIF 422
AFase_E3 GTSLVTLVDGPKYDSR----------DYTDVPYVEAAAVYREDVGELTVF 429
AFase_H4 NIALDVFWTGDTFSTA----------EHAGLRVLDVSATLDDRARKLTVF 412
Thermotoga_petrophila EKLVKTHVESETYNIEGVMFINKMPFSVENAPFLDAAASISEDGKKLFIA 408
Thermotoga_maritima EKLVKTHVESETYNIEGVMFINKMPFSVENAPFLDAAASISEDGKKLFIA 408
AFase_D3 AEALDVFTQSECYEMDGES----------IPQLSVSASRNSDG-IIHIS 415
Thermobacillus_xylanilyticus AELLDTWESVERTGPEGE----------LPKVSVSASRAADG-KIHIS 417

```
Geobacillus_stearothermophilus      AVNRDMEDALLLECDVRSFEDY-RVIEHIVLEHDNVKQTNSAQSSP-VVP 472
Geobacillus_caldoxylosilyticus      AVNRDLEDALLLECDIRNFDGY-QVIEHIILEHENVKQTNSATHSP-AVP 475
Bacillus_subtilis                   AVN-KAEDQMETEISLRGFESY-QIAEHIVLEHQDIKATNQHNRKN-VVP 469
AFase_E3                            AVNRHLHEPIVLEGKLDGFENV-RVVEHLVLDHPDLKAANTVDAPHRVVP 478
AFase_H4                            VVNRSQTDEMETTITLDGGQFA-GTGQAWVVNGPDIKAENSFDAPDRVSA 461
Thermotoga_petrophila               VVNYRKEDALKVPIRVEG-LGQ-KKATVYTLTGPDVNARNTMENPN--VV 454
Thermotoga_maritima                 VVNYRKEDALKVPIRVEG-LGQ-KKATVYTLTGPDVNARNTMENPN--VV 454
AFase_D3                            LCNVHHDREAVLTIDLRGTDGSDQVIEGTVLRASELNAHNTFTQPNTVQP 465
Thermobacillus_xylanilyticus        LCNLDFETGASVDIELRGLNGG-VSATGTTLTSGRIDGHNTFDEPERVKP 466
                                      *           : .           :    :.  *

Geobacillus_stearothermophilus      HRNGDAQLSD-RKVSATLPKLSWNVIRLGKR-- 502
Geobacillus_caldoxylosilyticus      HSNGNAHLSD-GKVVAQLPKLSWNVIRLAKK-- 505
Bacillus_subtilis                   HSNGSSSVSE-NGLTAHFTPLSWNVIRLKKQS- 500
AFase_E3                            RRLSDAAAEN-GLLTAQLPAASWNVIRLARRPS 510
AFase_H4                            TR---SAVTA-GLTSATYTFEPHSVTALSFDL- 489
Thermotoga_petrophila               DITSETITVD-TEFEHTFKPFSCSVIEVELE-- 484
Thermotoga_maritima                 DITSETITVD-TEFEHTFKPFSCSVIEVELE-- 484
AFase_D3                            TSLTSIVRKDDGTITVTMPPASVGILRLG---- 494
Thermobacillus_xylanilyticus        APFRDFKLEG-GHLNASLPPMSVTVLELTAG-- 496
                                                  .  :  :
```

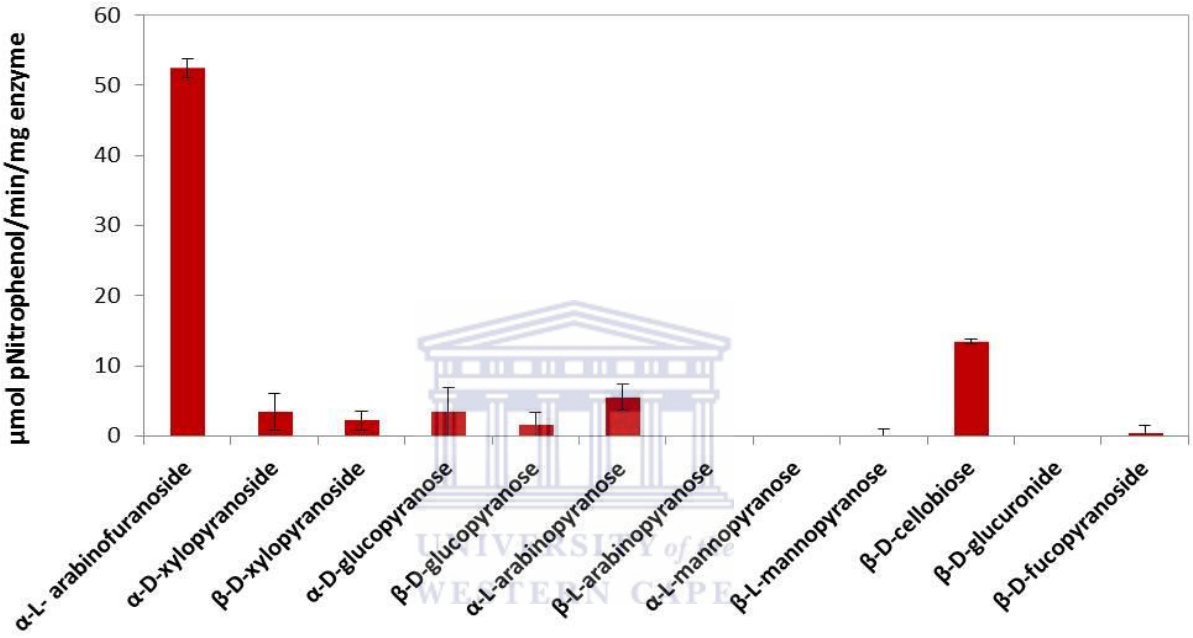
Figure 3.18 The AFase amino acid sequence of AFase_H4, AFase_E3 and AFase_D3 were aligned with six AFase sequences, retrieved from NCBI, using ClustalW (www.ebi.ac.uk/Tools/msa/clustalw2). The glutamic acid residues involved in the catalysis are highlighted in green, while substrate binding amino acids residues are indicated in blue. Furthermore, the cysteine residues that have previously been implicated in thermostability are shown in yellow and the amino acid residues that are proposed to play a role in the interactions between oligomers that contributes to thermostability are shown in red. The numbers at the end of each row indicates the amino acid position relative to its start codon (Methionine). The single and double dots refer to the absence of similar amino acid residues and two similar amino acid residues at an alignment position, respectively. The star symbolises three of the same amino acid residues at an alignment position.

3.4.3 Substrate specificities of AFases

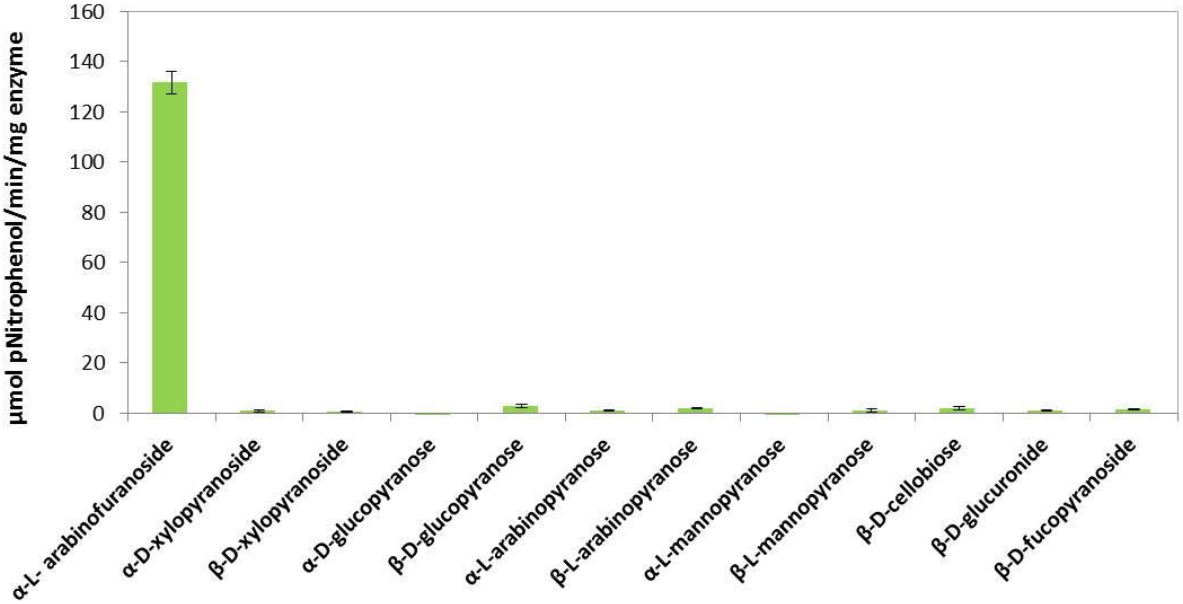
Substrate specificity was determined using synthetic *p*NP-glycosides and natural polymeric substrates. The AFases characterised in this study displayed differences in their affinities for hydrolysing the various *p*NP-glycoside bonds as shown in Figure 3.19. All three AFases displayed the greatest hydrolytic activity towards *p*NP-α-L-arabinofuranoside, this activity is common to AFases belonging to the GH51 family (Wagschal *et al.*, 2008; Wagschal *et al.*, 2009; Ohta *et al.*, 2013). Hydrolysis of the *p*NP-α-L-arabinofuranoside bond resembles the hydrolysis of the α-1.3 arabinofuranosyl bonds within L-arabinose-containing-polysaccharides (Saha, 2000; Taylor *et al.*, 2006). No hydrolytic activity was measured for the *p*NP-β-D-xylopyranoside (Figure 3.19). This presumably means that the three AFases are unable to hydrolyse two glycoside bonded xylose residues located within the hetero- or homoxylans backbone of hemicellulose. The enzymatic hydrolysis of *p*NP-β-D-xylopyranoside is a functionality of the GH43 family of AFases, which are known for their dual activity, possessing both β-xylosidase and the conventional α-L-arabinofuranosidase activity (Canakci *et al.*, 2007). Only one GH51 AFase has been shown to hydrolyse this bond, that from *Geobacillus caldoxylolyticus* TK4 (Canakci *et al.*, 2007). AFase_E3 displayed some hydrolysis against *p*NP-β-D-glucopyranoside and *p*NP-β-L-arabinopyranoside, while AFase_D3 has very low activity on *p*NP-β-D-glucopyranoside and *p*NP-α-L-mannopyranose. AFases have not previously been reported to have the capability of hydrolysing these synthetic substrates. AFase_H4 indicated was 10-fold less active on *p*NP-α-L-arabinopyranose than its preferred substrate. Activity on this substrate has only been reported from an AFase purified from *Caldicellulosiruptor saccharolyticus* which exhibited an equivalent preference to the conventional *p*NP-α-L-arabinofuranose (Lim *et al.*, 2010).

AFase_H4 displayed significant activity for *p*NP-β-D-cellobioside. This is the first report of this activity for a functionally characterized AFase belonging specifically to the GH51 family, and in fact has not been reported for any other family of AFases. This could be attributed to the cellulose-like domain present within this AFase as indicated from Figure 3.5.1.

A)



B)



C)

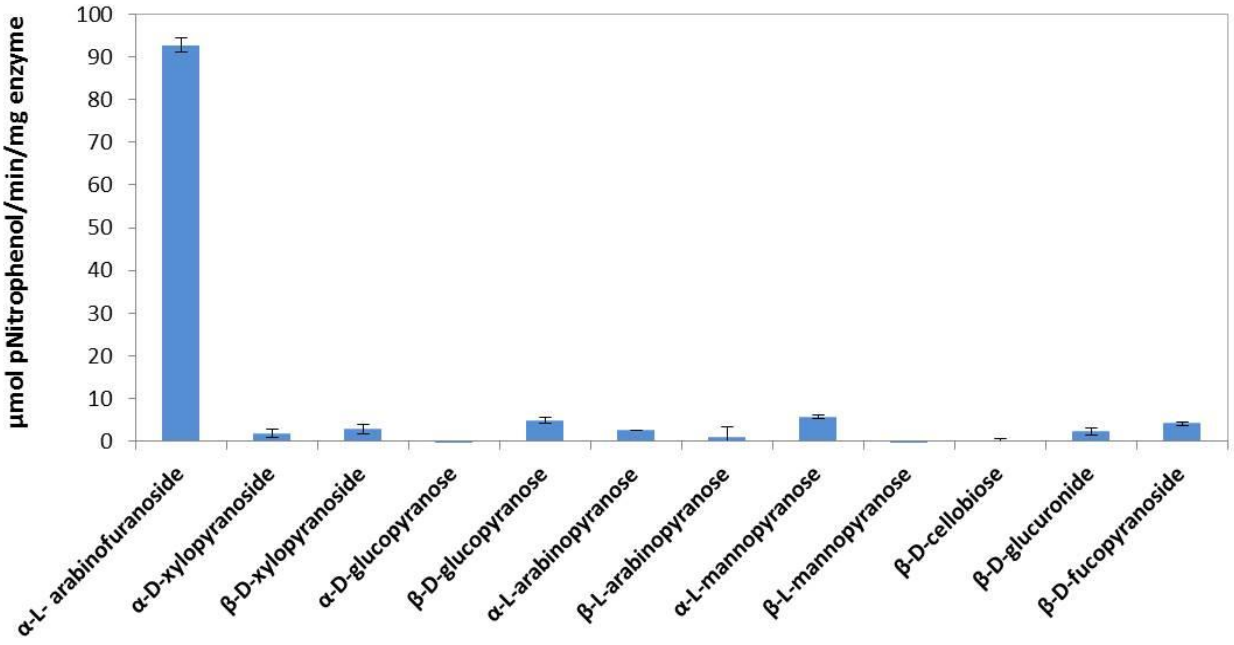


Figure 3.19 An array of pNP-glycosides assays has been executed to determine the mode of action of the A) AFase_H4, B) AFase_E3 and C) AFase_D3. Assays for the AFase activity occurred using 2 mM of each *p*NP-glycoside in 5 min reaction at their respective pH and temperature optimum. The release of *p*-nitrophenol, by enzymatic hydrolysis, was measured spectrophotometrically at 410 nm (Spectrostarnano,BMG Labtech). The percentage residual activity was calculated from a triplicate sample of purified protein. Data represents the average of three replicates ± Standard error (*n*=3).

The activities of AFase_H4, AFase_E3 and AFase_D3 were tested on the naturally occurring L-arabinose-containing polysaccharides: arabinoxylan, arabinan and linear arabinan. Following hydrolysis of these substrates, a DNS assay was performed to determine the release of reducing sugars from the polymeric substrates. All three AFases indicated different affinities for hydrolysing arabinofuranosyl bonds within the arabinoxylan, arabinan and linear arabinan (Figure 3.20). AFase_E3 indicated a higher affinity for these substrates when compared to AFase_H4 and AFase_D3. The hydrolysis of arabinan has been a common functional characteristic for GH51 AFases (Mckee *et al.*, 2012) but, AFase_H4 showed no

hydrolysis of arabinan. Similarly, an AFase from *Streptomyces sp.* and *Penicillium purpurogenum* was unable to hydrolyse this substrate (Fritz *et al.*, 2008; Yan *et al.*, 2010). It could be possible that the arabinan binding and catalytic determinants are not encoded by the AFase_H4 catalytic domain, but this property has not been experimentally proven. AFase_E3 and AFase_D3 had a lower hydrolysing effect on the linearized arabinan and it is known that GH51 AFases weakly hydrolyze the α-1,5-L arabinofuranosyl bonds, whereas this is a capability of the GH 43 family AFases (Zhou *et al*., 2012; Mckee *et al.*, 2012). A similar linear arabinan hydrolysis profile has been reported from Fritz *et al.*, 2008 confirming that AFases such as AFase_H4, AFase_E3 and AFase_D3 are typically exo-acting enzymes contrary to GH43 AFases which readily hydrolyze the α-1,5-L arabinofuranosyl bonds (Wagschal *et al.*, 2009; Zhou *et al.*, 2012). None of the AFases were able to release reducing sugars from arabinoxylan (Figure 3.20). This activity is characteristic of some GH51 AFases, but not all (Taylor *et al.*, 2006; Paes *et al.*, 2008; Remond *et al.*, 2008; Canakci *et al.*, 2008; Dumbrepatil *et al.*, 2012; Miyazaki *et al.*, 2005; Canakci *et al.*, 2007; Shi *et al.*, 2010; Cartmell *et al.*, 2011). Substrate specificity towards arabinoxylan has been attributed to specific residues in the catalytic domain in the *Thermotoga spp.* and *Thermobacillus spp.* of AFases (Paes *et al.*, 2008; Dumbrepatil *et al.*, 2012). The AFase arabinoxylan binding residue is a key xylan-binding residue that should be located in the sequence of the AFase_H4, AFase_E3 and AFase_D3 which is a tryptophan that is located approximately at the 96-100th position on the catalytic domain (Paes *et al.*, 2008; Souza *et al.*, 2011; Dumbrepatil *et al.*, 2012). However only AFase_E3 has two of these residues required for an electrostatic active site to liberate L-arabinose from the xylan backbone. Figure 3.20 shows no liberation of L-arabinose from the xylan backbone and this could be due to the absence of the residues required for an adequate electrostatic surface within the active site (Souza *et al.*, 2011).

The inability to liberate the L-arabinose from the xylan backbone has also been correlated to the ratio of the α-1,3-L and α-1,2-L arabinofuranosyl bonds that differs within different arabinoxylan-containing materials. There are particular arabinoxylans that are highly saturated with either α-1,3-L or α-1,2-L bonds (Caffell and Mohnen, 2009). Therefore, it is plausible that the arabinoxylan used in this experiment is saturated with more of α-1,2-L bonds. This conclusion is consistent owing to the AFases showing an affinity for the *p*-nitrophenyl-α-L-arabinofuranoside bond which resembles the α-1,3-L arabinofuranosyl bonds (Saha, 2000; Taylor *et al.*, 2006; Caffell and Mohnen, 2009). It has been demonstrated that the α-L positions of these arabinofuranosyl moieties can be confirmed by the principles and application of MALDI-TOF MS (matrix-assited laser-induced desorption ionisation time-of-flight mass spectrometry) (York *et al.*, 1996; Vierhuis *et al.*, 2001), and it is suggested that this be employed to confirm the nature of the arabinoxylan used.

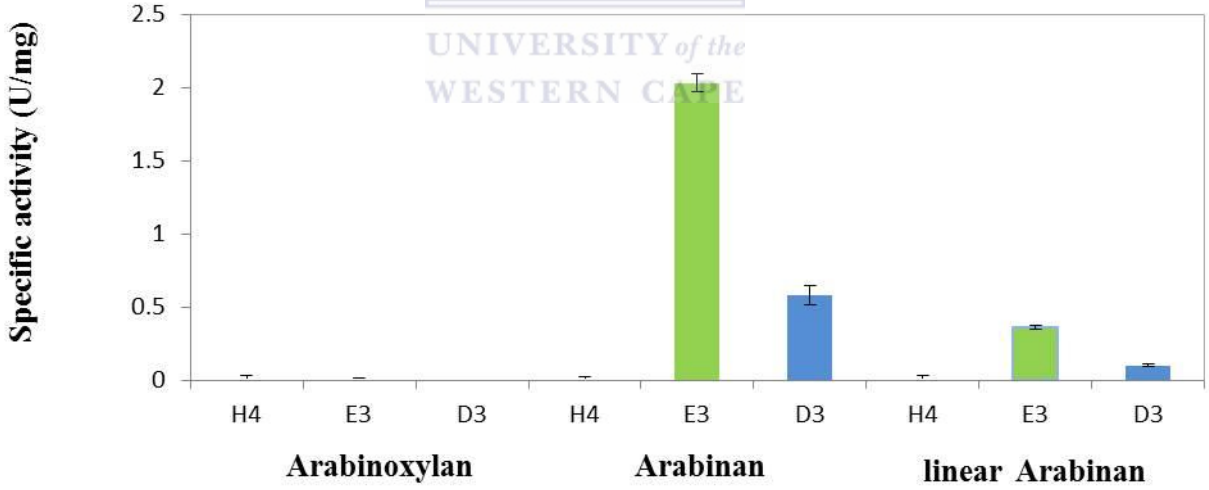
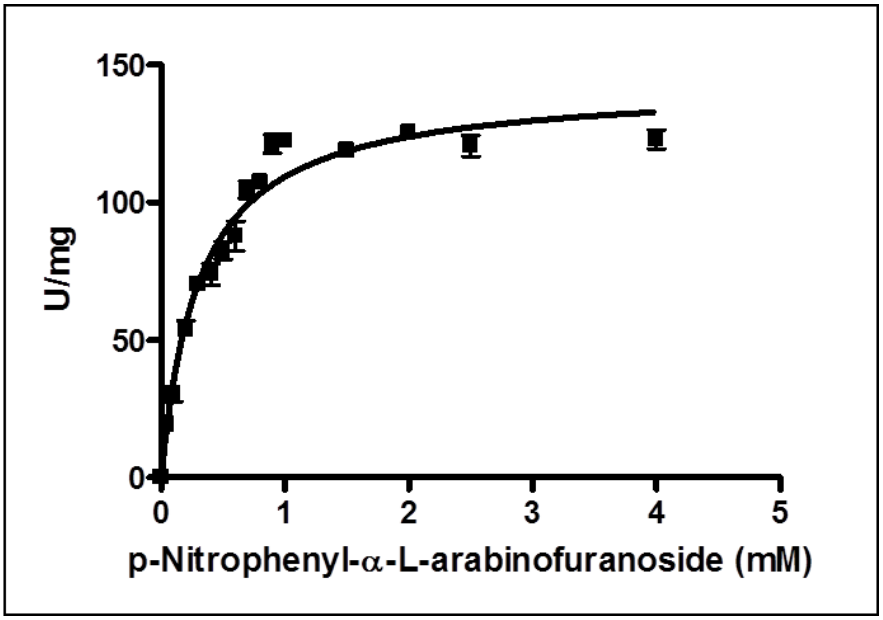


Figure 3.20 Three L-arabinose-containing-polysaccharides, arabinoxylan, arabinan and linear arabinan, were used in a glycoside hydrolysing assay for each purified AFase: AFase_H4 (■), AFase_E3 (■) and AFase_D3 (■). The absorbance of the DNS-reducing sugar complex was measured spectrophotometrically at 510 nm (Spectrostarnano,BMG Labtech). The pH and temperature optimum were incorporated in this reactions. The specific activity was determined in U/mg. (n = 3) ± standard error.

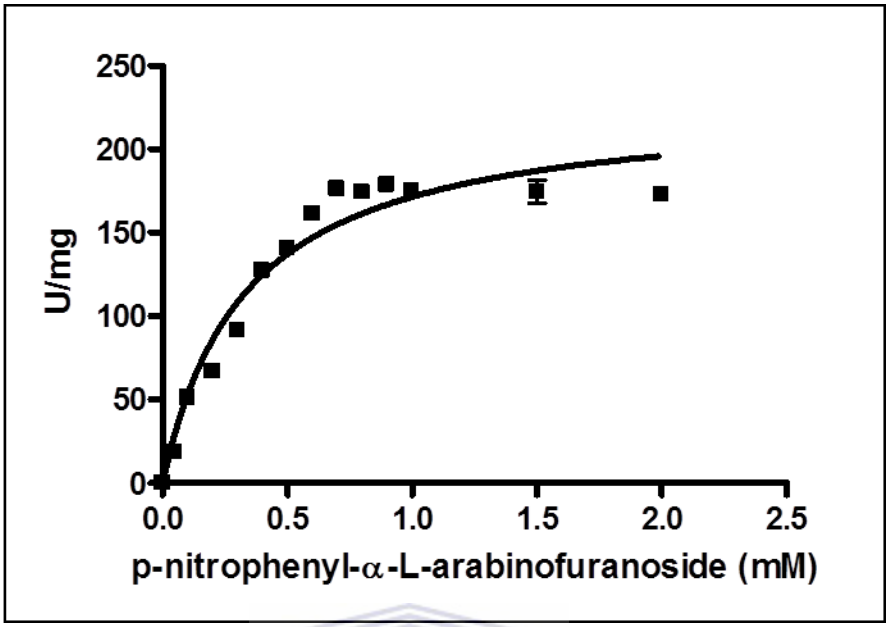
3.4.4 Enzyme kinetic characterization

Enzyme kinetics for AFase_H4, AFase_E3 and AFase_D3 were determined using pNP-α-L-arabinofuranoside as substrate. By increasing the substrate concentrations Michealis-Menten graphs were generated and analysed to determine the kinetic data (K_m, V_{max} and K_{cat}) for each AFase (Figure 3.21). AFase_H4 and AFase_E3 displayed lower Km values when compared to AFase_D3. This indicates that AFase_D3 has a lower affinity for *p*NP-α-L-arabinofuranoside compared to AFase_H4 and AFase_E3. AFase_E3 had the highest maximum velocity (V_{max}) and the highest turnover number (K_{cat}). The data correlates to *p*NP-α-L-arabinofuranoside and L-arabinose-containing-polysaccharide hydrolysis data obtained for substrate specificity in which AFase_E3 indicated the highest hydrolysing activity. In table 3.6, the kinetic data of the three AFases have been compared to other functionally characterized thermostable AFases from the GH51 family. All three AFases have similar overall catalytically efficiencies to the published kinetically characterized AFases (Table 3.6). The kinetic data of AFase_E3 is most similar to AFases characterised from *Geobacillus caldoxylolyticus* TK4 and *Anoxybacillus kestanbolensis*. Both these AFases could be suggested for use as industrial catalysts for highly thermophilic ethanologenesis process (Table 3.6).

A)



B)



C)

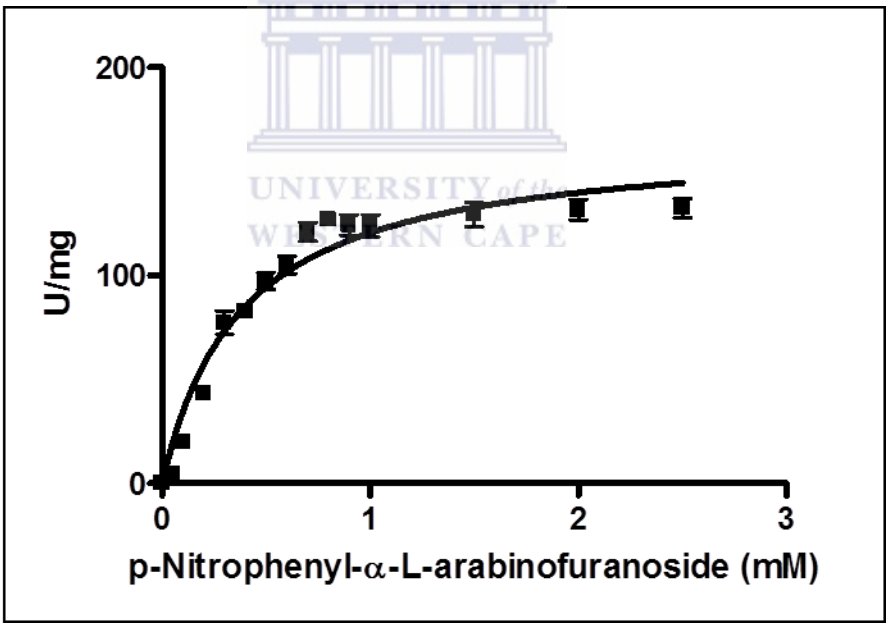


Figure 3.21 Michaelis-Menten plots of A) AFase_H4, B) AFase_E3 and C) AFase_D3 activity. The kinetics of each enzyme was determined in their respective pH and temperature optima's with increasing amounts of *p*NP-α-L-arabinofuranoside (mM). (n = 3) ± standard error.

Table 3.6 Comparative enzyme kinetics of AFases characterised in this study and thermostable AFases obtained from literature.

AFase	**Micro-organsim**	**Reference**	**K_m (mM)**	**V_{max} (U/mg)**	**K_{cat} (s^{-1})**	**K_{cat}/K_m**
AFase_H4	Unknown	This study	0.31	143.10	21.82	70.7
AFase_E3	Unknown	This study	0.33	228.60	35.88	107.7
AFase_D3	Unknown	This study	0.43	175.60	27.17	63.3
abf51S9	*Streptomyces sp* S9	Shi *et al.*,2010	1.45	221.00	50.83	35.1
AbfATK4	*Geobacillus caldoxylolyticus* TK4	Canakci *et al.*, 2007	0.17	588.20	142.15	836.2
AbfAC26Sari	*Anoxybacillus kestanbolensis*	Canakci *et al*., 2008	0.14	1.02	0.25	1.8
Tm-Afase	*Thermotoga maritima*	Miyazaki., 2005	0.42	-	21.70	52.2
AFase (C. Sacchrolyticus)	*Caldocellulosiruptor sacchrolyticus*	Lim *et al.*,2010	1.29	-	285.00	220.9
abfB	*Auriobasidium pullulans* ATCC 20524	Ohta *et al.*,2013	6.27	78.10	16.10	2.6

UNIVERSITY of the
WESTERN CAPE

Chapter 4:
General Conclusion

Chapter 4: General conclusion

Liquid biofuels are increasingly being noticed as an alternative to conventional fuels and the increasing environmental hazardous factors caused by GHG emissions. The common candidate for a renewable source of transport fuel is bioethanol generated from first generation processes (Gray *et al.*, 2006; Bhalla *et al.*, 2013). Currently the favourable fermentation process uses food-based products as substrates to yield biofuel products, however, developing economies, such as South Africa, cannot rely on staple food products (maize, vegetable oil, sugar and animal fats) as feedstock to yield bioethanol (Shafiee and Topal, 2009). These products allow the effortless yielding of biofuels, due to the fermentable sugars being readily available (Amigun *et al.*, 2008). Lignocellulose-based bioethanol was therefore proposed as an alternative to eradicate the economic issues arising from using food-based feedstocks (Sims *et al.*, 2010). The fibrous inert network of cellulose encompassed by hemicellulose and lignin prevents access to the fermentable monosaccharides by anaerobic micro-organisms (Remond *et al.*, 2008). A major worldwide research focus over several years has been to search, isolate and characterize thermostable lignocellulosic enzymes to synergistically and extensively hydrolyse the crystalline structure of lignocellulose. Thermostability is a key characteristic owing to its numerous advantages over its mesophilic counterparts, making it one of the economically feasible ways of commercialising second generation biofuels. The objectives of the overall project were to screen for novel genes encoding thermostable enzymes for efficient hydrolysis of a number of different crystalline feedstocks, using a functional metagenomic approach. The principles of metagenomics have permitted the harnessing of the novel genes encoding saccharolytic thermostable enzymes that can potentially hydrolyse the pre-treated lignocellulose feedstock. α-L-arabinofuranosidases (AFases) have been targeted as one of the enzymes needed for an

efficient hydrolytic cocktail owing to its synergistic properties with endoxylanases, which in turn has steric liberating properties for subsequent cellulase hydrolysis.

In this study three AFases from a compost metagenomic library were characterized and evaluated for their potential inclusion in a thermal simultaneous saccharification and fermentation process, as well as application in other industries. All three AFases displayed different biochemical characteristics, despite all showing conserved overall structural similarity, containing the typical domains of a GH 51 family AFase and a homo-hexameric structure. AFase_H4 and AFase_E3 displayed thermostable characteristics while AFase_D3 performed optimally under mesophilic temperatures, however also displayed a rare ability to maintain hydrolytic activity across a wide range of temperatures.

Source of compost for lignocellulosic enzymes

Isolating both mesophilic and thermophilic AFases from a metagenomic library represents the process of composting which undergoes both mesophilc and thermophilic stages. Mesophilic micro-organisms dominate during the early stages of composting which lasts a few days. The activity of the micro-organisms's metabolism causes an increase in temperature of the compost which stimulates the growth and activity of thermophiles, and then follows with a cooling and final maturation phase (Goyal *et al.*, 2005; Sundberg *et al.*, 2004). Using a metagenomic approach it can be concluded that the isolation of both mesophilic and thermophilic enzymes from a compost sample is expected, and was paralleled by the results of this study. In further

support of this, the same compost metagenomic library was used to screen for three lignocellulosic classes of enzymes (other parallel work within IMBM), and for each class thermostable and mesophilic representatives were isolated. Several other studies have used compost as a source for thermophilic glycoside hydrolytic enzymes, using both a culturing approach (Zambare *et al*., 2011), as well as the functional screening of metagenomic compost libraries (Dougherty *et al.*, 2012; Pang *et al.*, 2009; Kwon *et al.*, 2010). In all these studies, including the findings reported here, novel thermophilic lignocellulosic enzymes were obtained, demonstrating that both the source of compost and the metagenomic approach were appropriate strategies to achieve the objectives of this study.

Suitability of metagenomic assay for thermophilic enzymes

The thermostability characteristics for each AFase of this study differed. Although AFase_E3 was the most thermostable AFase of the three, it was not as thermally stable as some previously reported AFases which could maintain residual activity at 90^{o}C for more than 24 hours (dos Santos *et al.*, 2011; Miyazaki, 2005). Isolating hyper-thermal AFases is sample-site-specific at which the highest temperature recorded will determine the extent of the thermostability i.e. *Thermotoga spp* were isolated from a Japanese oil reservoir that have *in situ* temperatures higher than 90^{o}C and *Geobacillus spp* were isolated from Yellowstone hot springs where temperatures exceed 95^{o}C (Miyazaki, 2005; Hovel *et al.*, 2003). AFases from both these micro-organisms were discovered to be the most thermostable to date. In comparison to the sample from which the

metagenomic library of this study was constructed the compost temperature was limited to 70^{o}C. Therefore it is not unexpected that the AFases screened from this metagenomic library are not as thermally robust as the *Thermotoga sp* and *Geobacillus sp* derived AFases. Other investigations into screening for adequate thermally stable AFases employing a mesophilic gene expression system have harnessed lignocellulosic enzymes with a similar biochemical characteristic as seen for AFase_H4 and AFase_E3. A study executed by Verma *et al*, (2013) demonstrated that screening for thermostable lignocellulolytic enzymes from a compost metagenomic library in *E. coli* DH10B, resulted in the isolation of xylanases that yield temperature optima and thermostability profiles similar to AFase_H4 and AFase_E3. This finding has been also substantiated by Dougherty *et al*, (2012) that isolated thermostable glycoside hydrolases from a compost metagenomic library.

The approach used to screen a metagenomic library is a very important aspect which needs to be considered at the start of the study. The most common approach is utilizing *E. coli* as the host for screening of the metagenomic library this is due to the extensive knowledge of this microorganism that has led to the development of numerous genetic manipulative tools available for small (<15 Kb) and larger (Cosmids; 15-40Kb, Fosmids; 25Kb-45Kb) insert cloning (Angelov *et al*., 2009; Kakirde *et al*., 2011; Ekkers *et al*., 2012).

However, because of the enormous effort and expense of screening large libraries, a more stringent approach should be designed in order to maximise efforts. In this particular study, this is a very important aspect to consider when evaluating the level to which thermostable enzymes were isolated. The library was screened for AFases under mesophilic conditions (mesophilic host

and screened for activity at 37^{o}C). This approach will have neglected the thermostable enzymes present in the library which were either inactive or expressed at very low relative activity as a result of being inflexible for catalysis at lower temperatures. Furthermore, other heterologous expression limitations which may have impaired the enzyme from folding into its native state, such as lack of appropriate chaperones and incorrect codon usage, this limitation is also accompanied by the experimentally proven fact that only 40% of foreign genes are adequately expressed in *E. coli* (Feller, 2012; Gaobor *et al*., 2004). Therefore it can be expected that the number of thermophilic AFases in the library was under-estimated.

The initial functional screening method of the supernatants in this study should have been tailored to more adequately and specifically identify thermostable lignocellulosic enzymes from the compost metagenome. To clarify the proposed idea, a trial experiment was executed by Angelov *et al*, 2009 using a pCC1fos fosmid library constructed with genomic DNA from *Spirochaeta thermophila* (*S. thermophila*). Comparative functional screening between the fosmid library hosts, *Thermus thermophilus* (*T. thermophilus*) and *E. coli*, showed that 50% of the fosmids that showed xylanase activity when screened in *T. thermophilus* were not active in *E. coli*. This finding indicated that a higher number of *S. thermophila* genes were more frequently functionally expressed in *T. thermophilus* in comparison to screening the library in *E. coli*. To use *T. thermophilus* as an alternative host-vector system may prove particularly valuable for the functional screening of complex metagenomic libraries for thermostable lignocellulosic enzymes A second alternative is the use of multiple-hosts, in sequence or parallel, to increase the likelihood of metagenomic gene expression owing to the diverse expression machinery. Broad-host range screening uses shuttle vectors and this has been executed by Aakvik *et al*., 2009,

demonstrating that fosmid libraries (constructed from broad-host-range fosmid and the BAC vector pRS44) could stably hold up to 200kb of insert and be transformed into a wide range of hosts such as *Pseudomonas fluorescens* and *Xanthomonas campestris*. A further consideration is the initial library screening (not conducted in this study), where the library clone supernatants were screened at 37^{o}C. Although the clones were still required to be expressed in *E. coli* at a mesophilic temperature, the screening assay was independent of growth and could have been conducted at higher temperatures in order to selectively target the clones expressing the more thermostable enzymes. No recorded assay of this nature has been investigated using metagenomic libraries constructed from the commercial vector, pCC1 Fos (Epicentre).

Use of metagenomics to access novelty- sequence and function.

The use of a metagenomic approach as a revolutionary molecular tool to harness products for industrial applications was an objective that has been achieved for this project through the isolation of AFases displaying differential biochemical activities. However, the metagenomics approach is also expected to have the potential to access novel activities or properties.. Sequence-based screening is biased to already known sequences available on databases, relying on the alignment of probes or primers to predetermined conserved regions. Function-based screening on the other hand, possesses the practical potential of isolating novel genes encoding novel activities which cannot necessarily be predicted through sequence-based screening (Simon and Daniel, 2011). The AFases isolated from the compost in this study were novel at the

sequence level, displaying less than 70% sequence identity to anything available on the databases. Particularly for AFAse_H4, this enzyme branched very deeply with the closest relatives. However, at the functional level, these AFases were largely similar to other biochemically characterized GH 51 family AFases. This is a common theme in many metagenomic studies, particularly for the screening of hydrolytic enzymes, where sequence novelty is achieved, while promiscuous or novel activities are less often identified. Exceptions include the AFase characterized from *Caldicellulosiruptor saccharolyticus* and *Pleurotus ostreatus*; studies executed from Lim *et al*, (2010) and Amore *et al*, (2012), respectively.

It is now well recognised that the activities isolated in metagenomic studies are dictated by the novelty of the assay used, and the use of well-established assays places a substantial restriction on the level of functional novelty that can be acquired (Tuffin *et al.,* 2009).

The assays used in many metagenomic studies are generally limited to simple screening systems, typically using colourimetric, spectrofluorimetric or spectrophotometric synthetic compounds. The AFase activities characterised in this study were isolated based on the ability to hydrolyse the synthetic chromophore substrates provided. The synthetic *p*NP-α-L-arabinofuranoside used mimics the α-1,3-arabinofuranosyl bond within natural substrates. Using this synthetic substrate could have discriminated against novel AFases with the ability to hydrolyse glycosidic bonds other than the synthetic *p*NP-α-L-arabinofuranoside. It is therefore imperative to incorporate the right substrates to increase the chances of isolating the enzymes with the desired properties. This could be achieved by screening with multiple synthetic substrates that resemble all the possible glycosidic bonds in naturally occurring substrates to complete hydrolysis of all arabinofuranosyl

links, or alternatively to develop assays which would enable screening using natural substrates. This could increase the chances of identifying enzymes with novel (and multiple) functional motifs or domains.

However, alternative metagenomic approaches are emerging that can access higher diversity of enzymes and activities suited for industrial applications. Metatranscriptomics and metaproteomics are relatively new approaches that could enhance both sequence and functional base screening of novel biocatalysts. Briefly, the former can "differentiate between expressed and non-expressed genes" and not only the functional screening capacity of the microbial community of the sample, a technique that metagenomics cannot deliver with DNA-based analysis alone. (Frias-Lopez *et al*., 2008; De Long, 2009; Sorek and Cossart., 2010). The latter involves the analysis of the proteome and the catalytic potential of complex microbial communities, a large-scale characterization of an environmental sample i.e. compost (Welmes *et al*., 2008). As a future initiative, these two arms of metagenomics could aid in the search of thermophilic and mesophilic biocatalysts (specifically AFases) with novel properties suited for specific industrial applications.

Despite the limitations associated with the assay and heterologous expression, a range of AFase activities and properties was isolated using the simple and synthetic screening assay, which demonstrates the capability of metagenomics to yield a wide range of functionalities. Several additional functional properties of these enzymes validate this approach. AFase_D3 is the first GH 51 family AFase to function optimally at 25^{o}C. Furthermore, despite functioning optimally at a mesophilic optimum temperature, it maintains activity at temperatures as high as 80^{o}C and

90^{o}C (Figure 3.16), a generally rare property of enzymes. AFase_H4 is the first AFase to contain a potential dual activity within its catalytic domain, as a result of its ability to hydrolyse cellobiose. Furthermore no other AFase has been reported to contain the cellulose-like domain. Lastly, it is a common characteristic for GH 43 family of AFases to hydrolyze all the arabinofuranosyl bonds, and has not been shown for any other GH 51 family AFase. This study therefore has certainly delivered on novelty at both the sequence and functional level.

Properties of the AFases which makes them suitable for industrial application

Bioethanol from a lignocellulose feedstock requires a hydrolysis synergy amongst three classes of lignocellulosic enzymes to degrade the complex polysaccharides. The inclusion of small quantities of hemicellulases significantly enhances the activity of the cellulases (Yang *et al*., 2006; Naik *et al*., 2010; Van Dyk and Pletschke., 2012). AFases work in concert and synergistically with xylanases and xylosidases in the degradation of xylan biomass (Alvira *et al.*, 2011; Yang *et al.*, 2006). A future study can be executed to investigate the contribution which AFase_E3 and AFase_D3 may have on endo-xylanases. AFase_D3 would likely only yield a weak synergy, since it displayed only low hydrolytic activity on both arabinan and linear arabinan (Figure 3.20). AFase_E3 displayed a low hydrolytic activity on the linear arabinan but showed a high affinity for the arabinan substrate. Both AFases cannot readily hydrolyze the α-1,5-L-arabinofuranosyl bonds and therefore cannot provide adequate synergistic activity if only this bond is present in the pre-treated lignocellulosic substrate. AFase_E3 showed an extensive

hydrolytic activity on arabinan, hydrolysing α-1,2-L- and α-1,3-L-arabinofuranosyl bonds, suggesting that this enzyme is a candidate for synergistic studies in conjunction with endo-xylanases. AFase_H4 cannot hydrolyze the α-1,2-L-arabinofuranosyl bond and lacked the ability to hydrolyse any of the complex substrates used in this study, and would therefore likely be unsuitable for synergy experiments with endo-xylanase. Although it has the ability to hydrolyse the α-1,3-L-arabinofuranosyl bond, it is possible that the arabinan used in this study had a higher percentage of α-1,2-L-arabinofuranosyl bonds, and could account for the seemingly inability to hydrolyse the arabinan (Dumbrepatil *et al*., 2012; Paes *et al*., 2008; Hovel *et al*., 2003).

Although the AFases from *Thermotoga sp.* are hyperthermostable enzymes and therefore have higher potential as biocatalysts within thermal industrial processes, their kinetic properties suggest they will be poor candidates for consideration as industrial biocatalysts (Canakci *et al.*, 2007; Debeche *et al.*, 2000). AFase_E3 is less thermostable than AFases characterized from *Thermotoga maritime*, however AFase_E3 has a higher catalytic efficiency and could therefore be suitable for inclusion in a thermophilic ethanologenesis process. AFase_E3 showed a high catalytic efficiency on the α-1,3-L-arabinofuranosyl bonds. Its use as an industrial catalyst is therefore restricted to substrates saturated in α-1,3-L-arabinofuranosyl bonds. Substrates saturated with solely this bond are currently unknown.

Due to its temperature sensitivity and low temperature optimum, AFase_D3 has no potential application for a thermophilic bioethanol process. However, its ability to maintain activity at temperatures way above its optimum is a sought after property for industrial enzymes- this versatility means it can operate at a range of temperatures, and is likely a very stable protein

under a wide range of conditions. Although secondary bioethanol has been partially succeeded by the use of algae (third generation) and the production of biohydrogen (fourth generation), emerging industrial biotechnological applications may occur for AFase_D3. Recently, it has been demonstrated that AFases with this property can be potentially used in novel industrialized processes to yield new biocommodities. Buschke and co-authors in 2011 have demonstrated the need of a cold-adaptive or mesophilic AFases that can be applied in synergy with endo-xylanases and β-xylosidases for liberating xylose for the synthesis of bio-polyamides. The xylose residues are used in the pentose phosphate pathway for conversion to xylulose-5 phosphate and the *Corynebacterium glutamicum* metabolises the xylulose-5 phosphate to 1,5 diaminopentane, an important building block for bio-polyamides (Buschke *et al.*, 2011). Meiswinkel and co-authors in 2012 have shown that important bioproducts can be yielded from hemicellulosic waste which contains an abundance of arabinose and xylose. *Corynebacterium glutamicum* utilized the pentose sugars for generating pharmaceutical related bioproducts such as glumate, lysine, ornithine and diamine putrescine (Meiswinkel *et al.*, 2012). AFase_D3's thermolabile and mesophilic adaptive properties can be potentially used in emerging bioindustries where the need of high thermal energy is decreased or completely not required (Feller, 2012). AFase_D3 has a rare characteristic of being able to function at lower than mild temperatures and continuously be active at thermal temperatures.

Three AFases with novel biochemical characteristics have been harnessed through functional based metagenomics. Sequence-based metagenomics would not have detected novel biocharacteristics of these enzymes. Owing to the present limitations of functional assay screenings, untapped glycoside hydrolases are yet to be discovered that could possess enzymatic

functions suited for emerging industrial applications. To utilize the compost metagenome extensively for novel lignocellulosic enzymes i.e. AFases, host-vectors screening assays should be tailored to a new specific function identified for an industrial purpose.

Appendices

A1)

```
1       CCCATTTCCAGATTTTCCTGCACGGTCATCTGAGGGAAGTTGGACCGGCCACCGGGAATG
61      AAGGCTATGCCCCGTTTGAGCAGCTCAAGCGTTGGCAGACCGGTGATGTCTGCCCCTTCG
121     TACAGTATCTGGCCGCGCGATGGCTTCACCAGCCCGAAGATGGTGCGCAGCAGCGTCGTC
181     TTGCCCGCGCCGTTCGCGCCGATGACGGTGACGATCTGTCCGTGCTCGACGCGCAGCGAA
241     ACGTTCCTGAGGACTTCGATTTCCCCGTATCCGGCATCGACGTTGCGCAGTTCGAGGATG
301     ACGTTGTCAGTGGCCAAAGTATGCCTCCCGCACTTTCTCGTCCATCCGCACATCGCGTGG
361     CAGGCCGTCGGCGATCTTGCGGCCGTAGTCCAGCACGACGATGCGCTGCGAGATGTCCAT
421     GACGAGCCGCATGTTGTGTTCGACCAGCAGGATTGTTTTGCCCTGCCGGTTCAGGTTCAG
481     AATGTGGTTCATCATGTGCGTGACCATCTCGCCGCTGACGCCTGCCGCCGGCTCGTCGAG
541     CATGATGACGTCGGGGTCGGGCATGAGCACGGTGGCAAATTCCAGCAGTTTGCGCTGGCC
601     ATAGGAAAGATGGCGCGCTTCGAGATGCCGGACGTTATCCAGCCCCACCATGTCGAGCAG
661     TTCACGCGCCCGTTCGGTGGCGGTGTGTTCCAGATGGTGAATGCGGCTGGTCCCCAGCAT
721     ACGCGCGATCAGGTTTTCCTCCTGATGCTGCTGCACGGCAGCGAGCATGTTTTCGAGCAC
781     GGTCATTTCGGGGAACGTCCGCACGTCCTGAACGTGCGCCGGATGCCCTTGAGCGCGATT
841     TTGTCCGG
```

A2)

```
1       GCTTCTGCGCGAGCGCGAACGTATGGAATTGTGGGAAATGGAACAGGAAACGCTGGAGCT
61      TATGCGACGCAACGACGACAGCCCACCCGCGGAACCCGATTCCTCGCCCTTCTGACGGGC
121     AGACCAGCCCTATTCGTCCGGCAGTCGCTGTGCAGCTTCAGCGTTGTCGGTGGGAGTCAG
181     CAGAACGCCGCCCTCGACTCGAACCGGCACGACCTTTTCGATGCGCGGCTCGAAAACCAG
241     ATGTTTCAGCAGGATCGCAACCGTGCCGGCGACCGGGATGGAGAGCAGCGCGCCGGTTAT
301     GCCGTTAAGCGCGAAGCCCGCTGAGACGAACACGATGACAAGCAACGGATCGAACTCGAC
361     GCTGCGCGACATGATACGCGGCGACAGCACGTAATTTTCGAACTGCTGCAGGGCAGTCGT
421     CACGATAAGCACGGTCACGGCGTTAATGGGCGACGCGATCAGTGCCAGGAACGTGGATGT
481     GGCAACACCGATAAAGCCGCCGATGATCGGAATGGCCGTCGACAGTCCGTAGACGAAGCC
541     GATGGCGGCAGCGTTGGGCACGCGCAGAAGCGCCAGAATGATGAAACTGAGACCTCCAAC
601     AATCAGCGATACCAGTACCAAGCCGCGCAGATAAGAGCCCAGACCGACCTCGATTTCCTC
661     GATAATGACCCGAACCCGCTCGCGCCGGCTGAGCGGCACGAACTGCGACAGATAGCCGAC
721     GCCCCGGTCGCGGGCGGAAATCCAGTAAACGCCCATCACCAGAACCAGCACAAAGTCACC
781     GATGAACTCGCCCACATCGGAAACCAATGTCGGCGCG
```

B1)

```
1       CTTTTTCCTGTTTCTTATCCCTCCCGTACGGCATCCTCTACATTATAACACGACATCCGT
61      CTGCCATCTTGCGAAACCGATTTTTGTGCTATAATAAGTGGCGTGCGCGTGCAGGCATTG
121     AAAAAACGTTTATGCGCTTATAAAAAACTTTTTTCGTTTTTCGACGAGGTCCGCACTATG
181     CGATGCGGACATTTCATGATATCTTATTTGTCAATTCCCGATTCCGTTCAGAAAGGTTGT
241     CCGGAAATGGAACAGACGCATCTGAAAACAAAAGGGCTATACGATCCCTGTTTCGAACAC
301     GACGCCTGCGGCATCGGATTTGTGGCGCATCTGAAAGGGGACGCTTCGCACGACATCGTC
361     CTGGACGCTCTGACGGTTTTGCGAAATCTCGACCATCGCGGCGCAAGGGGAGCGGAAGCG
421     GACGTCGGGGACGGAGCGGGTATTCTCGTTCAGATACCGGACGCGTTTCTCCGCCACCAA
481     TGCGAGCAAATGGGTATCGAACTTCCGGCCCGCGGCGATTACGCGGTCGGGATGATTTTC
541     CTGCCCCGGGACGACCGTCTGCGGGAAACGTGCGAAGCGAAAATCGAACAGATTGTCGCC
601     GAAACGGGATTGCCCGTCCTCGGCTGGCGGACGGTGCCGACCGACGACACGACGCTGGGC
661     AAGACAGCCAGGGATAGCCAGCCTTTCATTCGGCAACTGATCGTCGGCAAGCCGGAGGAA
721     ATCGAAGAAGAACTCGCGTTCGAACGCCGTCTTTTCGTCGCCCGGAAACGTTGTGAACGC
781     GCAGCGAAGCATGCGGGCGGCAAGCCTCTTTTTTATTTCGCCAGTTTTCTTCGCGCACCG
841     TTGTTTATAAGGCATGCTGACGCCGGAACAGTTGGCGGACTATTACCTCGATCTGGGCGA
```

```
901        TCCGACTTTCGCCAGTGCGATCGCCCTTGTCCATTCGCGGTTCAGCACGAACACGTTTCC
```

B2)

```
1          GGTGTTTTGGACTGCGCGGCGGAGGGCTGACGGCGGCAGGACGGAGGCTGTTGGAAGAAT
61         GCCGCGCTTGGAGCGTCGCAATCGACGTTTCCCATCTCGGCGAAAACGCGTTCCACGAGA
121        CGATCGACTGGGCGGAAGCGAGCGGGCATCCGGTATTCGCGTCGCACGCCAACGCGCGAT
181        CGGTCTGCGGCCATCCGCGCAATTTGAGCGACGAACAGCTGATCCGGCTCTTTTCGCTCG
241        GAGGAGTCGTCGGCCTGACGTTCGTGCCGGCGTTTTTGAAAAACGACGGCAACGCCTCGA
301        TCGACGACGTACTTCGGCATCTGGAACACGTCTGCGAACTCGGAGGGCGACGACATGTCG
361        GTTTCGGCTCGGATTTCGACGGAACACCCGGCTATGTCCGCGGGCTGGAACATCCCGGCA
421        AGTTCGCGGTGCTCGCCGACGTACTCCTGCGGTATTATTCCGACGACGACGTGCAACACT
481        TTCTTTACGGCAATTGGCGGTCGTTCTGGCATCGCAATCTGCCCGATTGATCGCGAAACG
541        GCATGAATCCGATCGAAAATCGCTATCGTCTCCTGTGAAGGAAGTTCTTGCGCTCACACA
601        CCGTACAGCCCTAGAATTTTAATGAAGGGTGACATCGCCAAGCGGCGAAAGGGAGTGTTC
661        TTGCGTGATCAGCCAGTTGTCGCTGAAAATCGGCGGCCAGCAGGGCGAGGGCGTCGAATC
721        GACGGACAAAATTTTTTCGACGGCGCTCAACCGGCTGGGGTATTA
```

C1)

```
1          CCGTGATTTCAGTAGGTCCTTCTCCGAACGTACGCCAGAAGGTCGTGTCATAAATAGCTT
61         CACCATTGACCTTGAGCCAATCACCAATGGATAGTAGAATCTGCTTATCTTCATCTGGAA
121        TCGTTCCATCCGCTTTAGGTCCAATATTAAGCAGGAGATTGCCATTCTTGCTTACAATAT
181        CGACTAAATCCCTGATCAACTCTGTGGATGATTTATAATCGTTGTTTCGCGTATAACACC
241        ATGAATTTTTGGCGACAGCCGTATCCGTCTGCCAGAAGTAAGGCTTAAGCTCAGCGAATT
301        GACCTCTCTCAATATCCACAACAGCACTCCCTAACGCGAATGCATCATGCTTGTAGTTGA
361        TCGCAACAGGGTATCCCCATTCATCTGCGCAATTATAATAGTAGGCTGCAAAATTCTTGA
421        GGTAGGGCTTAAAAGCAGCCGTCTGAATCCACCAATCAAAATAAAATACCTTAGGCTTGT
481        AATGATCTACAAGCTCACAGCATCTCATTAACCAATCCTCTAAGTACTCTTGATTGGGCG
541        GTGAACCATACAGATCCTGATGATCGGGTTCGGGCATCGCTGGCCAATAGAAGTCCCCAC
601        ATACAAGCGGTTCACGAATATCTGAATCAAATTCCTTACCGTGTGACATAAAGAACCAGT
661        GCTCTGCTCGATGGGAGGAAACACATAATGCTAGATTTCGCTTCTCGAAGGCGGCTCTCA
721        TCTCAGCGAGAAGATCCCGCTTGGGCCCCATCTCAAATGCGTTATAGGTAGATAGCTCGC
781        TCTTATACATTTGGAATCCATCATGGTGCTCCGCTACAGGCATAACATATTTAGCGCCAG
841        ATTGGCTGACAGCTCTGCCCATTCGTCAGCATCGAATTTTTCCTCAGTAAACAGCGGTAT
901        GAAGTCTTTATACCCGAAGCTCTTATGTC
```

C2)

```
1          CCAAAATTTCGAGGTCTGCCTCATCGTCTACTACCAGTATCTTAGCCATGTTGAATCAAT
61         GTTTTGAGTTTTTCTTTGAGCAGGTTGAAATCCAGGGGTTTTGTCAAAAAATCATTTGCA
121        CCATTTTCCATTGCCTGTCGATGGTTTTCATCATCACCATAGGCAGTGATCATCATTACC
181        ACGGGCGGCGGCGCATGGAAGTCGTGACGGATGTGCGATAAAAGCTCGATGCCGCTCATT
241        CCCGGCATATTGATATCCGAAAGGATCAAAACCACTTCCGAGCTTTTCTGCCTTAAATAT
301        TCAAGCGCTTCTTCTCCTGAAAGCGCAAAATCGAATTCTATTTCGTGCTCGCGAAGCTCC
361        TTTCGGAAGCGTTGCAAAAAAAGTGGCTGAACATCTGCCTCGTCGTCGACGACTAATATT
421        TTCATAATCATGGTTTTATAACGTAAATATACTAATTCATTAATCTGTAGCAGTGTGCAG
481        CAGCACAATCGTAAACTGCGCAAACTCGCCTTCTTTGCTTTCCACTTTTATCGAGCCGTT
541        GTGCGCTCGAATGATGTCGTAACTCATACTCAATCCGAGGCCCGTTCCTTCCCCTGTTGG
601        TTTTGTTGTAAAAAATGGCTGCATGATCTTTTCGATTATATGCTGTGGGATTCCCATTCC
661        GTTATCGGCAACAATGATCGCAACTCCGTTTTTCAGCTGGTTTTGTAGAAATCGTAACGG
721        TAGGTTCGTAGCCTTCGGGGATTGAGCTTGCGCTTCTTCTCGG
```

Figure 3.8. The Forward and Reverse sequences from the fosmid A1, A2 pFos_H4; B1, B2 pFos_E3 and C1, C2 pFos_D3 generated from the T7 Promoter forward and pCC1Fos reverse primers.

Consensus sequence of transposon-treated Fos_AFase_H4:

A)

```
1       TGAACCGGCT CCGGCATCCA TGACGCCGGG GGCACCAGAT TCGCCTCGGC CAGCGCCGCC
61      AGGCAGCCCS GGAACCGATC TACCAGACTG CTGTACTTGG CCGGCCCTTT CAAGACAGCG
121     ATTTTCCGGT GTCCCTGATC CAATATGTAG CGCGCGATCT TATAACCGGC CATGAAGTTA
181     TCGCCCAACA CAGCGGGAAA AGGCTGGCCC TCAATATAGG TTTCGATCAA CACGACGGGC
241     ACATTCGGAA TCCGGAGTTG CATGACCATC TCAGGGTGAA TGTCGCCATC GTTGGCCACA
301     ACAAACCCAG CCACCTTGTG CGCAAAACCC GACTTCAACA GGTCCAGATC CTGCGTGTTG
361     CGATCGAAAG TCAACAGCAC AACGTGATAA CCCAGAGATT CGGCTTCAGA CTGAAAACCA
421     CGAATGACAT CGCCATAGAA CACATCGCTG ATCGCGGGTA TCGAGGACTG CTCGATGAGA
481     AGCCCGACCG ATTCCGCCTT TGAGCCGTTG TTATCTGTCT TAACAACGTA AGACAACTTG
541     TCCATCGCCG CCCGGACTTT CCTGACGGTT TCGGCGGGAA GGCCCGGTTT GTTATGAACG
601     ACCATCCACA CCGTTGTCGC AGATACACCA GCCGCACGAG CAACATCGCG CATGGTTGTT
661     TTCGACTTAG CCAACCGCTC TCTCCTCATA TAACAACGGC TGCGCGTTTA TTAAACAACA
721     CTCTCTTGTT TAATGTACAA GTATTTGACA GGTTTGGCAA AACGGCATAT ATATAGATTG
781     TGAAAGCGCT CGACTGAACG CACTTATGCT TAATTTAGCA TTCTAACCTT GCAGGCGATT
841     TCGGGAATTC GGCGCTGAAT GGGCCGTATA GATTTAGTAA ACACGCCGGC CAAGAATCAG
901     CTTTTTAGGG CGTATTTCGC CTGAAACTGG ACGTTGTTTT CATGAATCAC ATCAAGATTG
961     ATTTAGATCG TCAATCCGGC TGTATCCACC CTCATATTTT TGGCGGCTTC GTCGAACATC
1021    TGGGACGCTG CGTGTATGGC GGCATCTACG AGCCCGGTTC GCCATTGGCC GACAAACAGG
1081    GTTTTCGCCG GGATGTTCTG GAAGCCCTGC AGCGACTGAA CCTGTCTATT GTACGCTATC
1141    CCGGGGGCAA CTTCGTATCG GGCTATCGCT GGATGGATGG TATCGGTCCG GTCGAAGATC
1201    GCCCCACCCG GCCTGACCTC GCCTGGGGAG CCATCGAAAC GAATCACTTT GGAACGAACG
1261    AATTCATCCA GTTCTGCCGC ACGATCAACG CTGAACCCTA TCTGGTGGTC AATTGTGGCG
1321    ATGGGGACAT GCGCGAAGCG CGGGACTGGG TGGAATACTG CAACGGCACT CAGAACACAG
1381    CCCTCGCCAA CCTGCGCCGC CAGCACGGCT TCGACGCGCC TCATAACGTG AAATATTGGG
1441    GCATTGGCAA TGAGGTCGAT GGCGACTGGC AGATCGGCTA TAAGACCGCG CAGGAATACG
1501    CTCGCGCCTA TAAAGAATTC GCTAAAGTCA TGCGCTGGGT CGACCCATCG ATCAGCCTGC
1561    TGGCTTCAGC GGTATCCAGT TGGCGAACGG ACTTTGTCGA ACGTATCCAA CTTCTCCTCG
1621    ACCACGCGCC AGACCTGATC GACTATCTCG CCATCCACTG GTATGTCGGC AATCCAGAGG
1681    GCGATTTCGA AAAGTATATG GCCGTATCCG AGCTGATCGA GGAGCGGATC AGCGCTATCG
1741    AAGGCCTGAT CAGGGTGATG AAATTGCAGC GCAACATTCA GCGGCCGATT GCAATTGCGG
1801    TCGACGAATG GAACGTCTGG TACCGGACAC ATGGGGCAAC GCCGACCGGG CCGGACAACC
1861    TCGAAGAAAA ATACAATCTG GAGGACGCGC TGGTCGTCGC GATGCATTTC AATGCATTCA
1921    TTCGTCACGC CCGGTCCGTC AAAATGGCAA ATATCGCCCA GTTGGTCAAC GTCATCGCGC
1981    CGATCTTCAC CAACCCAGAC GGGCTCTTCT TGCAGACGAT CTTTTACCCC ATCGAAATCT
2041    ACCGCCAGTC GTGTGGTAAT ATCGCGCTAG ACGTATTTTG GACCGGCGAT ACGTTTTCAA
2101    CGGCGGAACA CGCCGGGCTG CGTGTGCTTG ATGTTTCGGC CACGCTTGAC GACCGCGCTC
2161    GAAAGCTGAC CGTCTTCGTG GTCAACCGAA GCCAGACCGA CGAGATGGAA ACCACCATCA
2221    CGCTGGACGG CGGCCAGTTC GCTGGCACAG GTCAGGCCTG GGTAGTCAAT GGCCCCGATA
2281    TCAAAGCCGA AAACTCGTTC GACGCTCCTG ACCGTGTGTC CGCAACCAGA TCGGCAGTGA
2341    CCGCCGGCTT AACCTCTGCC ACCTACACCT TCGAGCCGCA TTCGGTAACT GCGCTCAGCT
2401    TTGACTTATG ACCCCGAAAG GAGGTGAAGC GCAGGCACAT AACACAATAG CTGGTGTCAC
2461    ACAATTTGCA TTTTCATGTT CGTAAAAGGA GCTTGAGATG TCTAAACGCA GATTCATGAT
2521    GTTGCTTGTT TTGCTTGCCA TCAGCCTGGG CGCGGTTTCG CTGGTGAGCG CGCAGGAAGC
2581    CACCGATCTA TCGCTGTGGG TGTTTGTTGA GCGACACGGC ACCTTCATGC AGCATCAGGC
2641    AGAGCGCTGG AATGAATTGA ATCCAGACCG GCCCATCAAT CTCACGTTTG AG
```

Consensus sequence of the Transposon-treated Fos_AFase_E3:

B)

```
1       GCATGTCGAT CGAGACATAG GCGGCAAGGC TGCCGAGGGT AATGCCGGTG ATGTACTCGA
61      AAACGGACAA TTCCGTGATT TGGCGTTTGC CCAGAAGTCG CGTCACGACG AGCAACGCAA
121     GAATGGCGGC AAATGTTCGG ACGGCGACAA GCATCCACTC CGGCATGCGA TTCCCACCTT
181     ATTGATCGAT TGCCGACGCG CCGACGTGAA AATTGTGCGC GGCGACAGGA CGTTTTATTC
241     CCAGGCGACG GGCGRAATAT GCCGCGTGCT TTCGACGCAT GATGAATCCG TGAAGGAGGT
301     GACCGTCCGT GACCGTCGCC AGCCAGGTCA AACAATGCTT GGCGTCGCTG AAAAGCGCTC
361     AGGCCAATCT CGAACAGTTC GCGCTCAACA CGCAAAACCA GGAAGCCAAG TCGCTGTTTT
421     CGAGCGCAGC TCAAAGCACT CAGCAGATTG TTCAGAAACT GGAAAGCCGC ATTCAACAGA
481     TCGAGTTTGA AGAACCGCAA TACCGCGGGT TCTGACGCAC GCAAGCGTAA AAACGGCGCG
541     GTCCCGATCG TCGGGATCGC GCCGTTTTGC TGTTCTGAAT TGACGGACCG CCGCCGCTGT
601     GGTTTAATAT GTATGGTTAT GTTTTTATTA TTTAATTCAT ATTTATATTA ATCAAATGGA
661     CGGAGGCGCA TGCGTCATGG CATCGGCCAA ACTTCGCGTC GATCGCGCGT ATACGATCGG
721     CGATACGGAT CCGCGGTTGT TCGGGGCGTT CGTTGAACAT CTCGGCCGGG CGGTGTACGG
781     CGGGATTTAC GAGCCGGGGC ATCCGGAAGC TGACGAACAA GGATTCCGGC GGGACGTGCT
841     GGAGCTCGTC CGCGAGTTGG GCGTCCCGAT CGTTCGTTAC CCGGGCGGAA ATTTCGTGTC
901     GGGATATAAC TGGGAAGACG GCGTCGGCCC GAAGTCCGAG CGCAAACGCC GTCTTGAGCT
961     CGCCTGGCGA TCGATCGAGA CCAACGAATT CGGGACGAAC GAATTCGTCG ATTGGTGCAA
1021    AAAAGCAGGG GCCGAACCGA TGCTCGCAGT CAACCTCGGC ACGCGCGGCA TCGACGAGGC
1081    CCGCAATCTG GTGGAATATT GCAATTACCC CTCAGGCACG TATTGGAGCG ACTTGCGGGT
1141    GCGGCACGGC TACCGCGAAC CGCATCGCGT CAAAGTCTGG TGTCTCGGCA ATGAAATGGA
1201    TGGATCCTGG CAAATCGGCA GCAAAACGGC CGACGAATAC GGACGGTTGG CCTGCGAAAC
1261    TGCCAAGGCG ATGAAGTGGG TCGACCCGTC GATCGAACTC GTCGCCTGCG GCAGCTCGAG
1321    TCCTTCCATG CCGACGTTTC CGGAATGGGA AAGGATCGTT CTCGAACATA CGTACGATCA
1381    CGTCGACTAC GTGTCGCTGC ACATCTATTT CGGCAATCCG GAAAACGACA CGGCAAACTT
1441    TCTCGCCAAG CCGCTGGATA TGGAAAGGTT TATCCGGATC GTCAAAGCGA CGTGCGATTA
1501    CGTCAAAGCG AAGAAGCGCG GTAAAAAGGA TCTTTACATT TCGTTCGATG AGTGGAACGT
1561    CTGGTATCAT TCCCATGAAG CGGATAAACA ACAAAAACCG TGGCAAATTG CGCCCCCGCT
1621    TCTCGAAGAC GTCTACAATC TGGAAGATGC TCTTGTCGCC GGTTGCATGT TGATCACGTT
1681    GCTCCGTCAC GCCGACCGTG TCAAGATCGG CTGTCTTGCG CAACTCGTCA ACGTGATCGC
1741    CCCGATCATG ACGAAGACGG GCGGCCCGGT CTGGAGACAG ACGATTTTTT ATCCGTTCCT
1801    CCACGCATCC CGCTACGGAC GCGGCACCTC GCTCGTCACA CTCGTCGACG GGCCGAAGTA
1861    CGACAGCCGC GACTACACGG ACGTGCCGTA TGTCGAAGCG GCGGCAGTGT ACCGAGAAGA
1921    CGTCGGTGAG TTGACCGTGT TCGCGGTCAA CCGGCATTTG CACGAGCCGA TCGTGCTGGA
1981    AGGAAAATTG GACGGTTTTG AAAATGTTCG AGTGGTCGAG CATCTTGTGC TCGACCATCC
2041    CGATCTGAAA GCGGCCAACA CCGTCGACGC CCCGCATCGC GTTGTGCCTA GGCGACTATC
2101    GGACGCCGCC GCGGAAAACG GTCTTTTGAC GGCGCAGTTG CCCGCAGCGT CCTGGAACGT
2161    CATCCGCCTT GCCCGCCGAC CGTCCTGACG CGCAACGAGC GGATCCGATT CAGCGCGGCG
2221    GCAAGCTCGA ACGGCCGCGG CATCGCCAGA TCGCCGGGAT GCCGCCCGGG ATGCGGCGAC
2281    AGCAGATCGA GACCTTCGCG ACGAAGCCGC TCCCCGATCG CCGCGATCAG CTCGGCGACG
2341    GTACGCCGGC CGTCCATCG
```

Consensus sequence of Fos_AFase_D3 following transposon mutagenesis:

C)

```
1       TGCTTCGGTA ATGATCCATT CTCAAAAAAG TGAAATCGAA GCCTTTCCCC CACTTTCTAC
61      GCATCGTGTG CCCCCCATCA CATGACCTCA AACAGGGTCG TCCTCCGAAT TCTATCGTTC
121     AAGCTTATGC GTACGATGAG GAAAAAAGCA CATCTACTCT GCTAGCCCAG ATGGCATAAA
181     AAATTTGCAG TTGTTTTATT CTGCCATTTG AGATTAAATA GTATTATCTA TTACGTTATT
241     TCATTAATCA ATTAATATTA TTATTAATCC AAAGGAGAAG ATGAAATGAA CAATGTCGTC
301     ATCAATGTGG ATCATTCCAG AGGAAGCATT AATCGCAACA TCTACGGTCA CTTTGCAGAG
361     CACCTCGGCA GATGTATATA TGAAGGTATT TGGGTCGGGG AGGATTCACC AATTCCGAAT
421     ATACAAGGTA TCCGCACTGA TGTGTTAGCG GCGTTACGTC AATTGAAAAT TCCCGTGTTG
481     CGCTGGCCGG GGGGCTGTTT CGCCGATGAA TACCATTGGA TGGATGGGAT TGGACCTCGT
541     GAAACGCGAA AACGAATGGT GAATACCCAT TGGGGCGGTG TAGTGGAGAA TAACCATTTC
601     GGAACTCATG AATTTCTGCT GCTCTGTGAG TTGTTAGGCT GTGAGCCTTA TATTAACGGT
661     AACGTTGGCA GCGGAACGGT GAAAGAGATG CAGGAGTGGA TTGAATATAT GACCTTCGAT
721     GGTGATTCGC CTATGGCGAA TTTGCGCAGA GCCAATGGTC GCGAGAAGCC ATGGAAGGTG
781     AGCTACTTCG GAGTAGGTAA TGAGAATTGG GGCTGCGGCG GCAACATGCG TCCAGAATAC
841     TATGCCGATC TGTACCGGCG TTATCAGACA TACGTGCGAA GCTACGGCGA CAATCGTATC
901     GCGAGAATTG CTTGCGGTGC CAGCGATTAC GACACGAATT GGACTGAAGT GCTAATGCGT
961     GAATCCGCAA GGTATATGGA TGCACTCACC CTTCATTACT ATACCATTCC AAATACGTGG
1021    AAAGAGAAAG GCTCGGCAAC ACAATTCGGT GAACGGGACT GGTTCATCAC GCTTCGCAAG
1081    GCAATTGCCA TGGACGAGCT GTTGAATAAG CATGAGACGA TCATGAATAA ATACGATCCA
1141    GATAAGCGGG TAGCGCTAAT CGTTGACGAA TGGGGTACTT GGTTCGATGT CGAGCCAGGG
1201    ACGAACCCAG GCTTTCTGTA TCAGCAAAGT ACGATGCGTG ATGCCTTGGT TGCCGCTTTA
1261    ACGCTGCATA TTTTCCACGG TCATTGCGAC CGCGTACAGA TGGCCAACAT CGCGCAGACC
1321    GTTAACGTGC TCCAATCCGT TGTATTGACG GAAGGCGCAG CAATGACACT TACCCCGACC
1381    TATCATGTGT TCGATATGTT TAAGGTACAT CAGGATGCGG AAGCGCTCGA TGTATTCACC
1441    CAGTCTGAAT GCTACGAGAT GGATGGTGAG TCGATTCCCC AGCTAAGTGT CTCGGCTTCC
1501    CGCAATTCAG ATGGGATCAT CCATATCAGC CTGTGTAACG TTCATCATGA CCGCGAGGCT
1561    GTATTGACAA TTGATCTAAG AGGCACTGAT GGAAGCGATC AAGTCATAGA GGGAACCGTT
1621    CTTCGTGCTT CGGAGCTCAA TGCTCACAAT ACCTTCACTC AACCGAATAC CGTGCAGCCC
1681    ACTAGTCTCA CCAGCATCGT TCGCAAGGAT GACGGTACGA TCACGGTTAC GATGCCTCCG
1741    GCTTCAGTCG GCATTCTAAG ATTAGGTTGA TCAAAGTGGC CTCAGGACAA AGTCGAGAAG
1801    ACAATTGTAA AGGCTGTAGT GCGAATGTTC ATGTTACCGA AGCTCAGCTC GATCGAATAT
1861    TGCGCAAGCT CTCTCTTCAT GCTGAAGACT GTGTGACTGG AGAGCAATAT ACGAAACGGC
1921    TTGAGCAATG CGTAAAGTGT CCCGCTCTTC TATATGGAAC CACCTGCACC TATTGTGGTT
1981    GCTTTGTCCG GGTAAGGGCT AAATTGCTAG ACAAGCATTG TCCAAATCCC GCAGGCCAAC
2041    TATGGGAGAC ACTTTCAGTA TCAGAAGAAA CCGCACTGAA ATGAACATGA ATATAACAAT
2101    AGGGAAATGG TTGGGCAGAG ATCAATTCTG CTCAATCATT TCTCTTAGCC ATCTCTAAGG
2161    CTAAGACGTT TGCCTTCTAA ATTCAGTGGG CGTAAGCCCG ACATTCTTCT TAAACTGCTG
2221    AGCTTACTCT TGGCGAGTAG TTTCGCATAA ATCGCATCAG AATCAGGTCA GATAAGTAGT
2281    TCGCAAGCAG CGAATGATGC TTCATCATAA GATTCCACCT TAGAAAGTTA TTATTGTACA
2341    TAAGAGAGTA AGTGTCCTAA TTGTATTTCA ATACACTAAT AATAGTAAAT TGAGAGGATG
2401    ATATTGGCTA TGGATAAGCA TGAGTATTTG AAAATAATTA ATGAAACAAT TTCTAAAGGT
2461    AGATTTAAAG ACAATTGGGA CTCTCTTAGT GAATTTCAAG TCCC
```

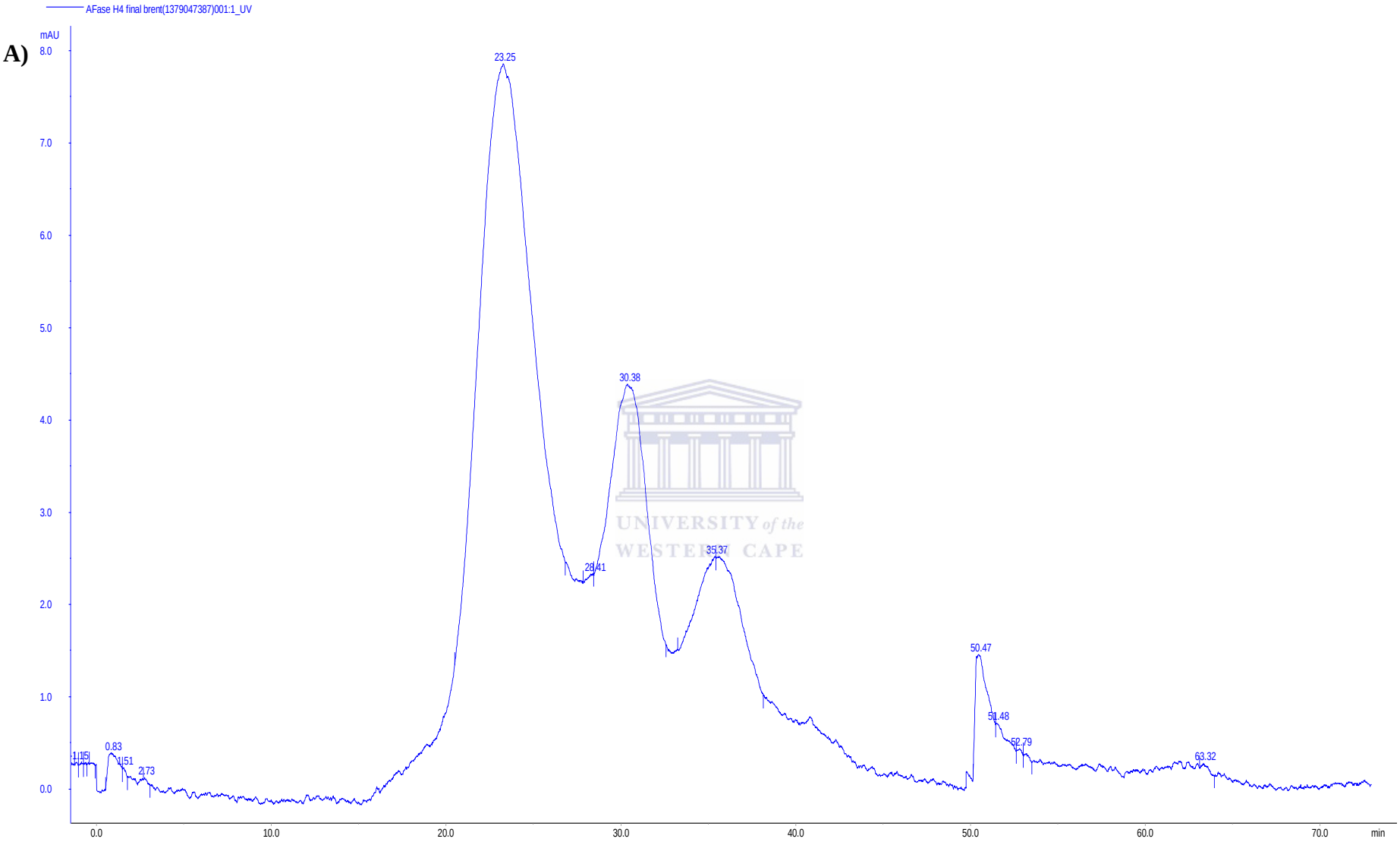
Figure 3.9: The nucleotide consensus sequence generated from the transposon mutagenesis primers MuKan FP and MuKan RP containing the ORF of A) *AFase_H4*, B) *AFase_E3* and C) *AFase_D3*. Other ORF's, mention in this transcript, are present as well.

Table A1: The standard protein markers and the AFase proteins with the respective retention time through the FPLC column as a function of its molecular weight (MW).

Standard protein markers and AFase proteins	MW	log(MW)	Retention time
b-amylase	200	2.30	24.6
alcohol dehydrogenase	150	2.18	26.4
albumin	66	1.82	27.88
Carb anhydrase	29	1.46	31.85
AFase_D3	316.2	2.5	22.95
AFase_E3	478.6	2.68	21.3
AFase_H4	288.4	2.46	23.25

Determining the quarternary structure of the native protein was executed through the FPLC (Fluorescent Pressure Liquid Chromatography). The retention time of the different protein standards were obtained under the same physio-chemical conditions as the purified AFase proteins. Table A1 indicates the different retention times which is the function of the logarithmic molecular weight. The Retention time of the three purified AFases (Table A1) confirmed the quarternary structure to be a Homo-hexamer or a dimer of trimers. Hexameric structures that have been determined by the principles of FPLC have been reported as well (Hoffman *et al.*, 2013), however other quarternary forms of the structure have been reported such as tetramers and octamers. Homotetrameric structures have been reported in AFases purified from *Streptomyces sp*, *Geobacillus caldoxylolyticus* TK4 and *Anoxybacillus kestanbolensis* AC26Sari (Shi *et al.*, 2010; Canakei *et al.*, 2007; Canakei *et al.*, 2008) and an experimental study performed by Lim and co-authors in 2010 have calculated a homo-octameric AFase structure isolated and purified from *Caldicellulosiruptor saccharolyticus*. The hexameric structures of the three purified AFases are consistent with the published crystallography experiments performed by Im *et al.*, 2012; Dumbrepatil *et al.*, 2012; Hovel *et al.*, 2003; Paes *et al.*, 2008; Souza *et al.*, 2011 and Taylor *et al.*, 2006 and each study has isolated and purified AFases from *Thermotoga maritima, Thermotoga petrophila, Geobacillus stearothermophilus*, *Thermobacillus xylanilyticus* and *Clostridium thermocellum*. Each momeric unit of the homo-hexamer consists of a $(\alpha/\beta)_8$ barrel catalytic domain and 12 stranded β-sandwich domain (Im *et al.*, 2012; Dumbrepatil *et al.*, 2012; Hovel *et al.*, 2003; Paes *et al.*, 2008; Souza *et al.*, 2011 and Taylor *et al.*, 2006).

A)



B)

Each peak fraction was tested for AFase activity. The peak with the retention time of 21.30 minutes indicated positive for AFase activity. Enzymatic data not shown.

C)

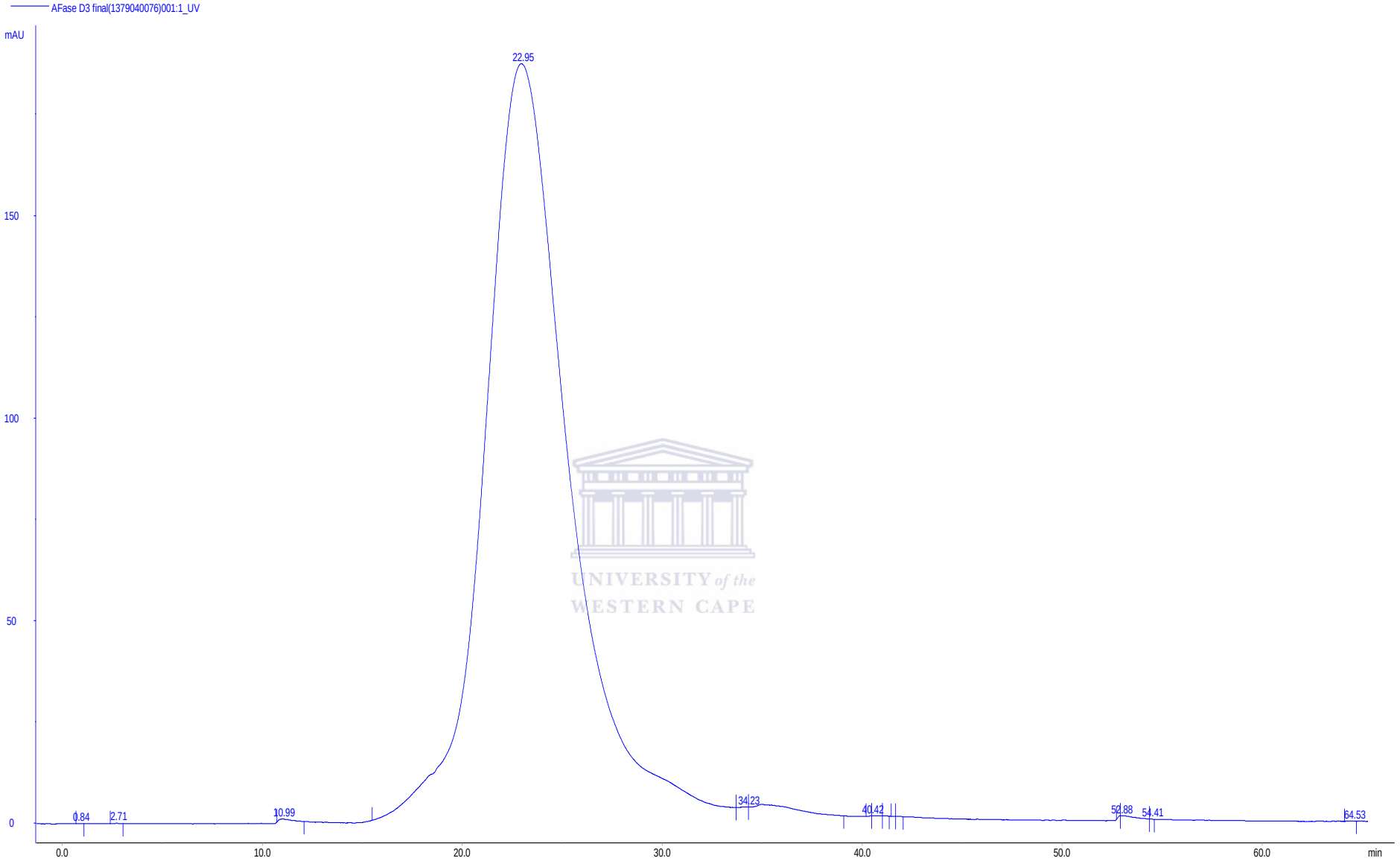


Figure A1. The Chromatograms indicating the milliabsorbance units (mAU) peaks measured as a function of retention time in minutes for proteins A) AFase_H4, B) AFase_E3 and C) AFase_D3.

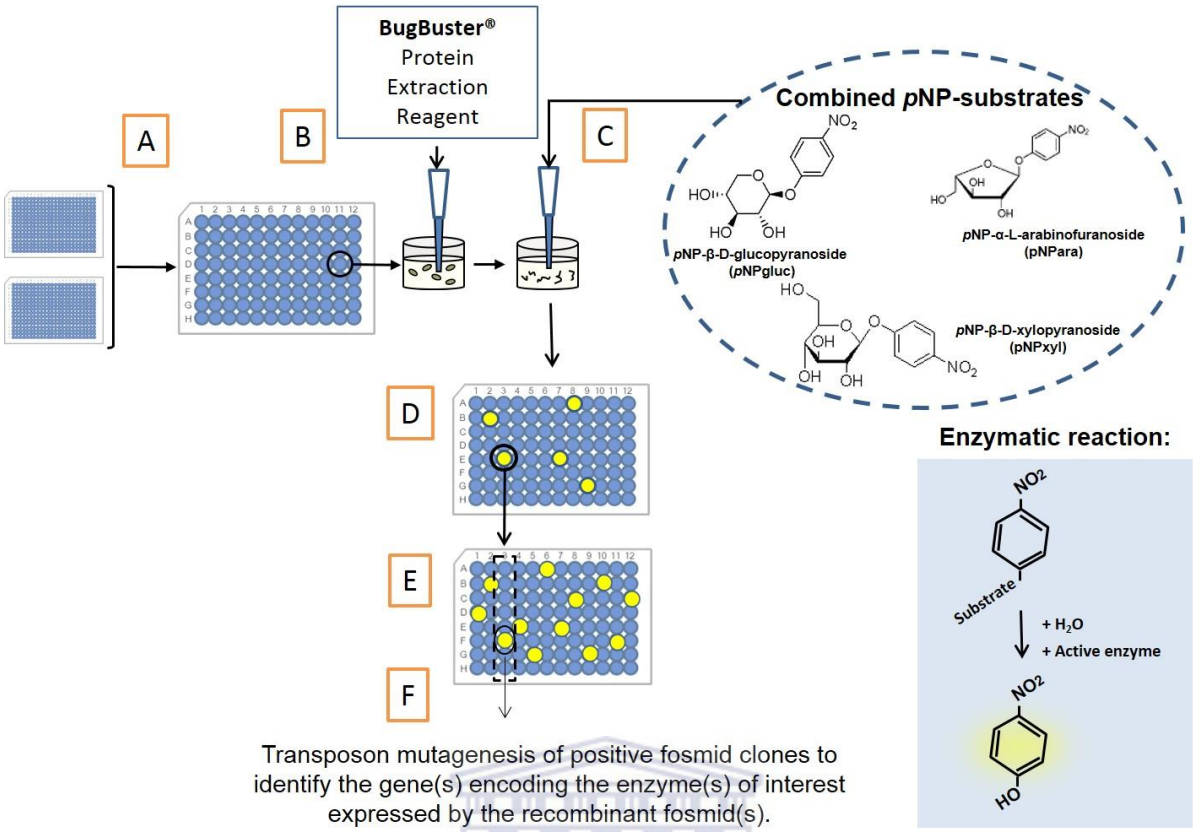


Figure A2. Diagrammatic representation of the novel high-throughput functional screening method used by Smart *et al*. (manuscript in preparation) to identify putative α-L-arabinofuranosidase, β-glucosidase or β-xylosidase enzymes encoded by a recombinant fosmid. (A) Eight *E. coli* fosmid clones were multiplexed, from two 384-well microtiter plates, into one well of a 96-well microtiter plate and incubated overnight. (B) The cells were lysed, thereby providing to access the soluble cell-free extracts (potentially containing the enzymes of interest). (C) Three para-nitrophenol (*p*NP)-linked substrate; *p*NP-α-L-arabinofuranoside, *p*NP-β-D-glucopyranoside and *p*NP-β-D-xylopyranoside were combined and added to each well, and the plates incubated to allow the enzymatic reaction to take place. (D) Positive hits were identified by the development of an intense yellow colour, indicating the enzymatic cleavage of the substrate from the *p*NP chromogen. (E) Individual fosmid clones were singularly inoculated and assayed as depicted in steps A, B and C to identify the clone(s) encoding the enzyme activity. Positive clones were screened with individual *p*NP substrates to determine specific enzyme activities, before being subjected to further analysis, including transposon mutagenesis.

References

Aakvik, T., Degnes, F.K., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Volker., U, Ellingsen., T.E. & Valla, S.(2009). A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS microbiology letters*, 296 (2), 149–58.

Agarwal, A. K. (2007). Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. *Progress in Energy and Combustion Science*, 33 (3), 233–271.

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A. & Levin, D.B. (2011). Biomass pretreatment: Fundamentals toward application. *Biotechnology Advances* 29: 675-685.

Ali, Z., Cousin, S., Fruhling, Brambilla, E., Schumann, P., Yang, Y. and Stackebrandt, E. (2009). Flavobacterium rivuli sp. nov., Flavobacterium subsaxonicum sp. nov., Flavobacterium swingsii sp. nov. and Flavobacterium reichenbachii sp . nov., isolated from a hard water rivulet. *International Journal of Systematic and Evolutionary Microbiology*, 59, 2610–2617.

Almarsdottir, A.R., Sigurbjornsdottir, M.A. & Orlygsson, J. (2012). Effect of various factors on ethanol yields from lignocellulosic biomass by Thermoanaerobacterium AK_{17}. *Biotechnology and bioengineering*, 109(3), 686–94.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215, 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25 (17), 389–402.

Alvira, P., Negro, M. J., & Ballesteros, M. (2011). Effect of endoxylanase and α-L-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw. *Bioresource Technology*, 102 (6), 4552–4558.

Alvira, P., Tomás-Pejó, E., Ballesteros, M., & Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology*, 101 (13), 4851–61.

Amann, R.I., Luwig, W. & Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation . Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. *Microbiology reviews,* 59 (1), 143-169.

Amigun, B, Sigamoney, R., & Von Blottnitz, H. (2008). Commercialisation of biofuel industry in Africa: A review. *Renewable and Sustainable Energy Reviews*, 12 (3), 690–711.

Amigun, Bamikole, Musango, J. K., & Stafford, W. (2011). Biofuels and sustainability in Africa. *Renewable and Sustainable Energy Reviews*, 15(2), 1360–1372.

Amore, A., Amoresano, A., Birolo, L., Henrissat, B., Leo, G., Palmese, A. & Faraco, V. (2012). A family GH51 α-L-arabinofuranosidase from Pleurotus ostreatus: identification, recombinant expression and characterization. *Applied Microbiology and Biotechnology*, 94 (4), 995–1006.

Angelov, A., Mientus, M., Liebl, S. & Lieb, W. (2009). A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Systematic and Applied Microbiology*, 32 (3), 177–85.

Antizar-Ladislao, B. & Turrion-Gomez, J.L. (2008). Second-generation biofuels and local bioenergy systems. *Biofuels, Bioproducts and Biorefining*, 2, 455–469.

Bastien, G., Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., Noirot, C., Dumon, C., and O'Donohue, M. (2013). Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics. *Biotechnology for Biofuels*, 6(1), p.78.

Bayer, E. A., Chanzy, H., Lamed, R., & Shohamy. (1998). Cellulose, cellulases and cellulosomes. *Current opinion in structural biology*, 8, 548-557.

Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G. & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design & Selection*, 17 (4), 49–56.

Bendtsen, J.D., Kiemer, L., Fausboll., A. & Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiology*, 5 (58), 1-13.

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K.M. & Sani, R.K. (2013). Bioresource Technology Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology*, 128, 51–759.

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, 248–54.

Buschke N., Schroder H., & Wittmann C. (2011). Metabolic engineering of *Corynebacterium glutamicum* for production of 1,5-diaminopentane from hemicellulose. *Biotechnology Journal*, 6, 306-317.

Caffall, K.H. & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*, 344 (14), 1879–1900.

Canakci, S., Kacagan, M., Inan, K., Belduz, A.O. and Saha, B.C. (2008). Cloning, purification, and characterization of a thermostable alpha-L-arabinofuranosidase from Anoxybacillus kestanbolensis AC26Sari. *Applied Microbiology and Biotechnology*, 81 (1), 61–68.

Canakci, S., Belduz, A., Saha, B.C., Yasar, A., Ayaz, F.A. and Yayli, N. (2007). Purification and characterization of a highly thermostable alpha-L-Arabinofuranosidase from Geobacillus caldoxylolyticus TK4. Applied Microbiology and Biotechnology, 75 (4), 813–820.

Cartmell, A., Mckee, L.S., Pen, M.J., Larsbrink, J., Brumer, H., Kaneko, S., Ichinose, H., Lewis, R.J., Vikso-Nielsen, A., Gilbert, H.J. & Marles-Wright, J. (2011). The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *The Journal of biological chemistry*, 286 (17), 15483–51495.

Chandel, A. K., Es, C., Rudravaram, R., Narasu, M. L., Rao, V. & Ravindra, P. (2007). Economics and environmental impact of bioethanol production technologies : an appraisal, 2, 14–32.

Chang, C., Ng, C.C., Wang, C.Y. & Shyu, Y.T. (2009). Activity of Cellulase from Thermoactinomycetes and Bacillus spp. Isolated from Brassica Waste Compost. *Scientia Agricola*, 66 (3), 304–308.

Chang, T. & Yao, S. (2011). Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives. *Applied Microbiology and Biotechnology*, 92, 13-27.

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

Charbonneau, D.M., Mouelhi, F.M., Boissinot, M., Sirois, M. & Beauregard, M. (2012). Identification of thermophilic bacterial strains producing thermotolerant hydrolytic enzymes from manure compost. *Indian Journal of Microbiology*, 52 (1), 41–7.

Chen, F. & Dixon, R. A. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology*, 25 (7), 759–761.

Chen, H., & Qiu, W. (2010). Key technologies for bioethanol production from lignocellulose. *Biotechnology Advances*, 28 (5), 556–62.

Christakopoulos, P., Koullas, D. P., Kekos, D., Koukios, E. G. & Macris, B.J. (1991). Direct ethanol conversion of pretreated straw by Fusarium oxysporum. *Bioresource Technology*, 35 (3), 297-300.

Chu, X. Haoze, H., Guo, C. & Sun, B. (2008). Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Applied Microbiology and Biotechnology*, 80 (4), 615–25.

Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., & Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annual review of chemical and biomolecular engineering*, 2, 121–45.

Chung, E.J., Chung, E.J., Lim, H.K., Kim, J.C., Choi, G.J., Park, E.J., Lee, M.H., Chung, Y.R., & Lee, S.W. (2008). Forest soil metagenome gene cluster involved in antifungal activity expression in Escherichia coli. *Applied and Environmental Microbiology*, 74 (3), 723–30.

Claassen, P. A. M., Van Lier, J. B., Lopez Contreras, a. M., Van Niel, E. W. J., Sijtsma, L., Stams, a. J. M., De Vries, S. S., Weusthuis, R.A.(1999). Utilisation of biomass for the supply of energy carriers. *Applied Microbiology and Biotechnology*, 52 (6), 741–755.

Cowan, D. & Fernandez-Lafuente, R. (2011). Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization. *Enzyme and Microbial Technology*, 49(4), pp.326–46.

Dagert M. and Ehrlich S.D. (1979) Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene*. 6 (1): 23-28.

Damasio, A.R.D.L., Pessela, B.C., Segato, F., Prade, R.A., Guisan, J.M. & Polizeli, M. D. L. T.M. (2012). Improvement of fungal arabinofuranosidase thermal stability by reversible immobilization. *Process Biochemistry*, 47 (12), 2411–2417.

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

De Long, E.F. (2009). The microbial ocean from genomes to biomes. *Nature*, 459, 200–6.

Debeche, T.N, Cummings,N., Connerton, I., Debeire, P. & O'Donohue, M.J. (2000). Genetic and Biochemical characterization of a highly thermostable of α-L-arabinofuranosidase from Thermobacillus xylanilyticus. *Applied Environmental and Microbiology*, 66, 1734–1736

Delgenes, J. P., Moletta, R., & Navarro, J. M. (1996). Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by Saccharomyces cerevisiae, Zymomonas mobilis, Pichia stipitis, and Candida shehatae. *Enzyme and Microbial Technology*, 19 (3), 220–225.

Demaneche, S., David, M.M., Navarro, E., Simonet, P. & Vogel, T.M. (2009). Evaluation of functional gene enrichment in a soil metagenomic clone library. *Journal of Microbiological Methods*, 76 (1), 105–107.

Demirbas, A. (2005). Bioethanol from Cellulosic Materials: A Renewable Motor Fuel from Biomass. *Energy Sources*, 27 (4), 327–337.

Demirbas, A. (2009). Political, economic and environmental impacts of biofuels: A review. *Applied Energy*, 86 (1), 108–117.

Dos Santos, C. R., Squina, F. M., Navarro, A. M., Oldiges, D. P., Leme, A. F. P., Ruller, R., Mort, A. J., Prade, R., Murakami, M.T. (2011). Functional and biophysical characterization of a

hyperthermostable GH51 α-L-arabinofuranosidase from Thermotoga petrophila. *Biotechnology Letters*, 33 (1), 131–137.

Dougherty, M.J., Dhaeseleer, P., Hazen, T.C., Simmons, B.A., Adams, P.D. & Hadi, M. Z. (2012). Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnology*, 12, 38.

Ebrahimi, M., Lakizadeh, A., Agha-Golzadeh, P., Ebrahimie, E. & Ebrahimi, M. (2011). Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PloS one*, 6 (8), e23146.

Ekkers, D.M., Cretoiu, S.M., Kielak, A.M. & Van Elsas, J.D. (2012). The great screen anomaly--a new frontier in product discovery through functional metagenomics. *Applied microbiology and Biotechnology*, 93 (3), 1005–20.

Elcock, A.H. (1998) The stability of salt bridges at high temperatures: implication for hyperthermophilic proteins. *Journal of Molecular Biotechnology*, 284 (2), 489-502.

Fatih Demirbas, M. (2009). Biorefineries for biofuel upgrading: A critical review. *Applied Energy*, 86 (1), 151–161.

Feller, G. (2013). Psychrophilic Enzymes: From Folding to Function and Biotechnology. *Scientifica*, 20 (13), 1-28.

Feng, Y., Duan,C.J., Pang, H., Mo, X.C., Wu, C.F., Yu, Y., Hu, Y.L., Wei, J., Tang, J.L. & Feng, J.X. (2007). Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Applied Microbiology and Biotechnology*, 75 (2), 319–328.

Fong, J. C. N., Svenson, C. J., Nakasugi, K., Leong, C. T. C., Bowman, J. P., Chen, B., Glenn, D. R., Nielan, B.A. & Rogers, P.L. (2006). Isolation and characterization of two novel ethanol-

tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles : life under extreme conditions*, 10 (5), 363–72.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chrisholm, S.W. & DeLong, E.F. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (10), 3805–3810.

Fritz, M., Ravanal, M.C., Braet, C. & Eyzaguirre, J. (2008). A family 51 alpha-l-arabinofuranosidase from Penicillium purpurogenum: purification, properties and amino acid sequence. *Mycological Research*, 112 (8), 933–942.

Fujimoto, Z., Ichinose, H. & Kaneko, S. (2008). Crystallization and preliminary crystallographic analysis of exo-alpha-1,5-L-arabinofuranosidase from Streptomyces avermitilis NBRC14893. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 64 (11), 1007–1009.

Gabor, E.M., Alkema, W.B.L. & Janssen, D.B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, 6(9), 879-86.

Galbe, M. & Zacchi, G. (2007). Pretreatment of lignocellulosic materials for efficient bioethanol production. *Biotechnology Advances,* 108, 41-65.

Galbe, M. & Zacchi, G. (2012). Pretreatment: The key to efficient utilization of lignocellulosic materials. *Biomass and Bioenergy*, 46, 70–78.

Georgieva, T. I., & Ahring, B. K. (2007). Evaluation of continuous ethanol fermentation of dilute-acid corn stover hydrolysate using thermophilic anaerobic bacterium Thermoanaerobacter BG1L1. *Applied Microbiology and Biotechnology*, 77 (1), 61–8.

Ghai, R., Cuadrado, A.B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., Verdu, M., Deschamps, P., Moreira, D., Lopez-Garcia, P., Mira, A. & Rodriguez-Valera, F. (2010). Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *The ISME Journal*, 4 (9), 1154–66.

Gloux, K., Leclerc, M., Iliozer, H., Haridon, R.L., Manichanh, C., Corthier, G., Nalin, R., Blottie`re, H.M. and Dore., J. (2007). Development of high-throughput phenotyping of metagenomic clones from the human gut microbiome for modulation of eukaryotic cell growth. *Applied and Environmental Microbiology*, 73 (11), 3734–3737.

Goncalves, T., Damasio, A.R.L., Segato, F., Alvarez, T.M., Bragatto, J., Brenelli, L.B., Citadini, A.P.S., Murakami, M.T., Ruller, R., Paes Leme, A.F., Prade, R.A. & Squina, F.M. (2012). Functional characterization and synergic action of fungal xylanase and arabinofuranosidase for production of xylooligosaccharides. *Bioresource Technology*, 119, 93–9.

Gong, X., Gruniniger, R.J., Forster, R.J., Teather, R.M. & McAllister, T.A. (2013). Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Applied Microbiology and Biotechnology*, 97 (6), 2423–2431.

Goyal, S., Dhull, S.K. & Kapoor, K.K. (2005). Chemical and biological changes during composting of different organic wastes and assessment of compost maturity. *Bioresource Technology*, 96, 1584–1591.

Gray, K. A, Zhao, L., & Emptage, M. (2006). Bioethanol. *Current Opinion in Chemical Biology*, 10 (2), 141– 6.

Guan, C., Ju, J., Borlee, B.R., Williamson, L.L., Shen, B., Raffa, K.F. & Jo Handelsman, J. (2007). Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Applied and Environmental Microbiology*, 73 (11), 3669–3676.

Gupta, R., Mehta, G. & Chander Kuhad, R. (2012). Fermentation of pentose and hexose sugars from corncob, a low cost feedstock into ethanol. *Biomass and Bioenergy*, 47, 334–341.

Gupta, R., Sharma, K. K. & Kuhad, R. C. (2009). Separate hydrolysis and fermentation (SHF) of Prosopis juliflora, a woody substrate, for the production of cellulosic ethanol by Saccharomyces cerevisiae and Pichia stipitis-NCIM 3498. *Bioresource Technology*, 100 (3), 1214–1220.

Hahn-Hägerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I., & Gorwa-Grauslund, M. F. (2007). Towards industrial pentose-fermenting yeast strains. *Applied Microbiology and Biotechnology*, 74 (5), 937–53.

Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms Metagenomics : Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*,68 (4), 669-685.

Handelsman, J. (2007). Metagenomics and Microbial Communities. *Encylcopedia of Life Sciences*, Wiley & Sons, University of Madison-Wisconsin, USA.

Hashimoto, K.H., Oshida, M.Y. & Asumi, K.H. (2011). Isolation and Characterization of CcAbf62A , a GH62 - L -Arabinofuranosidase , from the Basidiomycete Coprinopsis cinerea. *Bioscience, Biotechnology and Biochemistry*, 75 (2), 342–345.

Hasunumaa, T. & Kondo, A. (2012). Consolidated bioprocessing and simultaneous saccharification and fermentation of lignocellulose to ethanol with thermotolerant yeast strains. *Process Biochemistry*.

Havlík, P., Schneider, U. a, Schmid, E., Böttcher, H., Fritz, S., Skalský, R., Aoki, K., De Cara, S., Kindermann, G., Kraxner, F., Leduc, S., McCullam, I., Mosnier, A., Sauer, T. & Obersteiner, M. (2010). Global land-use implications of first and second generation biofuel targets. *Energy Policy*, 39 (10), 5690–5702.

Heinze,T., Schwikal,K. & Barthel, S. (2005) Ionic liquids as reaction medium in cellulose functionalization. *Macromolecular Bioscience* 5 (6), 520-525.

Helm, C. V, Magalhães, W. L. E., De Lima, E. a, Silva, P. R., Hoffmann, K., Higa, A., & Mendes, D. (2011). Time influence in the enzymatic saccharification of cellulose pulp samples. *BMC Proceedings*, 5 (7), 112-113.

Hendriks, A.T.W.M. & Zeeman, G. (2009). Bioresource Technology Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology*, 100, 10–18.

Henrissat, B & Davies, G.J. (2000). Glycoside hydrolases andglycosyltransferases. Families, modules, and implications for genomics. *Plant Physiology*, 124, 1515-1519.

Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., & Foust, T. D. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315, 804–807.

Hoffman, Z.B., Oliveira, L.C., Cota, J., Alvarez., T.M., Diogo, J.A., Neto, M.D.O., Citadini, A.P.S., Leite, V.B.P., Squina, F.M., Murakami, M.T. & Ruller, R. (2013). Characterization of a Hexameric Exo-Acting GH51 α-L-Arabinofuranosidase from the Mesophilic Bacillus subtilis. *Molecular Biotechnology*. 1-8.

Hove-jensen, B., Rosenkrantz, T.J. & Wanner, B.L. (2003). Escherichia coli phnN , Encoding Ribose (Phosphoribosyl Diphosphate Forming): Dual Role in Phosphonate Degradation and NAD Biosynthesis Pathways Escherichia coli phnN , Encoding Ribose 1 , 5-Bisphosphokinase Activity (Phosphoribosyl Diphosphate Forming): Dual Role in Phosphonate Degradation and NAD Biosynthesis Pathways. *Journal of Bacteriology*, 185 (9), 2793–2801

Hovel, K., Shallom, D., Niend, K., Belakhov, V., Shoham, G., Baasov, T., Shoha, Y. and Schomburg, D. (2003). Crystal structure and snapshots along the reaction pathway of a family 51 α -L-arabinofuranosidase. The EMBO Journal, 22 (19), 4922-4932.

Howard, R. L., Abotsi, E., Jansen van Rensburg, E.L. & Howard, S. (2003). Lignocellulose biotechnology : issues of bioconversion and enzyme production. *Journal of Biotechnology*, 2, 602–619.

Hu, Y., Zhang, G., Li, A., Chen, J. & Ma, L. (2008). Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Applied Microbiology and Biotechnology*, 80, 823-830.

Ilmberger, N., Meske, D., Juergensen,J., Schulte, M., Barthen, P., Rabausch, U., Angel Angelov, A., Mientus, M., Liebl, W., Schmitz,R.A. & Streit, W.R. (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids--linking thermostability and halotolerance. *Applied Microbiology and Biotechnology*, 95 (1), 135–146.

Im, D.H., Kimura, K.I., Hayashaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S.I., Miyazaki, K., Wakagi, T. & Fushinobu, S. (2012). Crystal Structures of Glycoside Hydrolase Family 51 α-L-Arabinofuranosidase from Thermotoga maritima. *Bioscience, Biotechnology, and Biochemistry*, 76 (2), 423–428.

Inacio, J.M., Correia, I.L. & de Sá-Nogueira, I. (2008). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in Bacillus subtilis. *Microbiology*, 154 (9), 2719–29.

Iyer, P. V. & Ananthanarayan, L. (2008). Enzyme stability and stabilization—Aqueous and non-aqueous environment. *Process Biochemistry*, 43 (10), 1019–1032.

James .D. (2001) Chromas version 2.01. Department of microbiology, University of Cape Town

Jendresen, C.B., Martinussen, J. & Kilstrup, M. (2012). The PurR regulon in Lactococcus lactis - transcriptional regulation of the purine nucleotide metabolism and translational machinery. *Microbiology*, 158 (8), 2026–38.

Jeon, J.H., Kim, J.T., Kim, Y.J., Kim, H.K., Lee, H.S., Kang, S.G., Kim, S.J. & Jung-Hyun Lee, J.H. (2009). Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Applied microbiology and biotechnology*, 81 (5), 865–874.

Jones, M.V., Spencer, W.N. and Edwards, C. (1984). Temperature-dependent Azide Sensitivity of Growth and ATPase Activity in the Facultative Thermophile , Bacillus cougulans. *Journal of General Microbiology*, 130, 95–101.

Kadar, Z., Szengyel, Z., & Réczey, K. (2004). Simultaneous saccharification and fermentation (SSF) of industrial wastes for the production of ethanol. *Industrial Crops and Products,* 20 (1), 103–110.

Kakirde, K.S., Wild, J., Godiska, R., Mead, D.A., Wiggins, A.G., Goodman, R.M., Szybalski, W. & Liles, M.R. (2012). Gram negative shuttle BAC vector for heterologous expression

Kang, C.H., Oh, K.H., Lee, M.H., Oh, T.K., Kim, B.H. and Yoon, J.H. (2011). A novel family VII esterase with industrial potential from compost metagenomic library. *Microbial cell factories*, 10 (1), 41.

Karakashev, D., Thomsen, A. B., & Angelidaki, I. (2007). Anaerobic biotechnological approaches for production of liquid energy carriers from biomass. *Biotechnology Letters*, 29 (7), 1005–1012.

Karmakar, M., & Ray, R.R. (2011). Current Trends in Research and Application of Microbrial Cellulases. *Research Journal of Microbiology*, 6 (1), 41-53.

Kazimierczak, K. A., Scott, K.P., Kelly, D. & Aminov, R.I. (2009). Tetracycline resistome of the organic pig gut. *Applied and Environmental Microbiology*, 75 (6), pp.1717–1722.

Khan, M. A. (2010). Master's Thesis: Hydrolysis of Hemicellulose by Commercial Enzyme Mixtures Hydrolysis of Hemicellulose by Commercial Enzyme Mixtures.

Kim, E.Y., Oh, K.H., Lee, M.H., Kang, C.H., Oh, T.K. & Yoon, J.H. (2009). Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Applied and Environmental Microbiology*, 75 (1), 257–260.

Kim, J.-H., Block, D. E., & Mills, D. A. (2010). Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. *Applied microbiology and biotechnology*.

Koenig, J.E., Sharp, C., Dlutek, M., Curtis, B., Joss, M., Boucher, Y. & Ford Doolittle, W. (2009). Integron Gene Cassettes and Degradation of Compounds Associated with Industrial Waste: The Case of the Sydney Tar Ponds. *PLoS ONE*, 4 (4), 1-9.

Koh, L. P., & Ghazoul, J. (2008). Biofuels, biodiversity, and people: Understanding the conflicts and finding opportunities. *Biological Conservation*, 141 (10), 2450–2460.

Koseki, T., Okuda, M., Sudoh, S., Kizaki, Y., Iwano, K., Aramaki, I. & Matsuzawa, H.(2003). Role of Two a-L-Arabinofuranosidases in Arabinoxylan Degradation and Characteristics of the Encoding Genes from Shochu Koji Molds *Aspergillus kawachii* and *Aspergillus awamori*. *Journal of Bioscience and Bioengineering*, 96 (3), 232–241.

Kota, S., Kumar, C.V. & Misra, H.S. (2010). Characterization of an ATP-regulated DNA processing enzyme and thermotolerant phosphoesterase in the radioresistant bacterium Deinococcus radiodurans. *The Biochemical journal*, 431 (1), 149–57.

Kouker, G. and Jaeger, K. E. (1987). Specific and sensitive plate assay for bacterial lipases. *Applied and Environmental Microbiology*, 53, 211–213.

Krasikov, V.V., Karelov, D.V. & Firsov, L.M. (2001). α-Glucosidases. *Biochemistry*, 66 (3), 267-281

Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science*, 324, 255-277.

Kumar, P., Barrett, D. M., Delwiche, M. J., & Stroeve, P. (2009a). Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. *Industrial & Engineering Chemistry Research*, 48 (8), 3713–3729.

Kumar, R., Mago, G., Balan, V., & Wyman, C. E. (2009b). Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. *Bioresource Technology*, 100 (17), 3948–62.

Kumar, S., Singh, S. P., Mishra, I. M., & Adhikari, D. K. (2009c). Recent Advances in Production of Bioethanol from Lignocellulosic Biomass. *Chemical Engineering & Technology*, 32 (4), 517–526.

Kumar, S., Tsai, C. & Nussinov, R., (2000). Factors enhancing protein thermostability. *Protein Engineering*, 13 (3), 179–191.

Kuo, C.-H., & Lee,C.-K. (2009). Enhanced enzymatic hydrolysis of sugarcane bagasse by N-methylmorpholine-N-oxide pretreatment. *Bioresource Technology*, 100 (2), 866–71.

Kwon, E.J., Jeong, Y.S., Kim,Y.H., Kim, S.K., Na, H.B., Kim, J., Yun, H.D. & Kim, H. (2010). Construction of a Metagenomic Library from Compost and Screening of Cellulase- and Xylanase-positive Clones. *Journal of the Korean Society for Applied Biological Chemistry*, 53 (6), 702–708.

Lakizadeh, A., Agha-Golzadeh, P. & Ebrahimi, M. (2011). Engineering Thermostable Enzymes ; Application of Unsupervised Clustering Algorithms. *Advance Studies in Biology*, 3 (2), 63–78.

Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. & Ruppa, S. (2007). Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology*, 127 (4), 575–592.

Langridge, G.C., Phan, M-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J. & Turner, A.K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, 19 (2), 2308-2316.

Lee C.M., Yeo Y.S., Lee J.H., Kim S.J., Kim J.B., Han N.S., Koo B.S. & Yoon S.H. (2008). Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome. *Biochemical and Biophysical Research Communications* 370, 322-326.

Lee, H.S., Kwon, K.K., Kang, S.G., Cha, S.S., Kim, S.J. & Lee, J.H. (2010) Approaches for novel enzyme discovery from marine environments. *Current Opinion in Biotechnology*, 21 (3), 353–357.

Lee, W.-J., Kim, M.-D., Ryu, Y.-W., Bisson, L. F., & Seo, J.-H. (2002). Kinetic studies on glucose and xylose transport in Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology*, 60 (1-2), 186-191.

Lim, Y.R, Yoon, R.Y., Seo, E.S., Kim, Y.S., Park, C.S. & Oh, D.K. (2010). Hydrolytic properties of a thermostable α-L-arabinofuranosidase from Caldicellulosiruptor saccharolyticus. *Journal of Applied Microbiology*, 109 (4), 1188–97.

Limayem, A. & Ricke, S.C. (2012). Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science*, 38 (4), 449–467.

Lynd, R., Currie, D., Ciazza, N., Herring, C. & Orem, N. (2008). Consolidated bioprocessing of cellulosic biomass to ethanol using thermophilic bacteria. *Bioenergy*, 55–73.

Lynnon Biosoft (1999) DNAMAN version 4.13. Department of microbiology, University of Cape Town, South Africa

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S. Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D. & Bryant, S.H. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research*, 33, 192–196.

Mathews, C.K., Van Holde, K.E. & Ahern, K.G. (2000). "Biochemistry 3rd edition". Addison-Wesley, San Francisco.

Matsumura, K., Obata, H., Hata, Y., Kawato, A., Abe, Y., & Akita, O. (2004). Isolation and characterization of a novel gene encoding alpha-L-arabinofuranosidase from Aspergillus oryzae. *Journal of Bioscience and Bioengineering*, 98 (2), 77–84.

Maynard, J., Adams, E.J., Krogsgaard, M., Petersson, K., Corey, W., Liu, K. & Garcia, C. (2005). High-level bacterial secretion of single-chain alphabeta T-cell receptors. *Journal of Immunological Methods*, 306, 51–67.

Mayumi, D. & Akutsu-shigeno, Y. (2008). Identification and characterization of novel poly (DL -lactic acid) depolymerases from metagenome. *Applied Microbiology and Biotechnology*, (79), 743–750.

Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura., N. & Nakajima-Kambe, T. (2008). Identification and characterization of novel poly (DL-lactic acid) depolymerases from metagenome. *Applied Microbiology and Biotechnology*, 79 (5), 743–750.

McKee, L.S., Pena, M.J., Rogowski, A., Jackson, A., Lewis, R.J., York, W.S., Krogh, K.B.R.M., Vikso-Nielsen, A., Skjot, M., Gilbert, H.J. & Marles-Wright, J. (2012). Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (17), 6537–6542.

McMillan, J.D., 1994. Pretreatment of lignocellulosic biomas: Enzymatic Conversion of Biomass for Fuels Production. *American Chemical Society*. 292–324.

Meiswinkel, T.M., Gopinath, V., Lindner, S.N., Nampoothiri, M. & Wendisch, V.F. (2012). Accelerated pentose utilization by Corynebacterium glutamicum for accelerated production of lysine , glutamate , ornithine and putrescine. *Microbial Biotechnology*, 6, 131-140.

Miller,G.L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry* ,31, 426–428

Mirete, S., de Figueras, C.G. & González-Pastor, J.E. (2007). Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. *Applied and Environmental Microbiology*, 73 (19), 6001–6011.

Miyazaki, K. (2005). Hyperthermophilic alpha-L-arabinofuranosidase from Thermotoga maritima MSB8: molecular cloning, gene expression, and characterization of the recombinant protein. *Extremophiles life under extreme conditions*, 9 (5), 399–406.

Moore, D., Robson, G.D., Trinci, A.P.J (2011). *21st Century Guidebook to Fungi*. Cambridge University Press: ISBN: 9780521186957.

Mori, T., Mizuta, S., Suenaga, H. & Kentaro Miyazaki, K. (2008). Metagenomic screening for bleomycin resistance genes. *Applied and Environmental Microbiology*, 74 (21), 6803–6805.

Morimoto, S. & Fujii, T.(2009). A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Applied Microbiology and Biotechnology*, 83, 389–396.

Mu, X., Sun, W., Liu, C., & Wang, H. (2011). Improved efficiency of separate hexose and pentose fermentation from steam-exploded corn stalk for butanol production using Clostridium beijerinckii. *Biotechnology Letters*. 5. 44-53.

Nacke, H., Engelhaupt, M., Brady. S., Fischer, C., Tautzt, J. & Daniel, R. (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnology Letters*, 34 (4), 663–75.

Nacke, H., Engelhaupt., M, Brady, S., Fischer, C., Tautzt, J. & Daniel, R. (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnology Letters*, 34 (4), 663–675.

Naik, S. N., Goud, V. V, Rout, P. K., & Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 14 (2), 578–597.

Naumoff, D G. (2011a). Hierarchical classification of glycoside hydrolases. *Biochemistry*, 76 (6), 622–635.

Naumoff, D. G. (2011b). GHL1-GHL15: New families of the hypothetical glycoside hydrolases. *Molecular Biology*, 45 (6), 983–992.

Niemi, R.M., Heiskanen, I., Wallenius, K. and Lindstrom, K. (2001). Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *Journal of Microbiological Methods*, 45, 155–165.

Nigam, P.S. (2013). Microbial enzymes with special characteristics for biotechnological applications. *Biomolecules*, 3 (3), 597–611.

Numan, M.T. & Bhosle, N.B. (2006). Alpha-L-arabinofuranosidases: the potential applications in biotechnology. *Journal of Industrial Microbiology & Biotechnology*, 33 (4), 247–60.

Nurcholis, M., Nurhayati, N., Helianti, I., Ulfah, M., Wahyuntari, B. & Wardani, A.K. (2012). Cloning of α-L-arabinofuranosidase Genes and Its Expression in Escherichia coli: A Comparative Study of Recombinant Arabinofuranosidase Originatingin Bacillus subtilis DB104 and Newly Isolated Bacillus licheniformis CW1. *Microbiology Indonesia*, 6 (1), 1–8.

Ohgren, K., Bura, R., Lesnicki, G., Saddler, J., & Zacchi, G. (2007). A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover. *Process Biochemistry*, 42 (5), 834–839.

Ohta, K., Fujii, S. & Higashida, C. (2013). Characterization of a glycoside hydrolase family-51 α-l-arabinofuranosidase gene from Aureobasidium pullulans ATCC 20524 and its encoded product. *Journal of Bioscience and Bioengineering*, 116 (3), 287–92.

Okano, K., Kitagawa, M., Sasaki, Y., & Watanabe, T. (2005). Conversion of Japanese red cedar (Cryptomeria japonica) into a feed for ruminants by white-rot basidiomycetes. *Animal Feed Science and Technology*, 120 (3-4), 235–243.

Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y. & Tsuda, M. (2007). Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Applied Microbiology and Biotechnology*, 74 (2), 501–510.

Pace, C.N., Alston, R.W., Shaw, K.L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Science*, 9 (7), 1395-1398.

Pae, G., Skov, L.K., O'Donohue, M.J., Remond, C., Kastrup, J.S., Gajhede, M. & Mirza, O. (2008). The Structure of the Complex between a Branched Pentasaccharide and Thermobacillus xylanilyticus GH-51 Arabinofuranosidase Reveals Xylan-Binding Determinants and Induced Fit, *Biochemistry*, 47, 7441–7451.

Palackal, N., Lyon, C.S., Zaidi, S., Luginbühl, P., Dupree, P., Goubet, F., Macomber, J.L., Short, J.M., Hazlewood, G.P., Robertson, D.E. & Steer, B.A. (2007). A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Applied Microbiology and Biotechnology*, 74 (1), 113–124.

Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L. & Feng, J.X. (2009). Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Current Microbiology*, 58 (4), 404-408.

Park, H.J. & Kim, D. (2013). Draft Genome Sequence of a Humic Substance-Degrading Paenibacillus sp . Isolated from the Subarctic Grasslands at Low. *Genome Announcements*, 1 (1).

Peng, F., Ren, J.-L., Xu, F., Bian, J., Peng, P. & Sun, R.C. (2009). Comparative Study of Hemicelluloses Obtained by Graded Ethanol Precipitation from Sugarcane Bagasse. *Journal of Agricultural and Food Chemistry*, 57, 6305-6317.

Quaiser, A. Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., Eck, J. & Schleper, C. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain : evidence from environmental genomics. *Molecular Microbiology*, 50 (2), 563–575.

Quiroz-castañeda, R.E. & Folch-mallol, J.L. (2013). Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars. *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*, 119-155.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. a, Frederick, W. J., Hallet, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R. & Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. *Science*, 311, 484–9.

Ranjan, R., Grover, A., Kapardar, R.K., & Sharma, R. (2005). Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochemical and Biophysical Research Communications* 335, 57–65.

Remond, C., Boukari, I., Chambat, G. & O'Donohue, M. (2008). Action of a GH 51 α-l-arabinofuranosidase on wheat-derived arabinoxylans and arabino-xylooligosaccharides. *Carbohydrate Polymers*, 72 (3), 424–430.

Reysenbach, A.L., Liu, Y., Banta, A.M., Beveridge, T.J., Kirshtein, J.D., Schouten, S., Tivey, M.K., Von Damm, K.L. & Voytek, M.A. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature*, 442 (7101), 444–447.

Riaz, K., Elmerich, C., Moreira, D., Raffoux, A., Dessaux, Y. & Faure, D. (2008). A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases. *Environmental Microbiology*, 10 (3), 560–70.

Riesenfeld, C.S., Goodman, R.M. & Handelsman, J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Enivronmental Microbiology*, 6 (9), .981–989.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*, 454, 841–845.

Ruiz, L., Motherway, M.O., Lanigan, N. & Van Sinderen, D. (2013). Transposon mutagenesis in Bifidobacterium breve: construction and characterization of a Tn5 transposon mutant library for Bifidobacterium breve UCC2003. *PloS ONE*, 8 (5), 1-10.

Saha, B C. (2001). Xylanase from a newly isolated Fusarium verticillioides capable of utilizing corn fiber xylan. *Applied Microbiology and Biotechnology*, 56 (5-6), 762–766.

Saha, B.C.(2000). α-L-Arabinofuranosidase, biochemistry, molecularbiology, and application in biotechnology. *Advance Biotechnology*, 18,403–423

Saha, Badal C. (2003). Hemicellulose bioconversion. *Journal of industrial microbiology biotechnology*, 30 (5), 279–291.

Sakamoto, T., Ogura, A. & Inui, M., (2011). Identification of a GH62 α -L-arabinofuranosidase specific for arabinoxylan produced by Penicillium chrysogenum. *Applied Microbiology and Biotechnology*, 90, 137–146.

Sambrook, J., & Russell, D.W. (2001). *Molecular Cloning*. Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sasson, A. (2012). Food security for Africa : an urgent global challenge, 1–16.

Schaffer, H.E. and Sederoff, R.R. (1981) Improved estimation of DNA fragment lengths from agarose gels. *Analytical Biochemistry*, 115 (1), 113-122

Scheller, H. V. & Ulvskov, P. 2010. Hemicelluloses. Annual Review of Plant Biology, 61, 263-289.

Schipper, C., Hornung, C., Bijtenhoorn, P., Quitschau, M., Grond, S. & Streit, W.R. (2009). Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in Pseudomonas aeruginosa. *Applied and Environmental Microbiology*, 75 (1), 224–233.

Schmeisser, C., Steele, H. & Streit, W.R. (2007). Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology* 75, 955–962.

Sedlak, M. & Ho, N. W. Y. (2004). Production of ethanol from cellulosic biomass hydrolysates using genetically engineered Saccharomyces yeast capable of cofermenting glucose and xylose. *Applied Biochemistry and Biotechnology*, 113-116, 403–16.

Seidman, C.E., Struhl, K., Sheen, J. & Jessen, T. (1997). Introduction plasmid DNA into Cells. *Current Protocols in Molecular Biology*, 1.8.1-1.8.10.

Shafiee, S. & Topal, E. (2010). A long-term view of worldwide fossil fuel prices. *Applied Energy*, 87 (3), 988–1000.

Shallom, D. & Shoham, Y. (2003). Microbial hemicellulases. *Current Opinion in Microbiology*, 6 (3), 219–228.

Shi, P., Chen, X., Meng, K., Huang, H., Bai, Y., Luo, H., Yang, P. and Yao, B. (2013). Distinct actions by Paenibacillus sp. strain E18 α-L-arabinofuranosidases and xylanase in xylan degradation. *Applied and Environmental Microbiology*, 79 (6), 1990-1995.

Shi, P., Li, N., Yang, P., Wang, Y., Luo, H., Bai, Y. & Yao, B. (2010). Gene cloning, expression, and characterization of a family 51 alpha-L-arabinofuranosidase from Streptomyces sp. S9. *Applied Biochemistry and Biotechnology*, 162 (3), 707–18.

Simon, C. & Daniel, R. (2011). Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology*, 77, 1153-1161.

Simon, C., Herath, J., Rockstroh, S. and Daniel, R. (2009). Rapid Identification of Genes Encoding DNA Polymerases by Function-Based Screening of Metagenomic Libraries Derived from Glacial Ice. *Applied and Environmental Micrbiology*, 75 (9), 2964–2968

Sims, R. E. H., Mabee, W., Saddler, J. N. & Taylor, M. (2010). An overview of second generation biofuel technologies. *Bioresource Technology*, 101 (6), 1570–1580.

Skovgaard, P.A. & Jørgensen, H., (2013). Influence of high temperature and ethanol on thermostable lignocellulolytic enzymes. *Journal of Industrial Microbiology & Biotechnology*, 40 (5), 447–456.

Sorek, R. & Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature Reviews| Genetics*, 11, 9-16.

Sorensen, H.R., Jorgensen, C.T., Hansen, C.H., Jorgensen, C.I., Pedersen, S. & Meyer, A.S. (2006). A novel GH43 alpha-L-arabinofuranosidase from Humicola insolens: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. *Applied Microbiology and Biotechnology*, 73 (4), 850–61.

Souza, T. A.C.B., Santos, C.R., Souza, A.R., Oldiges, D.P., Ruller, R., Prade, R.A., Squina, F.M. & Murakami, M.T. (2011). Structure of a novel thermostable GH51 α-L-arabinofuranosidase from Thermotoga petrophila RKU-1. *Protein science : a publication of the Protein Society*, 20 (9), 1632–1637.

Srinivasan, S. (2009). The food v. fuel debate: A nuanced view of incentive structures. *Renewable Energy*, 34 (4), 950–954.

Steele, H. L., Jaeger, K. E., Daniel, R., & Streit, W. R. (2008). Advances in recovery of novel biocatalysts from metagenomes. *Journal of molecular microbiology and biotechnology*, 16, 25-37.

Suenaga, H., Ohnuki, T. & Miyazaki, K. (2007). Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Enviromental Microbiology*, 9 (9), 2289–2297.

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

Sundberg, C. & Smars, S., Jonsson, H. (2004) Low pH as an inhibiting factor in the transition from mesophilic to thermophilic phase in composting, *Bioresource Technology*, 95, 145–150.

Svetlitchnyi, V. A., Kensch, O., Falkenhan,D.A., Korseska, S.G., Lippert, N., Prinz, M., Sassi, J., Schickor, A. & Curvers, S. (2013). Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria. *Biotechnology for Biofuels*, 6 (1), 31-46.

Szijarto, N., Siika-Aho, M., Tenkanen, M., Alapuranen, M., Vehmaanpera, J., Reczey, K. & Talebnia, F., Karakashev, D. & Angelidaki, I. (2010). Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation. *Bioresource Technology*, 101(13), 4744–53.

Taylor, E.J., Smith, N.L., Turkenburg, J.P., D'souza, S., Gilbert, H.J. & Davies, G.J. (2006). Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from Clostridium thermocellum. *The Journal of Biochemistry* 395 (1), 31–7.

Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., & Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (3), 1450–1455.

Taylor, T.J. & Vaisman, I.I. (2010). Discrimination of thermophilic and mesophilic proteins. *BMC Structural Biology*, 10 (1), 1–10.

Teather, R. M., & Wood, P. J. (1982). Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology*, 43, 777-78.

Tin, S., Bizzoco, R.W. & Kelley, S.T. (2011). Role of the terrestrial subsurface in shaping geothermal spring microbial communities. *Environmental Microbiology Reports*, 3 (4), 491–9.

Treangen, T.J. & Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature reviews| Genetics*, 13 (1), 36–46.

Tuffin, M., Anderson, D., Heath, C. and Cowan, D.A. (2009). Metagenomic gene discovery: How have we moved into novel sequence space?. *Biotechnology Journal*, 4 (12), 1671-1683.

Tymowska-lalanne, Z. & Kreis, M. (1998). Expression of the Arabidopsis thaliana invertase gene family. *Planta*, 207, 259-265.

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

Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, 23 (1), 88–93.

Van Dyk, J.S. & Pletschke, B.I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes conversion and synergy. *Biotechnology Advances*, 30 (6), 1458-1480.

Van Hellemond, E.W., Janssen, D.B. & Fraaije, M.W. (2007). Discovery of a novel styrene monooxygenase originating from the metagenome. *Applied and Environmental Microbiology*, 73 (18), 5832–5839.

Venter, J.C., Remington K., Heidelberg, J.F., Halpern, A. L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. & Smith, H.O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.

Verma, D., Kawarabayasi, Y., Miyazaki, K. & Satyanarayana, T.(2013). Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. *PloS ONE*, 8 (1), 1-9.

Vierhuis, E., York, W.S, Kolli, V.S.K., Vincken, J.P., Schols, H.A., Van Alebeek, G.J.W.M. & Voragena, A.G.J. (2001). Structural analyses of two arabinose containing oligosaccharides derived from olive fruit xyloglucan: XXSG and XLSG. *Carbohydrate Research*, 332, 285–297.

Viikari, L. (2007). Hydrolysis of amorphous and crystalline cellulose by heterologously produced cellulases of Melanocarpus albomyces. *Journal of Biotechnology*, 136, 140–147.

Villegas, A., Kropinski, A.M. & Kropinski, A. (2008). An analysis of initiation codon utilization in the Domain Bacteria – concerns about the quality of bacterial genome annotation. *Microbiology*, 154, 2559–2561.

Voget, S., Leggewie, C., Uesbeck, A., Raasch,C., Jaeger, K-E. & Streit, W.R. (2003). Prospecting for Novel Biocatalysts in a Soil Metagenome Prospecting for Novel Biocatalysts in a Soil Metagenome. *Applied and Environmental Microbiology*, 69(10), 6235–6242.

Wagschal, K., Heng, C., Lee, C. C., & Wong, D. W. S. (2009). Biochemical characterization of a novel dual-function arabinofuranosidase/xylosidase isolated from a compost starter mixture. *Applied Microbiology and Biotechnology*, 81 (5), 855–63.

Wagschal, K., Heng, C., Lee, C.C., Robertsons, G.H., Orts, W.J. & Wong, D.W.S. (2008). Purification and Characterization of a Glycoside Hydrolase Family 43 β-xylosidase from Geobacillus thermoleovorans IT-08. *Applied Biochemistry and Biotechnology*, 155, 304–313.

Wan, C., Chen, C-T., Chang, M. D-T., Lo, L-C. & Li, Y-K. (2007). Mutagenesis and mechanistic study of a glycoside hydrolase family 54 α-L-arabinofuranosidase from Trichoderma koningii. *Journal of Biochemistry*, 558, 551–558.

Wang, C.,Meek, D.J.,Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T. & Charles, T.C. (2006). Isolation of Poly-3-Hydroxybutyrate Metabolism Genes from Complex Microbial Communities by Phenotypic Complementation of Bacterial Mutants. *Applied and Environmental Microbiology*, 72 (1), 384–391.

Warren, L., Freeman, J.D., Levesque, R.C., Smailus, D.E., Flibotte, S. & Holt, R.A. (2008). Transcription of foreign DNA in Escherichia coli. *Genome Research* (604), 1798–1805.

Waschkowitz, T., Rockstroh, S. & Daniel, R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Applied and Environmental Microbiology*, 75 (8), 2506–2516.

Watanabe, K., Nagao, N., Yamamoto, S., Toda, T. & Kurosawa, N. (2007). Thermobacillus composti sp. nov., a moderately thermophilic bacterium isolated from a composting reactor. *International Journal of Systematic and Evolutionary Microbiology*, 57 (7), 1473–1477.

Welmes, P., Wexler, M. & Bond, P.L. (2008). Metaproteomics provides functional insight into activated sludge wastewater treatment. *PloS ONE*, 3(3), 1-11.

Wiley, J.M., Sherwood, L.M., Woolverton, C.J. (2008). "Prescott, Harley and Klein's Microbiology seventh edition". McGraw-Hill, 1221 Avenues of the Americas, New York.

Wilkinson, D.E., Jeanicke, T. & Cowan, D.A. (2002). Efficient molecular cloning of environmental DNA from geothermal sediments. *Biotechnology Letters*, 24 (2), 155-161.

Wyman, C. E. (2007). What is (and is not) vital to advancing cellulosic ethanol. *Trends in Biotechnology*, 25 (4), 153–157.

Yan, Q., Tang, L., Yang, S., Zhou, P., Zhang, S. & Jiang., Z. (2012). Purification and characterization of a novel thermostable α-L-arabinofuranosidase (α-L-AFase) from Chaetomium sp. *Process Biochemistry*, 47 (3), 472–478.

Yang, H.Y., Ichinose, H., Nakajima, M., Kobayashi, H. & Kaneko, S. (2006). Synergy Between an a-L-Arabinofuranosidase from Aspergillus oryzae and an Endo-Arabinanase from Streptomyces coelicolor for Degradation of Arabinan. *Food Science and Technology Research*, 12 (1), 43-49.

Yassin, A.F., Hupfer, H., Klenk, H.P. & Siering, C. (2009). Desmospora activa gen. nov., sp. nov., a thermoactinomycete isolated from sputum of a patient with suspected pulmonary tuberculosis, and emended description of the family Thermoactinomycetaceae Matsuo *et al*. 2006. *International Journal of Systematic and Evolutionary Microbiology*, 59 (3), 454–9.

Yeoman, C.J., Han,Y., Dodd, D., Schroeder, C.M., Mackie,R.I. & Cann, I.K.O. (2010). Thermostable Enzymes as Biocatalysts in the Biofuel Industry. *Advances in Applied Microbiology*, 70, 1-55.

York, W.S., Kolli, V.S.K., Orlando, R., Albersheim, P. & Darvill., A.G. (1996). The structures of arabinoxyloglucans produced by solanaceous plants. *Carbohydrate Research*, 285, 99–128.

Zambare, V., Zambare, A. & Muthukumarappan, K. (2011). Biochemical characterization of thermophilic lignocellulose degrading enzymes and their potential for biomass bioprocessing .*International Journal of Energy and Environment*, 2 (1), 99–112.

Zhang, L., Chou, C.P. & Moo-Young M. (2011). Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by *Escherichia coli* expression system. *Biotechnology Advances* , 29, 923–929.

Zhang, Y.-H. P. & Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and Bioengineering*, 88 (7), 797–824.

Zhao, C., O-Thong, S., Karakashev, D., Angelidaki, I., Lu, W., & Wang, H. (2009a). High yield simultaneous hydrogen and ethanol production under extreme-thermophilic (70°C) mixed culture environment. *International Journal of Hydrogen Energy*, 34 (14), 5657–5665.

Zhao, X., Cheng, K. & Liu, D. (2009b). Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology*, 82 (5), 815–27.

Zhou, J., Bao, L., Chang, L., Zhou, Y., & Lu, H. (2012). Biochemical and kinetic characterization of GH43 β-D: -xylosidase/α-L: -arabinofuranosidase and GH30 α-L-arabinofuranosidase/β-D-xylosidase from rumen metagenome. *Journal of Industrial Microbiology Biotechnology*, 39 (1), 143–152.