Development of an Actinobacteria based *in vitro* Transcription and Translation Systems

Takalani Whitney Maake

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

Department of Biotechnology,

University of the Western Cape,

Bellville

Supervisor: Prof. Marla (Trindade) Tuffin

Co-supervisor: Mr. Lonnie van Zyl

May 2015

Declaration

I, Takalani Whitney Maake, hereby declare that "**Development of an Actinobacteria based *in vitro* transcription and translation systems**" is my own original work and, that I have accurately reported and acknowledged all sources. Furthermore, this document has not previously, in part or, in its entirety been submitted at any university for the purpose of obtaining an academic qualification.

....

(Signature) (Date)

Abstract

Heterologous metagenomic screening strategies have relied largely on the construction of DNA libraries and screening in *Escherichia coli* to access novel enzymes. There is an increased demand for the identification of novel lignocellulose degrading enzymes with enhanced biochemical properties which are suitable for applications in industrial processes; biofuels being one of them. The use of heterologous gene expression in function based metagenomic studies has resulted in the discovery of enormous novel bioactive compounds. However, there are limitations associated with using *E. coli* as a heterologous host which does not allow transcription and translation of all genes in the metagenome. *E. coli* can only express 40% of the environmental DNA because of promoter recognition, codon usage, and host toxicity of gene products. Therefore alternative strategies for expressing or producing novel enzymes are needed, which can also be employed in metagenomic gene discovery. *In vitro* protein synthesis is an important tool in molecular biology and used to obtain proteins from genes for functional and expression studies. These systems may hold the key to unlock more of the potential in metagenomic DNA. The broader aim of the study is to develop non- *E. coli* based cell-free protein synthesis systems to further the metagenomics screening. In this study, *Rhodococcus erythropolis* H8 was evaluated for its suitability in cell-free expression. Crude extracts containing the macromolecular components (70S or 80S ribosomes, tRNAs, initiation, elongation and termination factors) from *R. erythropolis* were prepared using existing crude extract based cell-free protein synthesis (CFPS) protocols. Three genes were selected and used as templates for synthesis: *cell*11, *xp*12 and acetyl xylan esterase (*axe*10), all previously isolated from metagenomic libraries screened in *E. coli.* As judged by zymograms and enzyme assays, all enzymes were successfully expressed from their native promoters and in recombinants clones using the PtipA promoter, and were

active. Furthermore, the amounts of XP12 protein produced using pFos-XP_12 was 1.2mg/ml from *E. coli* and 1.67mg/ml from *R. erythropolis* CFPS, showing that the *R. erythropolis* machinery was more efficient in the expression of XP12 than the *E. coli* machinery. To the best of our knowledge this is the first demonstration of a cell-free expression using an actinomycete.

KeyWords: Metagenomics, Actinobacteria, *in vitro* protein synthesis

Acknowledgment

I would like to thank

- God for carrying me throughout my studies
- My supervisor Assoc. Prof. Marla Trindade (Tuffin) for allowing me to work under her supervision and for funding.
- My co-supervisor Mr. Lonnie van Zyl for constant help and moulding me into a scientist.
- Dr Bronwyn Kirby and Dr Heide Goodman for the day to day running of the lab.
- My family [mom, dad, brothers (Jury, Fidel), sisters (Rudzani, Phophi), husband (Lutendo)] and in-laws for encouragement and taking care of my son (Tendani) throughout my degree.
- To everyone for bringing laughter in the lab everyday; Timothy, Relebohile, Regina, Londiwe, Kulani, Lyle, Israel, Michael, Marriett, Manilize, Reyghana, Mapula, and Fazlin.
- Mr Walter Nevondo for the practical contribution to my project.

Table of Contents

List of Figures

CHAPTER 2

CHAPTER 3

CHAPTER 4

List of tables

CHAPTER 2

CHAPTER 3

CHAPTER 4

List of abbreviations

Aa	Amino acid
ACN	Acetonitrile
AFEX	Ammonium fiber explosion
AGC	Adaptive gain control
AIST	National Institute of advance Industrial
Amp	Ampicillin
Amp^R	Ampicillin resistance gene
APS	Ammonium persulphate
ASM	American society for microbiology
ATP	Adenosine triphosphate
AXE	Acetyl xylan esterase
BAC	Bacterial artificial chromosome
BG	β-glucosidases
BLAST	Basic local alignment search tool
bp	base pairs
CAF	Stellenbosch University central analytic facility
CAM	Chloramphenicol
cAMP	cyclic Adenosine monophosphate
CAZy	Carbohydrate-Active Enzymes database
CBHs	Cellobiohydrolases
cDNA	complementary DNA
CE	Carbohydrate esterases
CFPS	Cell-free protein synthesis
Chl^R	Chloramphenicol resistance gene
CMC	Caboxymethylcellulase
CO_2	Carbon dioxide
CTAB	Cetyl-trimethyl ammonium bromide
C-terminus	Carboxy-terminus
ddH2O	Deionised distilled water
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNS	Dinitrosalicylic acid
dNTPs	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
e.g.	Example
EGs	Endoglucanases

EPIs	Efflux pump inhibitors
ESI	Eletronspray ionization
FA	Formic acid
FACS	Fluorescence activated cell sorting
FAE	Ferulic acid esterase
FITC	Fluorescein isothiocyanate
FRET	Fluorescence energy transfer
GC	Gas chromatography
gDNA	genomic DNA
GFP	Green fluorescence protein
GHs	Glycoside hydrolases
HCl	Hydrochloric acid
HMW	High molecular weight
HPLC	High performance liquid chromatography
HTS	High throughput sequencing
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMBM	Institute for microbial Biotechnology and Metagenomics
IVC	*In vitro* compartmentalization
K_2HPO_4	Dipotassium hydrogen phosphate
kDa	kilo Dalton
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
LiP	Lignin peroxidase
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
MCS	Multiple cloning site
MES	2-(N-morpholino) ethanesulfonic acid
MMLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
MnP	Manganese-dependent lignin peroxidase
MUX	Methylumberliferyl-β-D xylopyranoside
MW	Molecular weight
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide
NGS	Next generation sequencing
NH_4HCO_3	Ammonium bicarbonate
$NiSO_4$	Nikel (II) sulphate
N-terminus	Amino-terminus
OD	Optical density
ORFs	Open reading frames
PCR	Polymerase chain reaction
PEG	Polyethyleneglycol

pNPX	p-nitrophenyl-β-D-xylopyranosidase
PVP	Polyvinylpyrrolidone
RBS	Ribosome binding site
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase PCR
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment
sec (s)	Second (s)
SIGEX	Substrate induced gene expression
SO_2	Sulphur dioxide
sp	Species
SSF	Simultaneous saccharification and fermentation
TAE	Tris acetate ethylenediaminetetraacetic acid
TCEP	Triscarboxyethyl phosphine
TE	Tris-EDTA
TEMED	N,N,N',N-tetramethylethylenenediamine
TFA	Trifluoroacetic acid
Tet^R	Tetracycline resistance gene
TGP	Tryptone glycerol pyruvate
$Thio^R$	Thiostrepton resistance gene
tRNA	transfer RNA
Tris-HCl	Tris (hydroxymethyl)methylamine hydrochloride
UV	Ultraviolet
w/o	water-in-oil
w/o/w	water-in-oil-water

CHAPTER 1:
General introduction and project aims

Chapter 1 General introduction and project aims

The biocatalytic conversion of lignocellulosic biomass into bioethanol may prove to be an alternative source of energy, and has made ethanol one of the most important products in biotechnology; in terms of market value (Taherzadeh and Karimi, 2007). The process of bioethanol production from lignocellulosic biomass is performed by pre-treatment and hydrolysis, which is followed by microbial fermentation of sugars. The aim of the pre-treatment step is to remove lignin and hemicellulose which can be carried out by enzymatic or physical hydrolysis. The enzymatic hydrolysis of cellulose is conducted by microbially-derived cellulase enzymes (filamentous fungi and actinomycetes). It is known that the synergistic action of multiple lignocellulolytic enzymes are required to facilitate the breakdown of the major structural components of lignocellulose; cellulose, hemicellulose and lignin to allow access to the large amounts of potential energy harboured within it (Kumar *et al*., 2008). Lignocellulose degrading enzymes include cellulases which act on cellulose and a large number of hemicellulases, complete list of these is provided in **Table 2.1**. There is a demand to identify new enzymes with enhanced biochemical properties for applications in industrial processes relying on lignocellulose degradation. Metagenomics is one of the ways which is used to access novel lignocellulases (Duan and Feng, 2010).

Metagenomics is a culture independent approach for analysing mixed microbial genomes directly from the environment. This approach has become a benchmark for exploring the inaccessible metabolic potential of microorganisms and understanding their genetic diversity. Since its introduction, metagenomics has facilitated significant developments in microbiology

and biotechnology, including the discovery of novel antibiotics (Singh and Macdonald, 2010; Craig *et al*., 2010), antitumor compounds (Piel, 2002), and industrially useful enzymes (Simon and Daniel, 2011). Metagenomics has also revealed an enormous phylogenetic diversity of microorganisms that were left out by classical microbiology techniques (Achtman and Wagner, 2008; Teeling and Glöckner, 2012). However, the scope and application of metagenomics is limited by numerous challenges, most of which are associated with heterologous expression of metagenomic DNA (mDNA). In addition, low throughput screening methods have relied largely on the construction of DNA libraries and screening *in vivo* in hosts such as *E. coli*. This does not allow transcription and translation of all genes from the metagenome because of the toxicity of some proteins resulting in the majority of the metagenomic genes remaining untapped. Another problem associated with heterologous protein expression, is microbial codon preferences. The frequency of codon usage in an organism is directly related to the number of corresponding tRNAs due to number of factors including (Stoletzki and Eyre-Walker, 2007); foreign genes with rare codons are most likely not to be expressed, expressed in low amounts or as an insoluble inclusion body or with defective protein folding because their expression is limited to the available aminoacyl-tRNA (Chen and Inouye, 1994; Angov, 2011).

Attempts to overcome the limitations associated with using *E. coli* as a surrogate host have included development of multiple host expression systems using shuttle vectors that allow metagenomic DNA to be screened in *Pseudomonas putida* and *Bacillus subtilis* (Martinez *et al.*, 2004; Hain *et al.*, 2008; Craig *et al.*, 2010; Troeschel *et al.*, 2012; Terron-Gonzalez *et al.*, 2013). These initiatives are intended to overcome stringency associated with inefficient transcription of target genes and improper assembly of corresponding enzymes. In addition, Leggewie *et al*.,

(2006) also developed a transposon, MuExpress which inserts randomly into environmental DNA thereby allowing gene expression from its inducible promoters. Even though these methods have helped gain access to a wider range of the sequence and functional space, the expression of metagenomic genes still relies on the capacity of the surrogate hosts (Terron-Gonzalez *et al.*, 2013).

As a result of these heterologous expression problems, the present study reports the development of an actinobacteria based *in vitro* transcription and translation system. The study aimed to demonstrate the proof of concept on the expression of a number of lignocellulolytic enzymes namely an acetyl xylan esterase, and two different metagenomic-derived cellulases. These enzymes were selected because they have previously been biochemically characterized and evaluated for their potential use for the hydrolysis of various biomass sources (sorghum and bagasse).

The specific objectives of this study included:

- To prepare crude extracts for *Rhodococcus erythropolis* H8 for *in vitro* protein synthesis.
- To clone and express genes encoding acetyl xylan esterase (AXE) and cellulase in *R. erythropolis* and compare expression using *R. erythropolis*-based cell-free expression to *E. coli* based cell-free expression.
- To confirm *in vitro* protein synthesis using MALDI-TOF mass spectrometry.

CHAPTER 2:
Literature Review

Chapter 2 Literature Review

2.1 The need for new biocatalysts to improve bioethanol production

It is estimated that the world oil reserves may be depleted by 2050, which means that significant efforts have to be made to develop renewable energy. One of the challenges facing modern society is meeting the demand for energy because it is important in transportation, industrial purposes and heating (Hahn- Hägerdal *et al.*, 2006). One of the more promising solutions to our energy problem is thought to be the development of biofuels (Demain, 2009). Biofuels refers to the following products; biodiesel, biogas, and bioethanol. The most common source for biofuels is photosynthetic plants; however, it can be produced from any biological carbon source.

Biofuels are either produced by 1st or 2nd generation processes (Naik *et al.*, 2010). 1st Generation processes refer to those in which biofuels are made from food crops such as; sugar cane, starch from corn using commercial yeast strains (*Saccharomyces cerevisiae* and *Schizossaccharomyces pombe*) (Sanchez and Cardona, 2008). However, the over-use of available land to cultivate these crops may result in deforestation (Gomez *et al.*, 2008) and also have a negative impact on food security especially in developing countries , where these food crops are used as staple food (Graham-Rowe, 2011), which will result in commodity prices rising. 2nd Generation processes appear to be a better alternative as it uses non-food crops and other lignocellulosic sources such as plant biomass (non-edible parts) which include straw waste, cornstalks, sugarcane bagasse, distillers dry grain, wood chippings or other organic material like municipal waste (Hamelinck *et al*., 2005; Hahn- Hägerdal *et al*., 2006).

2.2 Conversion of lignocellulosic biomass to bioethanol

Lignocellulose is a renewable organic material and it is composed of lignin, cellulose and hemicellulose (**Figure 2.1**) (Malherbe and Cloete, 2002; Howard *et al.*, 2004).

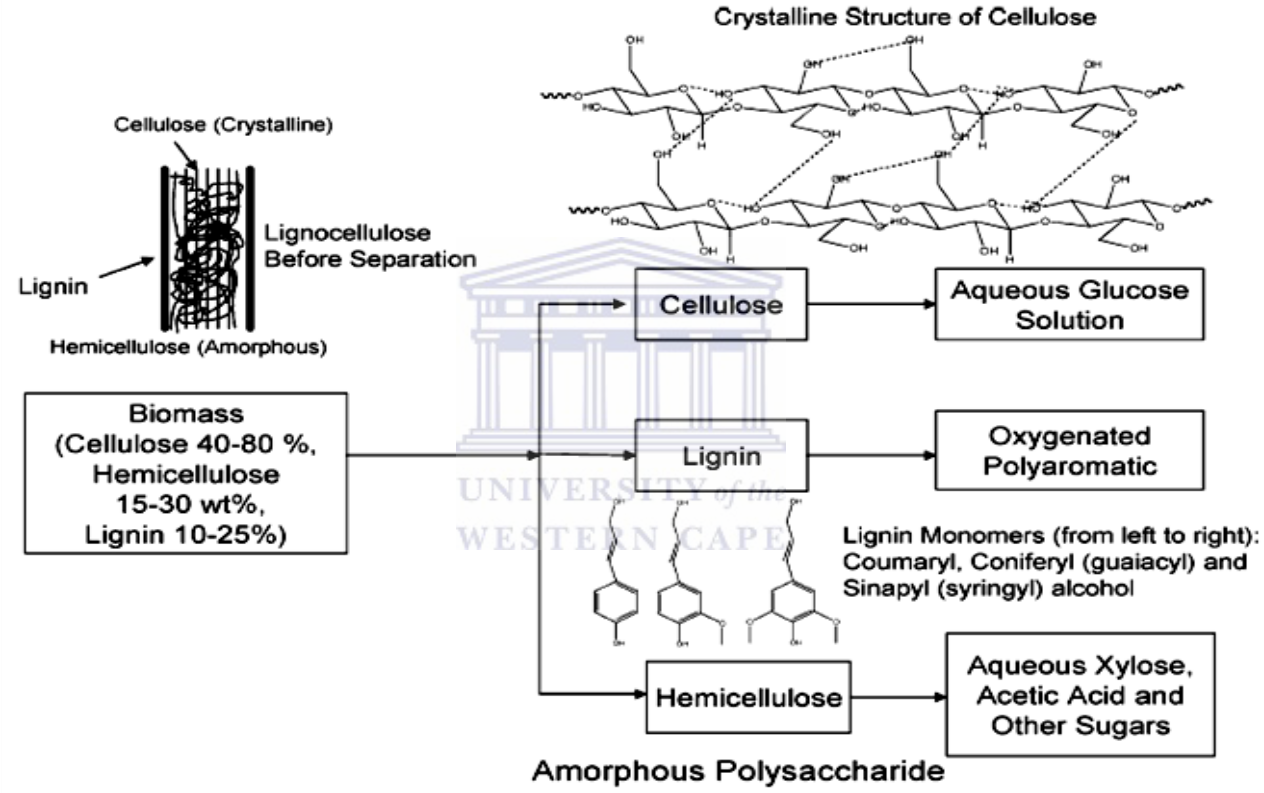


Figure 2.1: The polysaccharides component of lignocellulose, showing different subunits; lignin, hemicellulose and cellulose (Huber *et al*., 2006).

Cellulose is the most abundant polymer in nature and it is composed of thousands of glucose residues that are linked together by β-1, 4 glycosidic bonds. The association of cellulose with

lignin, hemicellulose, starch, protein and mineral elements makes lignocellulose a molecule resistant to hydrolysis. The process of hydrolysis breaks down polysaccharides to free sugar molecules in solution through saccharification (Hamelinck *et al.*, 2005).

Lignin is the most abundant aromatic polymer in nature and is a highly branched substituted mononuclear aromatic polymer which is found in plant cell walls (Huber *et al*., 2006). It consists of basic phenyl propane units (Zeng *et al*., 2013), also known as monolignols, which are produced through esterification and differ from each other in their degree of methoxylation; *trans*-*p*-coumaryl, *trans*-*p*-coniferyl and *trans*-*p*-sinapyl (Kirk *et al.*, 1977; Zaldivar *et al*., 2001; Liu *et al*., 2012), that are joined by different linkages and plays an important role in lignin biosynthesis. Plant lignin is divided into three classes; softwood (gymnosperm), hardwood (angiosperm), and grass plant (graminaceous). The polymerization of lignin is initiated by the oxidation of phenylpropane phenolic hydroxyl groups. The three lignin precursors undergo dimerization through dehydrogenation by an electron transfer mechanism, which gives rise to the resonance-stabilized phenoxy radicals (monolignol free radical) (Van holme *et al*., 2008). The monolignol free radical forms dilignols or dimers by coupling with another free radical, before the endwise polymerization, that are later incorporated into the polymer. These monomers are now called *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units (**Figure 2.2**) (Cesarino *et al.*, 2012). It is this cross-linking which makes lignin a complex substance, making it difficult to degrade (Hamelinck *et al*., 2005).

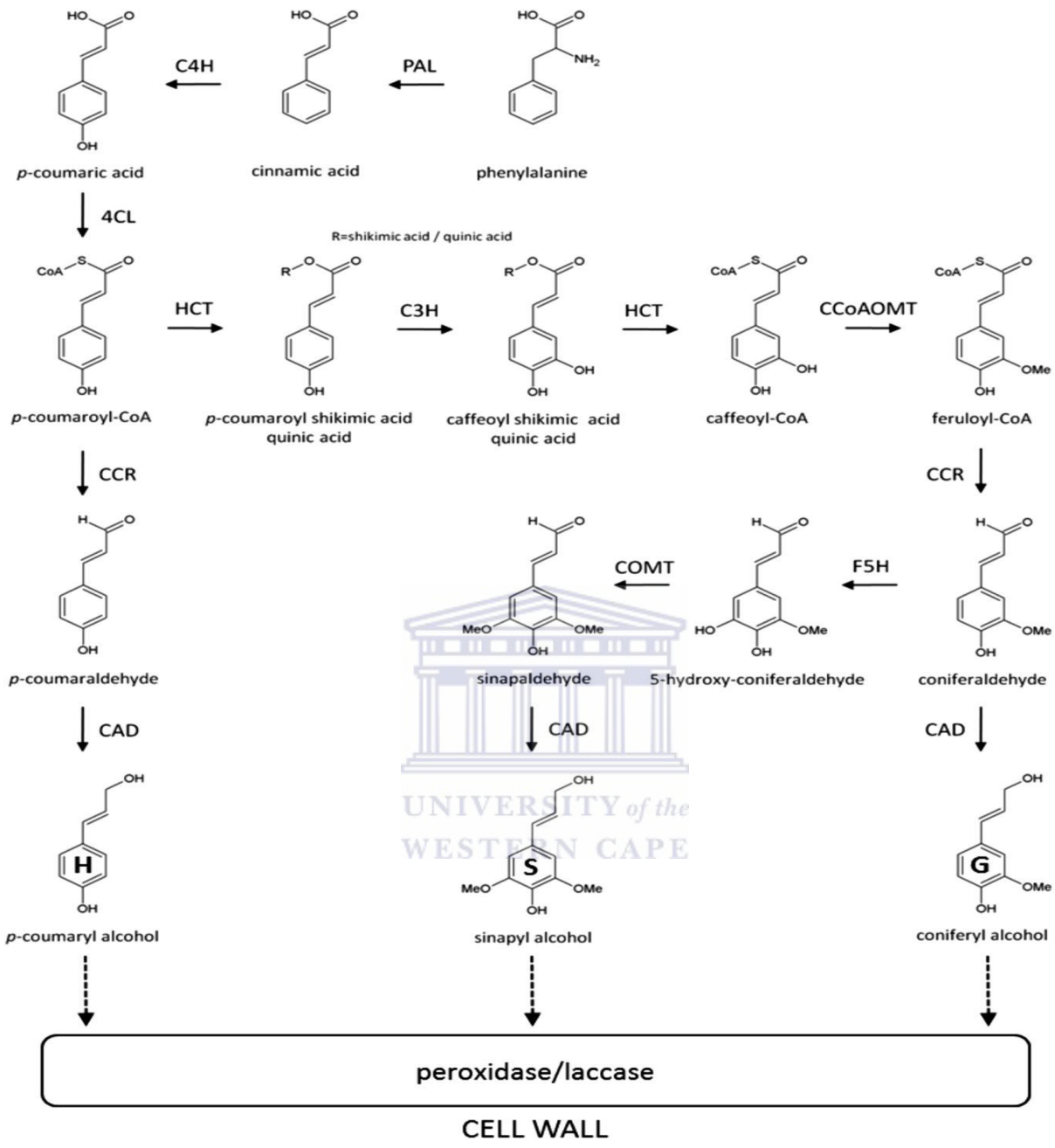


Figure 2.2: The biosynthetic pathway of three monolignols of lignin; p-coumaryl, sinapyl alcohols, and coniferyl. The monolignols are synthesized in the cytoplasm and then transported to the cell wall where they are oxidised by peroxidase or laccase. Abbreviations: PAL (phenylalanine ammonia-lyase); C4H (cinnamate 4-hydroxylase); 4CL (4-coumarate-CoA ligase); C3H (p-coumarate 3-hydroxylase); HCT (p-hydroxycinnamoyl-

CoA: quinate shikimate *p*-hydroxycinnamoyltransferase); CCoAOMT (caffeoyl-CoA O-methyltransferase); CCR (cinnamoyl-CoA reductase); F5H (ferulate 5-hydroxylase); COMT (caffeic acid O-methyltransferase); CAD (cinnamyl alcohol dehydrogenase); H (*p*-hydroxyphenyl), S (syringyl) and G (guaiacyl) (Adapted from Cesarino *et al*., 2012).

Hemicellulose is a heteropolymer that constitutes up to 50% of lignocellulose biomass, depending on the source and it is comprised of various sugar residues including: pentose (D-xylose and L-arabinose), hexose (D-galactose, L-galactose, D-mannose, L-fructose and L-rhamnose) and uronic acids (D-glucuronic acids), that are linked together by β-1,4 glycosidic bonds **(Figure 2.3)** (Zaldivar *et al*., 2001). The most abundant hemicelluloses in nature are xylans (Saha, 2003).

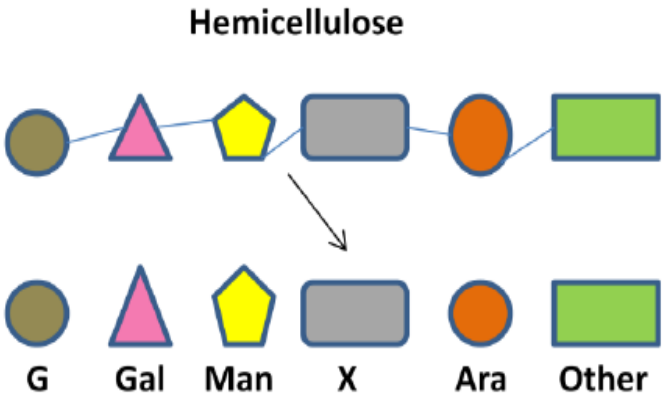


Figure 2.3: The various sugars that are found in a hemicellulose, which are linked together by β-1, 4 glycosidic bonds. Through hydrolysis they are broken down into G (glucose), Gal (galactose), Man (Mannose), X (xylose), Ara (arabinose), and other (uronic acids) (Zaldivar *et al*., 2001).

Xylans are found in grasses, softwood, and hardwood, which contain xylose, gluconic acid, acetic acid, ρ-coumaric, arabinose and ferulic acid. Depending on the source, the composition and the frequency of the branches differs, and their backbone has the following substituents; α-L-arabinofuranosyl, *O*-acetyl, 4- *O*-methylglucuronic or α-1, 2-linked glucuronic acid. As shown in **Figure 2.4**, almost 80% of the backbone is substituted with monomeric arabinose or glucuronic acid which is linked to the xylose residue on the *O*-2 or *O*-3 positions (Saulnier *et al*., 1995; Saha, 2003).

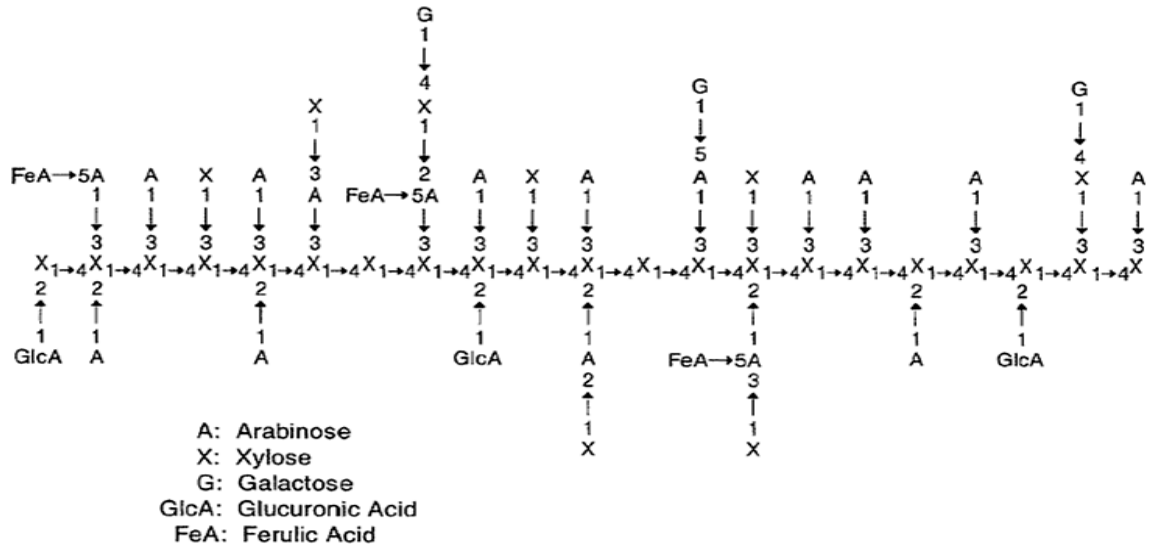


Figure 2.4: The structure of a heteroxylan of corn substituted on the second or third oxygen of the xylose by arabinose, glucuronic acid or galactose and cross-linked by ferulic cross-bridges (Saulnier *et al.*, 1995).

Bioethanol production from lignocellulosic biomass relies on pre-treatment and hydrolysis to extract sugars, which is followed by the microbial fermentation of sugars (Taherzadeh and Karimi, 2007; Olofsson *et al*., 2008).

2.1 Pre-treatment

Various pre-treatment options are available in order to accelerate the hydrolysis rate, which separate carbohydrates from the lignin matrix while minimizing the chemical destruction of fermentable sugars that are needed for ethanol production (Mielenz, 2001; Fernandez *et al*, 2009). These include: physical treatment such as mechanical comminution, pyrolysis and

irradiation; physico-chemical treatment such as steam explosion, ammonium fiber explosion (AFEX), carbon dioxide explosion (CO_2) explosion, and SO_2 explosion; chemical treatment such as ozonolysis, dilute-acid hydrolysis, alkaline hydrolysis, oxidative delignification and organosolv; and ionic liquids (Taherzadeh and Karimi, 2007; Brandt *et al*., 2012).

The selection of pre-treatment method should be compatible with the selection of hydrolysis, for example, when acid-hydrolysis is applied; pre-treatment with alkali may not be beneficial. The most commonly used method of chemical hydrolysis used is the dilute-acid hydrolysis because this method can be used as a pre-treatment which will be followed by enzymatic hydrolysis or as an actual method for hydrolysing lignocellulose to sugars (Taherzadeh and Karimi, 2007). The advantage of using dilute acid hydrolysis is that it allows a high level of xylan to xylose conversion (Sun and Cheng, 2002).

The ionic liquids that are used for pre-treatment have cations (phosphonium, sulfonium, and quartenised ammonium) and high hydrogen bonding basic anions (carboxylates, phosphonates, phosphates and chloride) (Brandt *et al*., 2012). The use of ionic liquids is new and it allows one to change the physiochemical properties of the biomass. This can be performed after biomass dissolution to perform different fractionation approaches. Specific macromolecular components provided by ionic liquids can be extracted (Lopes *et al.,* 2013). To date, 1-Ethyl-3-methylimidazolium ([emim] [CH3COO]) is the preferred ionic liquid for the pre-treatment of lignocellulose biomass because of the extra solvent power it has (Sun *et al*., 2009; Doherty *et al*., 2010; Lopes *et al.,* 2013; and Viell *et al*., 2013). Even though a lot of research has yet to be done

on the use of ionic liquid for this purpose, this method has been successfully used in biomass separation to obtain cellulose, hemicelluloses and lignin fractions (López-Barrón *et al.*, 2014).

2.2 Enzymatic hydrolysis

Enzymatic hydrolysis of cellulose is done by cellulose degrading enzymes that are highly specific (Beguin and Aubert, 1994).

In bioethanol production, according to Kristensen *et al*., 2009, certain targets for the sugar yield have to be met; for example, 8% (w/w) of sugar yield is needed in order to achieve 4% (w/w) of the final ethanol concentration in a bioreactor, and this reduces the energy requirement for distillation, hence the whole process becomes more economically feasible (Van Dyk and Pletschke, 2012).

The utility cost for enzyme hydrolysis is low as compared to the acid or alkaline hydrolysis because it is usually conducted under mild conditions (pH 4.5-5.0 and temperature 40-50 °C) (Sun and Cheng, 2002).

Factors that affect hydrolysis of lignocellulosic materials are; substrate concentration, applied pre-treatment method, cellulase activity, substrate quality and hydrolysis conditions (pH, temperature). Substrate (cellulose or hemicellulose) concentration is one of the main factors which affect yield and initial rate of enzymatic hydrolysis, because low substrate concentration results in low yield and high substrate concentration causes substrate inhibition that lowers the efficiency of hydrolysis (Taherzadeh and Karimi, 2007). The level of substrate needed is

dependant on which sugars will be utilized, meaning that the loading of substrate levels will vary depending on the lignin and carbohydrate composition (Van Dyk and Pletschke, 2012). Sugar yield after enzymatic action can be determined using (dinitrosalicylic acid) DNS assay, HPLC, and X- ray diffraction (Park *et al*., 2010). The problem with DNS assays is that it is an assay for determining reducing sugars, which can be used to measure the rate of enzyme activity, and is not an accurate reflection of the yield because it does not measure the monomer sugar production (Van Dyk and Pletschke, 2012). HPLC analysis determines the yield of sugar monomers (glucose, arabinose, mannose, xylose) produced during enzyme degradation by using appropriate standards (Park *et al*., 2010). X- ray diffraction measures the degree of crystallinity in crystalline cellulose, and also the contribution of specific enzymes during the removal of steric hindrances on a substrate.

Due to complexity of lignocellulose, it requires a variety of enzymes acting in synergy for complete hydrolysis. For example, the removal of acetyl group on acetylated xylan (Cybinski *et al*., 1999) or ferulic acid from linkages between hemicelluloses and lignin (Faulds *et al*., 2003) is required in order to give xylanases access to the xylan backbone. Evaluating the relationship between synergy and yield is useful for understanding the degradation of lignocellulose and in optimizing the enzyme combinations (Van Dyk and Pletschke, 2012).

2.2.1 Cellulases

The enzymatic hydrolysis of cellulose is carried out by specific enzymes called cellulases (glycosyl hyrolysases). Cellulases essentially refer to enzymes that belong to two classes: endo-1, 4-β-D-glucanases (EGs) (EC 3.2.1.4) and cellobiohydrolases (CBHs) also called exo-1, 4-β-D-

glucanases (EC 3.2.1.91) (Wyman, 1996; Sukumaran *et al*., 2009). The β-glucosidases (BG) are the third class of enzymes which work together with CBH and EG, and are often not referred to as real cellulases because they primarily hydrolyze glycosidic bonds of soluble cellooligosaccharides and cellobiose, not acting on cellulose itself.

Endoglucanases cleave low-crystallinity regions of the cellulose fiber and create free chain ends **(Figure 2.5)**, while opening them up for subsequent attack by the cellobiohydrolases (Lee, 1997). Cellobiohydrolases further degrade the sugar chain by removing glucose dimers (cellobiose units) from the free ends (Howard *et al.*, 2004). β-glucosidases complete the depolymerisation of cellulose to glucose. In the absence of β-glucosidase, end-product inhibition from cellobiose will result (Keshwani and Cheng, 2009). Many methods have been developed in order to overcome this inhibition such as increasing the amount of enzyme added and / or removal of sugar during hydrolysis by ultrafiltration or simultaneous saccharification and fermentation (SSF) (Taherzadeh and Karimi, 2007). The advantages of SSF includes: increased hydrolysis rate, lower enzyme requirement, and high product yields (Zaldivar *et al*., 2001).

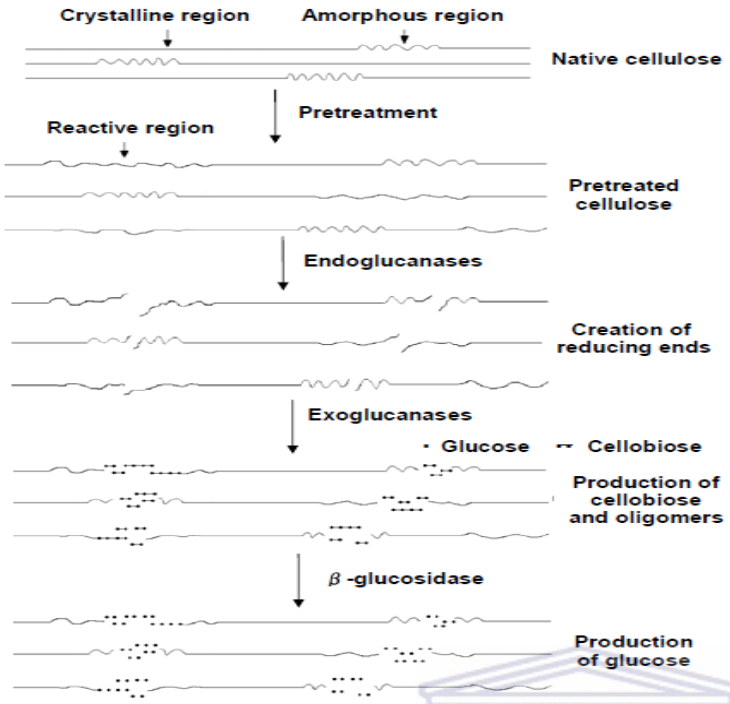


Figure 2.5: A schematic representation of the hydrolysis of cellulose by cellulases to glucose. The endoglucanase cleaves the crystalline and amorphous region, and then the cellobiohydrolases further degrade the sugar chain and release the cellobiose, and β-glucosidase completes the depolymerisation of cellulose to glucose (Taherzadeh and Karimi, 2007).

2.2.2 Hemicellulases

Hemicellulases are needed to complete degradation because of the complex chemical composition of lignocellulose. For example, for each structural feature in hemicellulose there is an associated enzyme that exists that can hydrolyse or chemically modify the feature. These include enzymes that breakdown β-1, 4 xylan (xylanases and β-xylosidases) and various side

chain (acetyl xylan esterase, α-galactosidases, α-glucuronidases, α-l-arabinofuranosidaseses) (**Table 2.1)** (Gray *et al.*, 2006; Howard *et al*., 2004).

Hemicellulases are either glycoside hydrolases (GHs) which hydrolyze glycosidic bonds or carbohydrate esterases (CEs) that hydrolyze ester linkages of acetate or ferulic acid side chains. GHs families are grouped according to the amino acid sequence and folding similarity by the Cabohydrate-Active Enzymes database (CAZy) (http:cazy.org/Glycoside-Hydrolases.html).

Hemicellulases can also be categorized as exo-acting enzymes, endo-acting enzymes or accessory (ancillary) enzymes, as shown in **Table2.1**. The endo-acting enzyme cleaves polysaccharides chains, while the exo-acting enzyme acts at either the reducing or non-reducing terminus (Himmel *et al*., 2010). The accessory enzymes include acetyl esterases and ferulic acid esterases (FAE).

Table 2.1: A summary of the major Hemicellulases (Howard *et al*., 2004).

Enzymes	Substrates	EC number	Family
Exo-β-1,4-xylosidase	β-1,4-xylooligomersxylobiose	3.2.1.37	GH 3, 39, 43, 52, 54
Endo-β-1,4-xylanase	β-1,4-Xylan	3.2.1.8	GH 5, 8, 10, 11, 43
Exo-β-1,4-mannosidase	β-1,4-Mannooligomersmannobiose	3.2.1.25	GH 1, 2, 5
Endo- β-1,4-mannanase	β-1,4-Mannan	3.2.1.78	GH 5, 26
Endo-α-1,5-arabinanase	α-1,5-Arabinan	3.2.1.99	GH 43
α – L-arabinofuranosidase	α-Arabinofuranosyl (1→ 2) or (1→ 3) xylooligomers α-1,5-arabinan	3.2.1.55	GH 3, 43, 51, 54, 62
α-Glucuronidase	4-*O*-Methyl- α-glucuronic acid (1→2) xylooligomers	3.2.1.139	GH 53
α-Galatosidase	β-1,4-Galactan	3.2.1.22	GH 4, 27, 36, 57
Endo-galactanase	α-Galactopyranose (1→6) mannooligomers	3.2.1.89	GH 53

β-Glucosidase	β-Glucopyranose (1→4) mannopyranose	3.2.1.21	GH 1 ,3
Acetyl xylan esterases	2-or 3-*O*-Acetyl xylan	3.1.1.72	CE 1, 2 ,3 ,4 ,5, 6
Acetyl mannan esterase	2-or 3-*O*-Acetyl mannan	3.1.1.6	CE 1
Ferulic and *p*-cumaric acid esterases	2-or 3-*O*-Acetyl mannan	3.2.1.73	CE 1

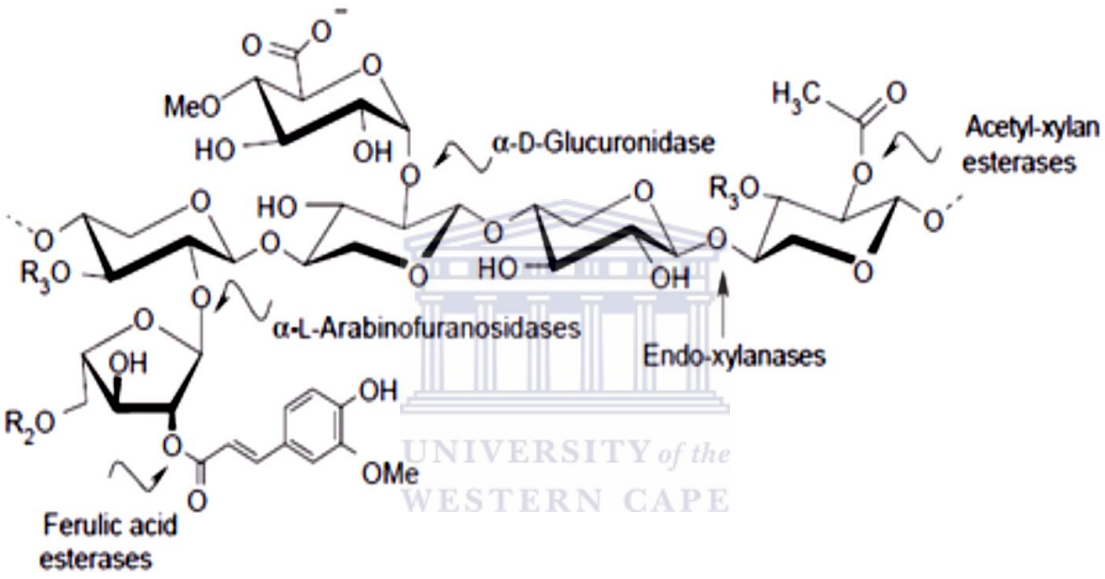


Figure 2.6: The hydrolysis of hemicelluloses involves the synergetic action with different enzymes like; acetyl xylan esterase, ferulic acid esterase, α-D-glucuronidase, α-L-arabinofurosidase and endo-xylanase (Krastanova *et al*., 2005).

2.2.2.1 Ferulic acid esterases

Ferulic acid esterases (FAE) are enzymes that cleave the ester link between polysaccharides and monomeric (**Figure 2.6**) or dimeric ferulic acid and therefore play a major role in the degradation of plant biomass (Topakas *et al*., 2007). FAE activity liberates phenolic acids (ferulic acids and *p*- coumaric acid) and their dimers from hemicellulose. This enzyme has synergistic effects on cellulose degradation. Many microorganisms have been reported to produce FAE, and this including fungi like *Aspergillus* (*A. flavipes* and *A. niger*) and *Trichoderma* species, and bacterial *Streptomyces* species *(S. olivochromogenes* and *S. thermophile), Bacillus* species, and *Lactobacillus* species (Ou *et al*., 2011). Most FAEs have been isolated from fungal as compared to bacterial sources (Topakas *et al*., 2007).

2.2.2.2 Acetyl xylan esterases

Acetyl xylan esterases (AXEs) (EC 3.1.1.72) are one of the accessory enzymes which are required to complete hydrolysis of xylan (Degrassi *et al*., 2000; Nisole *et al.*, 2006) by acting in synergy with exo-β-1,4-xylosidase and endo-β-1,4-xylanase. AXEs belong to the carboxylic ester hydrolase family, therefore they act on carboxylic ester bonds. Acetyl xylan esterases have a similar function to FAEs because some of the acetyl xylan esterases can produce ferulic acid, and are able to remove acetyl ester groups from positions C-2 or C-3 of D-xylopyranosyl residues in xylan chains (**Figure 2.7**) (Ghatora *et al.*, 2006). These enzymes have been categorized into seven subfamilies based on amino acid sequence and their affinity to polymeric and oligomeric substrates, for example; carbohydrate esterase (CE) 1 initially de-acetylates positions 2 and 3 on xylose, CE 4 performs a rapid double de-acetylation of position 2 and 3, CE

7 shows activity towards substrates like *p*-nitrophenyl acetate, α-naphthyl acetate, 7-aminocephylosporin-C, xylose tetra acetate and glucose penta acetate (Coutinho *et al*., 2003).

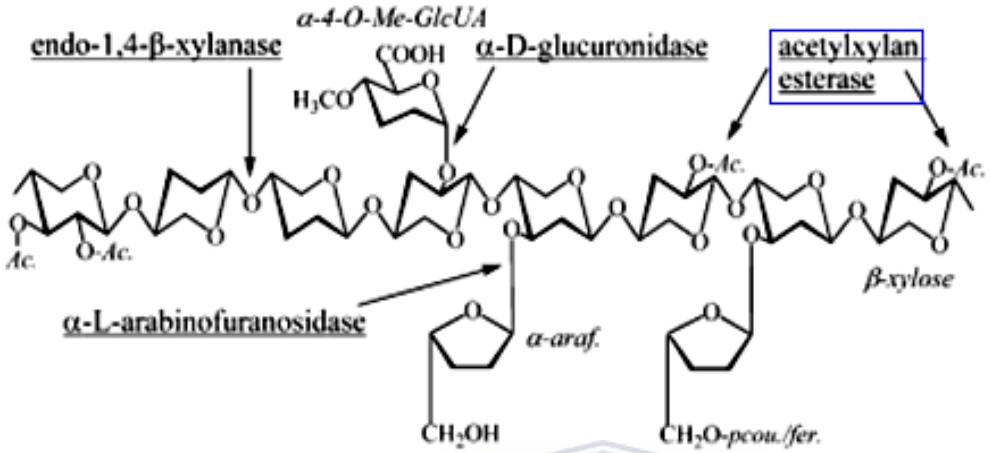


Figure 2.7: Schematic of the structure of xylan, showing the substrate backbone, that is comprised of 1, 4-β-linked residue attacked by acetyl xylan esterase and ferulic acid esterase or ρ-coumaric acid esterase (Sunna, 1997; Collins *et al*., 2004).

2.2.2.3 Exo- 1, 4-β Xylosidases

The exo-1,4-β xylosidases (EC 3.2.1.37) are produced by different organisms including fungi and bacteria, and they play a role in the degradation of xylan by hydrolysing xylooligosaccharides and xylobiose to xylose (Saha, 2003)

The intrinsic complexity of lignocellulosic biomass and its resistance to hydrolysis are the major factors that hamper its economic use for bioethanol production. This is the reason why industrial and competitive production of second generation bioethanol is not yet as developed as 1st generation processes. However a number of companies are exploiting facilities used for first generation bioethanol production to be upgraded for second generation bioethanol production from lignocellulosic biomass (Gibson and Hughes, 2009; Amore, 2013). Some examples of pilot plants include: (i) Logen plant in Canada, where cellulosic ethanol is produced from wheat straw, (ii) a pilot scale in York (Nebraska, US), (iii) a demonstrative plant in Salamanca (Spain) and (iv) a cellulose ethanol plant in Clariant (Germany), the biggest in Germany with a production rate of around 1,000 tonnes/year, using 4,500 tonnes/year of locally sourced agricultural waste as a feedstock. In Italy, the Mossi & Ghisolfi (M&G) group has built a semi-industrial plant for bioethanol production from *Arundo donax* (giant cane).

Considering that enzymatic hydrolysis, which liberates fermentable sugars, is one of the economic bottlenecks in bioethanol production using second generation processes, more effective enzyme cocktails could improve the process considerably. Metagenomic gene discovery has facilitated significantly in the discovery of lignocelulose degrading enzymes, and could be an approach to facilitate the development of more effective enzyme cocktail which could result in higher product yield than when an individual enzyme is used to degrade a complex substrate (Van Dyk *et al.*, 2012).

2.2 Metagenomics, a way of accessing novel enzymes

Metagenomics is the direct genetic analysis of genomic DNA from an environmental sample (e.g. soil), as opposed to analysis of genomic DNA from an individual organism (Handelsman *et al*., 1998). It is estimated that more than 99% of organisms have not yet been cultured using conventional methods (Pham and Kim, 2012); therefore metagenomics is applied in order to study the genomes of such uncultured microbes (Yun and Ryu, 2005).

Metagenomics provides access to the functional gene composition of microbial communities by providing information on novel enzymes (biocatalysts), showing genomic linkages between function and phylogeny for uncultured organisms, and also evolutionary profiles of community function and structure (Thomas *et al*., 2012). For gene discovery it can be divided into two areas: functional based and sequence based approaches which are followed by bioinformatic screening or mutagenesis to identify the gene of interest (Gilbert and Dupont, 2011).

The sequence based approach involves analysing metagenomic DNA to identify genes of interest, i.e. sequences that encode conserved protein domains for which complementary DNA or oligonucleotides probes are designed (Leis *et al*., 2013). Target open reading frames (ORFs) are identified by hybridization or PCR techniques, and the frequency of different target genes can be increased by using high-density arrays or degenerate primers. An unknown sequence can be determined by using *de novo* sequencing of the total metagenomic DNA or metagenomic

libraries, whereby large sequence data sets are collected for the expected coverage to be reached and then assembled and annotated for gene identification.

The functional based approach differs from sequence based because it relies on the success of the heterologous expression of gene products in a host organism (Simon and Daniel, 2009). There are three functional based screening approaches; 1-modulated detection which under selective conditions uses reporter systems that involves the growth of metagenomic library clones, 2-phenotypic detection which detects enzyme activities by using indicator reagents, 3- substrate- and product-based detection which focuses on catabolic gene expression. These techniques are elaborated below.

In phenotypic detection, screening is performed using agar plate or liquid assays. The library clones are plated on, or inoculated in, media containing the substrate or applying substrate overlays after the growth has formed (Leis *et al.,* 2013). The active clones are detected phenotypically through altered morphology, colony pigmentation or halo formation because of the released chromophores, pH or dye degradation. This type of screening detection is dependent on whether the cells are lysed or not (Cottrell *et al*., 1999), and accumulation of enzymes inside the host which can be toxic to the cells. Cells require permeabilization by chemical treatment (toluene or chloroform), enzymatic treatment (lysozyme) or soft detergents (Triton X or Tween) physical treatment (liquid nitrogen) (Taupp *et al*., 2011).

Modulated detection involves the indirect phenotypic screening of specific enzymatic activities. This detection method allows the metabolites for growth from even the weakly expressed genes, which will result in phenotypic changes of the reporter or screening (Leis *et al.*, 2013).

Substrate induced gene expression (SIGEX) (Uchiyama *et al*., 2005) is an approach which uses an operon-trap expression vector that carries a green fluorescent protein gene without a promoter. The green fluorescent gene is linked to a cloned gene that is under catabolic control of an inducible promoter. In the presence of the substrate, the recombinant gene is co-expressed with the green fluorescent protein gene. The advantage of this method is that it allows ultra-high throughput screening since the positive clone to be identified by fluorescent activated cell sorting (FACS) (**Figure 2.8**). The main limitations of this method is that it can only be used for discovery of inducible genes; it requires the fusion between the metagenomic DNA and the GFP to be such that GFP is successfully expressed and it relies on library construction, which introduces bias.

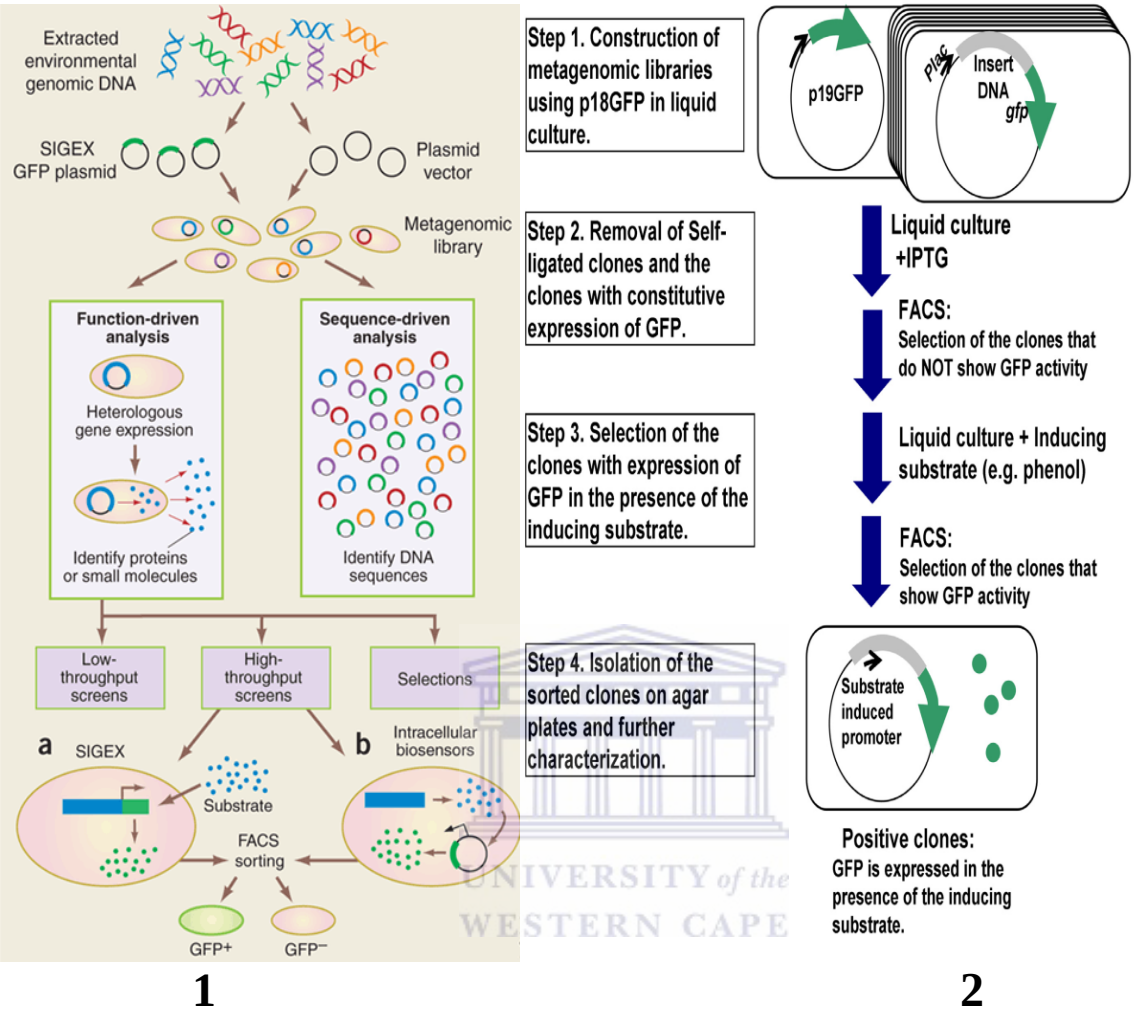


Figure 2.8: A schematic representation showing the construction of a metagenomic library and the use of function based, sequenced based and SIGEX (1). SIGEX traps the operon induced by a substrate and sort using FACS (2) (Handelsman, 2005; Yun and Ryu, 2005).

2.2.1 Extraction of mDNA

Sampling is the first step in any metagenomics project (Thomas *et al.,* 2012). The metagenomic DNA extraction protocol is the most important step because the DNA extracted should represent all members of a particular microbial population, many of which are more difficult to lyse than others.

Extraction methods are selected according to the sample and the purpose of the metagenomic study, and each extraction protocol may have one or all of the following elements; physical disruption (sonication, bead mill homogenization), chemical lysis (EDTA, sodium hydroxide, sodium dodecyl sulphate (SDS), and/or enzymatic lysis (lysozyme, proteinase K). Polyvinylpyrrolidone (PVP) can be used in DNA extraction to increase DNA purity, and chelating agents such as EDTA can be added in order to inhibit nuclease activity. DNA is recovered from the soil or cell debris using a phenol/chloroform extraction method (Zhou *et al*., 1996).

Extracting DNA is challenging because the process of removing contaminants can shear the DNA, resulting in poor DNA quality. This problem can be overcome by immobilizing the cells in an agarose matrix and subjecting it to a formamide and high salt treatment and through electrophoresis the high molecular weight DNA is separated from DNAses and humic acids (Liles *et al.,* 2008). A variety of different extraction methods such as FastDNA® Spin kit for soil followed by hexadecyltrimethylammonium (CTAB) to yield minimal contaminated DNA have

been successfully employed (Sabree *et al*., 2009). The validation of an extraction technique is dependent on whether or not the extracted DNA represents the entire microbial community and its suitability for library construction (Handelsman, 2005). Studies have shown that most DNA extraction techniques introduce biases as a result of incomplete cell lysis, DNA degradation and enzymatic inhibitors extracted from the sample (Miller *et al*., 1999).

The recovered DNA is then precipitated using isopropanol or ethanol. Recovery of metagenomic DNA that is suitable for restriction enzyme digestion, PCR and library construction is challenging because of polyphenolics; substances from soil and other environmental materials which can interfere with the downstream applications (Yeates *et al*., 1998; Pang *et al*., 2008), therefore DNA purification using chromatography or filtration may be necessary (Tebbe and Vahjen, 1993). The extraction of high molecular weight (HMW) metagenomic DNA from an environmental sample is important because 30-50kb size range is incorporated by fosmids or cosmids for generating large insert libraries. A large insert library enables the recovery of biosynthetic pathways (Bertrand *et al.*, 2005).

The extraction of RNA is challenging because of the short half-life of the prokaryotic mRNA, adsorption by soil particles, interference with high-quality mRNA due to co-extraction of other types of RNA and degradation by RNases (Simon and Daniel, 2009; Wang *et al*., 2012;. Leis *et al*., 2013). The extracted mRNA can be transcribed into cDNA in order for the transcripts to be cloned into metagenomic libraries.

2.2.2 Cloning and expression vectors

Cloning vectors are selected based on the desired target activities and library structure such as; number of clones and the size of the insert; and the method of screening. Small DNA fragments (<10 kb) can be cloned using standard cloning vectors (pBluescript SK+, pUC derivatives, pCF430, and pTOPO-XL), to identify the function encoded by a single gene; a wide range of enzyme classes including lignocellulosic enzymes (cellulases, hydrolases and amidases) have been identified from metagenomic libraries that use smaller insert (Lämmle *et al*., 2007; Sabree *et al*., 2009). Large DNA fragments (>20 kb) can be used to target functions encoded by multiple genes and these are cloned into Fosmids (25-40 kb), Cosmids (25-35 kb) (Voget *et al*., 2006) or Bacterial artificial chromosomes (BAC) (100-200 kb) (Béjà, 2004). Some of the commonly used vectors that are used for cloning larger DNA fragments include; pCC1FOS, pCCERI and pWE15 (Sabree *et al*., 2009; Trincone and Schwegmann-Weßels, 2014). The advantage of using pCC1FOS is that higher expression can be achieved because it has an additional replication origin, oriV, to increase the copy number when used with an appropriate host (*E. coli* Epi300) (Westenberg *et al*., 2010).

2.2.3 Expression host

Heterologous expression has provided the means to study proteins from prokaryotes and eukaryotes for the production of medicines and industrial biocatalysts. Most researchers use *E. coli* as a surrogate host, which has been engineered to lack genes for homologous recombination (*rec*A, *rec*BC) and restriction (*mcr*A, *mcr*BC), making it suited for cloning foreign DNA into *E. coli* (Uchiyama and Miyazaki, 2009). Different *E. coli* strains [BL21 (DE3);

DH5α, JM109 and others] are commercially available. *Streptomyces* species and *Pseudomonas* species have been used as alternative hosts to identify genes of interest because they are the major source of secondary metabolites (polyketides and nonribosomal peptides) (Brady *et al.*, 2001; Nakashima *et al*., 2005). Different heterologous genes have been expressed in *Thermus thermophilus* including dibenzothiophene monooxygenase, β-galactosidase, chlorolophenicol acetyl transferase, and (*p*-nitrobenzyl) PNB carboxy esterase (Park *et al*., 2004). *E. coli* Epi300 competent cells are suitable to use with all clone sizes and have high transformation efficiency (Rapa *et al*., 2014).

2.2.4 Construction of metagenomic libraries

Construction of metagenomic libraries involves the following steps: mDNA is isolated; suitable sized DNA fragments are generated (random shearing) and size fractionated; DNA end repair; cloning into a vector and screening for the desired gene (**Figure 2.9**). Many studies have described the construction of metagenomic libraries for the isolation of novel biocatalysts or their products for pharmaceutical and biotechnological applications (Wilson and Piel, 2013; Guazzaroni *et al*., 2014). Metagenomics has so far been the most successful approach in screening of novel lignocellulolytic activities (Simon and Daniel, 2009). However, the success of metagenomic-based discovery is limited by numerous challenges, most of which are associated with heterologous expression of metagenomic DNA and low throughput screening methods. *In vitro* metagenomics has been suggested as a solution for overcoming the heterologous expression problems (Ferrer *et al*., 2007).

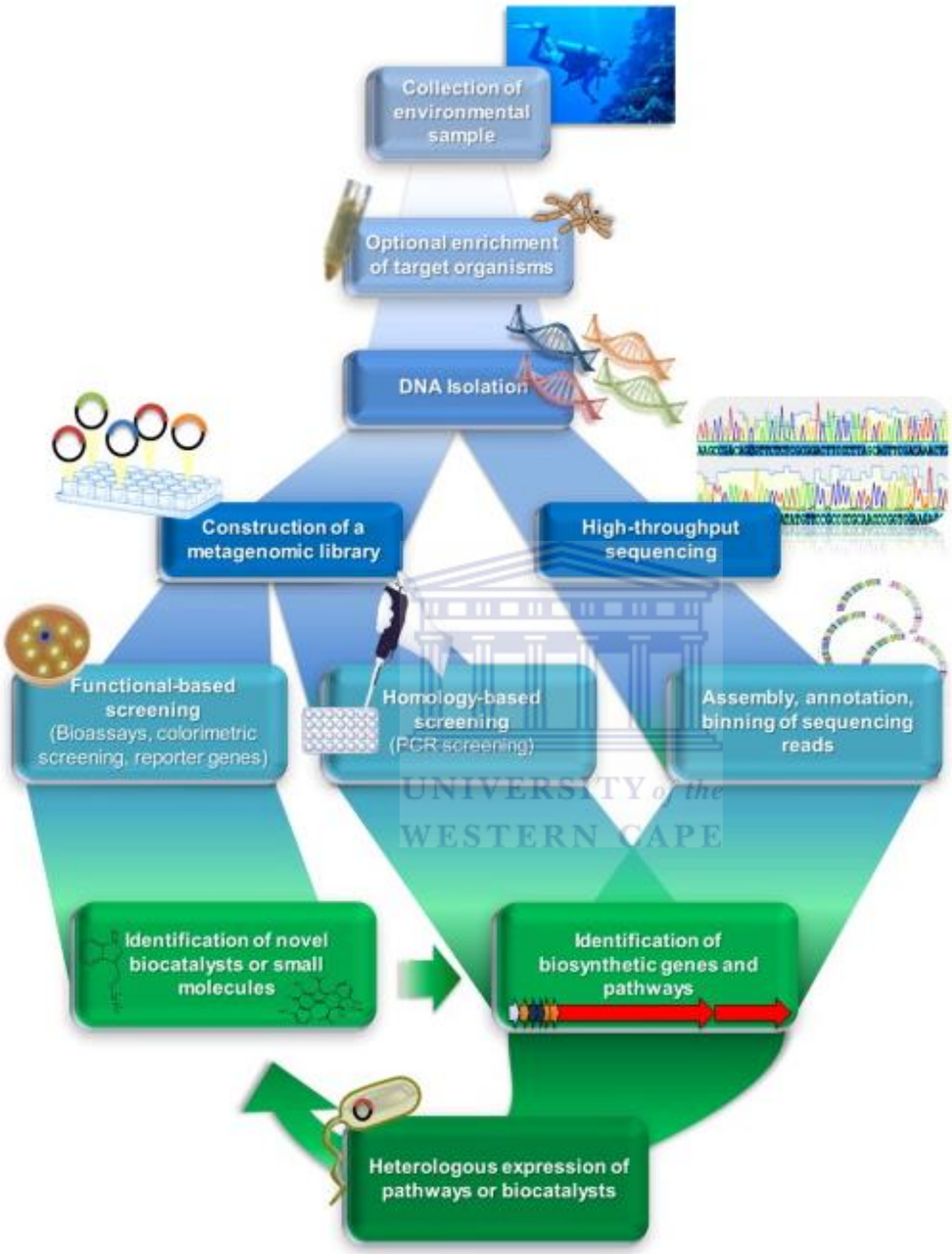


Figure 2.9: The workflow of metagenomics from mDNA extraction to studying the expression of activities by either sequence-based or function-based approaches (Wilson and Piel, 2013).

2.3 Alternative expression systems as solutions for heterologous expression

Numerous proteins which are difficult to express *in vivo* have been identified using *E. coli* cell-free machinery. These proteins include: toxic proteins (Avenaud *et al*., 2000) , membrane proteins (Isaksson *et al*., 2012), virus particles (Bundy and Swartz, 2011), and proteins with unusual amino acids (Bundy and Swartz, 2010).

2.3.1 *In vitro* protein synthesis

Cell-free protein synthesis (CFPS) also known as *in vitro* protein synthesis is a coupled transcription/translation system whereby protein synthesis is separated from metabolic activities associated with cell growth and maintenance, and thus allows the direct control of protein synthesis in a test tube wherein the transcription, translation, and protein folding components such as aminoacyl-tRNA synthetases, ribosomes, translation initiation factors, elongation factors, and chaperons are provided by the cell extract (Shrestha *et al*., 2012; Jewett *et al*., 2013). Protein synthesis is activated by addition of essential components, which include amino acids, an energy generation system (ATP, NADH, creatine phosphate, phosphoenolpyruvate), cofactors, salts and nucleotides (**Figure 2.10**) (Carlson *et al*., 2012). The CFPS system synthesizes only the target protein depending on what DNA is provided. The advantages of the CFPS system includes; the ability to produce cytotoxic proteins and the absence of a cell barrier which allows unnatural amino acids to be incorporated for protein design (Iskakova *et al*., 2006). There are several CFPS systems depending on the type of reactor used to perform the reaction namely; batch system continuous flow, semi-continuous flow and hollow fiber reactor. The commonly used one is the

batch, whereby the reaction mixture contains all the components necessary for transcription and translational system (Kim *et al*., 2006; Iskakova *et al*., 2006).

There are some toxic proteins from the cytolethal-distending toxins (CDTs) group produced by several gram positive bacteria which have shown to be difficult to express *in vivo*, (Avenaud *et al*., 2004; Ge *et al*., 2008), and also membrane proteins (Isaksson *et al*., 2012), virus particles (Bundy and Swartz, 2011), and proteins with unusual amino acids (Bundy and Swartz, 2010), which have been successfully expressed using the *in vitro* system. However, there still remain many enzyme activities which cannot be screened using *E. coli* machinery (Gabor *et al*., 2004), therefore cell free extracts from an organism other than *E. coli* must be used to overcome problems associated with poor protein folding, codon bias, proteolysis of the expressed protein, degradation of template DNA, mRNA instability and hairpin formation. This stringency could be overcome by using alternative bacterial hosts, like actinomycete species with a wider genetic background including a variety of different promoters, transcriptional and post-translational regulatory factors. These organisms are discussed in **Section 2.4**.

Since the inception of preparation of the cell extract by Moore (1967) through to current systems used by Hong *et al*. (2014), the *E. coli* based cell-free extracts have been commonly used. This is because of its well-known machinery and ease of use (Carlson *et al*., 2012). Improvements have been made to reduce costs and time, to facilitate the efficiency and yield of the system (Shrestha *et al*., 2012). These include; the use of fermenters to improve biomass yield (Liu *et al*., 2005), the use of sonication to help researchers without access to specialized equipment for cell

lysis (Shrestha *et al*., 2012) and low speed centrifugation (Kim *et al*., 1996) to reduce costs and different energy sources to improve yields (Kim and Kim, 2009).

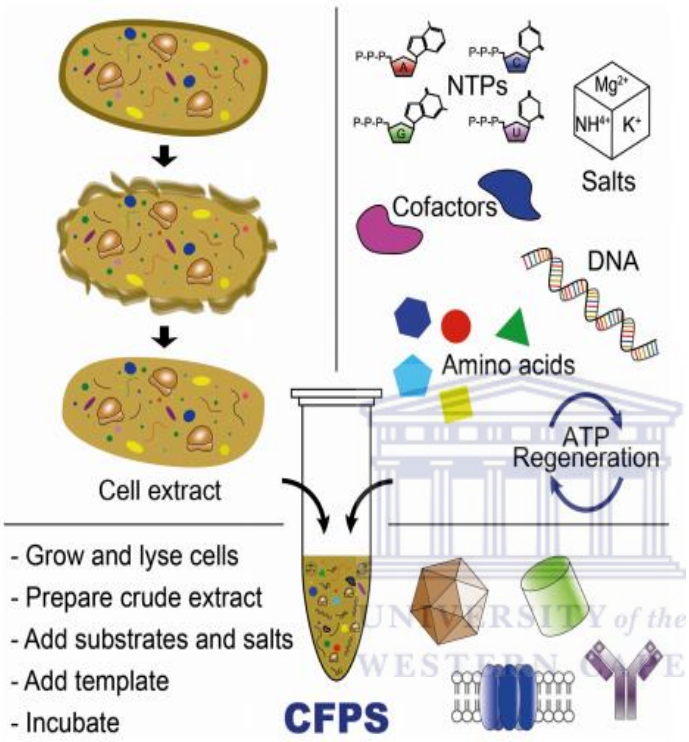


Figure 2.10: CFPS is performed using two steps: preparation of cell free extracts and protein expression. Extracts preparation involves growth of the organism, cell lysis and separation of extracts from cell debris. Protein expression involves addition of DNA and protein synthesis activation components followed by incubation to allow the production of desired protein (Adapted from Hong *et al*., 2014).

2.3.2 *In vitro* compartmentalization

The *in vitro* compartmentalization (IVC) is an emulsion technology that was first developed by Tawfik and Griffiths (1998) that uses aqueous droplets of water-in-oil (w/o) emulsion as cell like compartments. The aqueous compartments of w/o emulsion were designed by adding *in vitro* transcription-translation mixture to stirred mineral oil that contains surfactants. Its application enables the linkage of genotype (a nucleic acid that can be replicated) and phenotype (functional traits such as binding or catalytic activity) (**Figure 2.11**) (Griffiths and Tawfik, 2000).

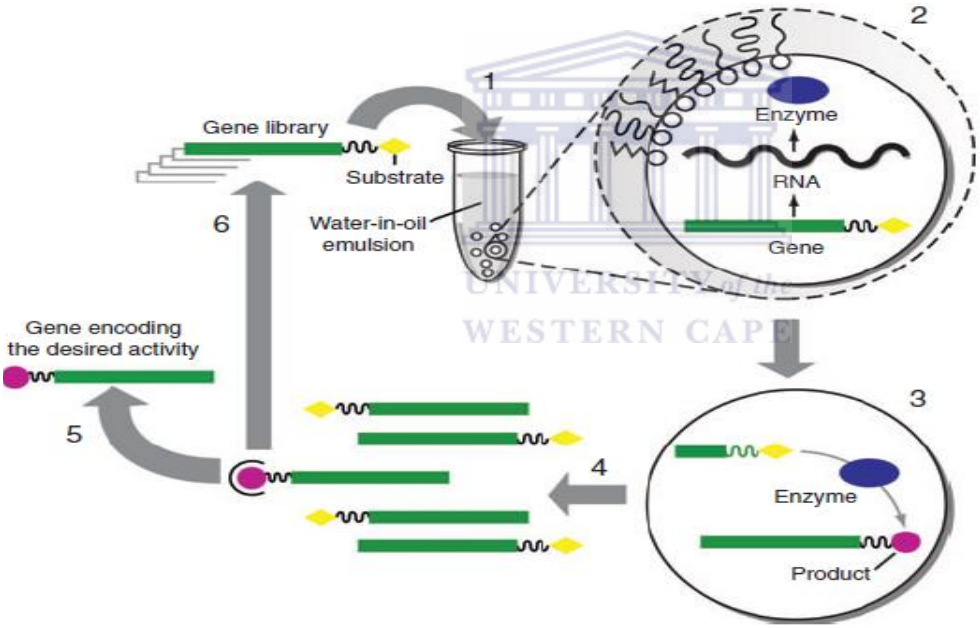


Figure 2.11: Representation of directed *in vitro* compartmentalization. Single genes are transcribed and translated inside the aqueous droplets of a water-in-oil emulsion, resulting in a gene encoding a desired activity (Adapted from Miller *et al*., 2006).

IVC provides a complete cell-free approach in order to produce novel targeted proteins by using aqueous droplets that contain DNA and components for protein production within w/o emulsions (Rothe *et al*., 2006). IVC has advantages over other *in vitro* selection techniques such as phage display, ribosome display, mRNA-peptide fusion and SELEX (systematic evolution of ligands by exponential enrichment) because it can select for properties rather than binding (e.g. catalytic or regulatory activities) and intermolecular catalysis *in trans* whereby the substrate is not linked to the catalyst, resulting in the selection of enzymes for multiple turnover. These enzymes include DNA methyltransferases, DNA polymerases, phosphotriesterases, restriction endonucleases, β-galactosidases and thiolactonases (Miller *et al*, 2006). When *in vitro* transcription-translation is replaced with only transcription systems, it is possible to select for ribozymes which catalyse Diels-Alder cycloaddition and RNA ligation *in trans* with multiple-turnovers (Agresti *et al*., 2005; Levy *et al*., 2005; Miller *et al*., 2006)

2.3.2.1 Fluorescence activated cell sorting (FACS)

FACS technology has also been used in conjunction with IVC. The FACS technology is a fully automated system which allows cells to be stained, analyzed and sorted. The aim of FACS is to capture as many particles as possible; 100 droplets per second are generated, measuring a number of parameters at the same time (**Figure 2.12**).

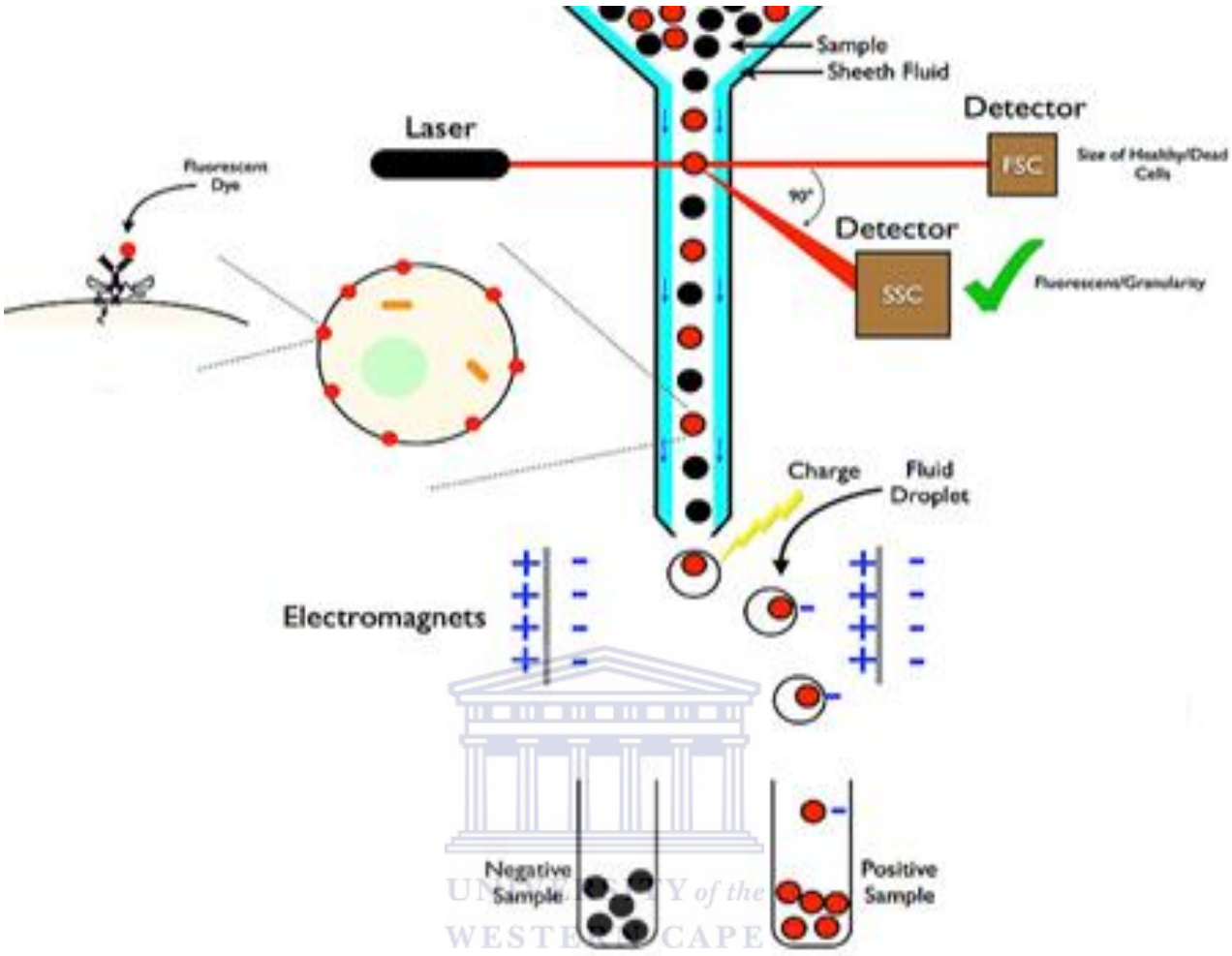


Figure 2.12: Schematic representation of how flow cytometry works (Adapted from Edwards *et al*., 2004).

It is a high-throughput technology that sorts more than 10^7 clones or emulsion droplets in one hour to allow individual cell / droplet analysis based on fluorescence and light scattering, by allowing the particles to be analysed to be introduced into a column containing pressurized sheath fluid and then pass through the laser beam, resulting in the excitation of the particle which releases the fluorescent light. The cytometer collects information about the fluorescence

characteristics released by the particle and then allows the particle to pass through the stream for the break-off distance where only particles with fluorescent signal will be charged. The charged drops pass through two high-voltage deflection plates and are deflected into collection vessels (Bernath *et al*., 2004). IVC in conjugation with FACS was used to screen for genes encoding for β-xylosidase activity in an *Actinobacteria* based cell free protein synthesis system, to demonstrate if the method works.

2.4 Actinomycetes

Actinomycetes are a group of gram positive bacteria with high guanine and cytosine content (>55%) in their genomic DNA (Ventura *et al.*, 2007). Even though most actinomycetes are free-living microorganisms, of which a few are pathogens (Willey *et al*., 2008; Hamedo and Makhlouf, 2013), they are widely spread in soil and water, as well as colonizing plants (Benizri *et al*., 2001). They resemble fungi because of their elongated cells that branch into filaments or hyphae (**Figure 2.13**). The hyphae grow on the surface of the substratum to form a dense mat of hyphae called substrate mycelium (Willey *et al.*, 2008). Actinomycetes have great ecological significance because they play an important role in the mineralization of organic matter (Bhattarai *et al*., 2007). They are prolific producers of secondary metabolites, making them useful as screening hosts for large biosynthetic pathways (Rajan and Kannabiran, 2014).

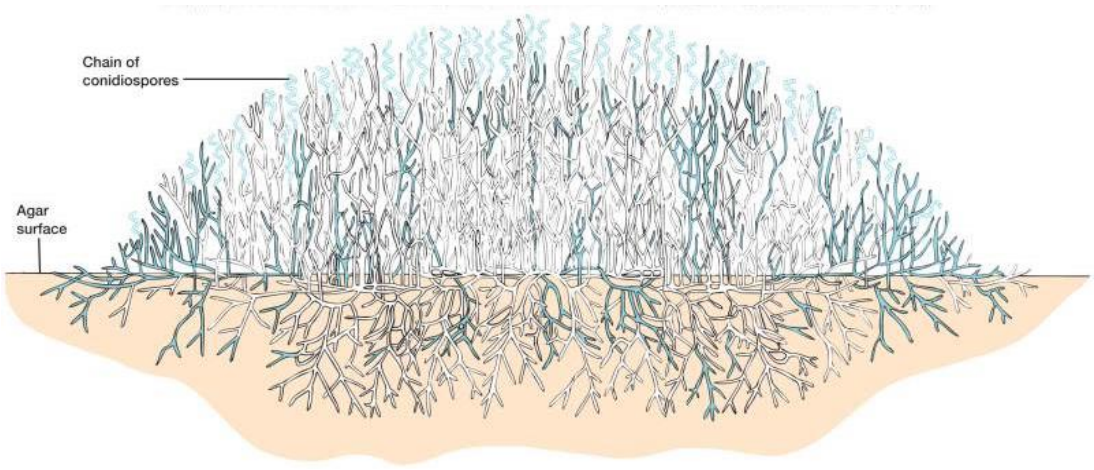


Figure 2.13: Cross section of an actinomycete colony growing on an agar plate (Adapted from Willey *et al*., 2008).

2.4.1 The genus *Rhodococcus*

The genus *Rhodococcus*, from the family *Nocardiacea* are gram positive, aerobic bacteria that are closely related to the genera *Mycobacterium* and *Corynebacterium*. Their cell wall consists of mycolic acid, arabinogalactan and peptidoglycan (Mitani *et al*., 2005). *Rhodococcus erythropolis* has been used as an alternative host for development of a protein expression system because they express proteins in a wide temperature range of 4-35°C and have successfully produced proteins that are difficult to express in *E. coli* systems (Nakashima and Tamura, 2004). For example, deoxyribonuclease I (DNAase I) has been recovered as a soluble protein when the expression is induced at 4˚C in *R. erythropolis* as compared to *E. coli* (mesophilic bacteria; 18-37˚C). *R. erythropolis* can also secrete expressed proteins into the culture medium such as members of the genus *Streptomyces*.

CHAPTER 3:
Materials and Methods

Chapter 3 Materials and Methods

3.1 General laboratory chemicals and reagents

The chemicals used in this study were supplied by Fluka, Merck Chemicals and Laboratory Supplies (Darmstadt, Germany), and Sigma Aldrich Chemical Company (Deissenhofen, Germany). The DNA size markers, protein size markers and restriction endonucleases were purchased from Fermentas Life Sciences Ltd (Vilnius, Lithuania). The oligonucleotides for polymerase chain reaction (PCR) were purchased from Inqaba Biotech (Johannesburg, South Africa).

3.2 Media

The media used are listed in **Table 3.1**. The recipes are from Sambrook and Russel (2001). All media were autoclaved at 121˚C for 20 minutes.

Table 3.1: The media used for culturing microorganisms

Constituent	1 Liter final volume
LB Medium (Lysogeny Broth)	
Tryptone	10g
Yeast extract	5g
NaCl	10g

Constituent	1 Liter final volume
TGP broth (Tryptone Glycerol Pyruvate)	
Tryptone	17g
Soy peptone	3g
K_2HPO_4	2.5g
NaCl	5g
pH to 7.3	
Addition post autoclaving	
Sodium pyruvate	4mg/ml (filter sterilize)
Glycerol	4mg/ml (filter sterilize)
CMC LB agar (Carboxymethyl cellulose)	
CMC	10g
Tryptone	10g
Yeast extract	5g
NaCl	10g
Agar	15g
CMC minimal agar	
CMC	10g
Agar	15g

After autoclaving, the media were cooled down to 55 °C before adding antibiotics **(Table 2.2)**.

Table 3.2: Stocks and final concentrations of antibiotics used in the study

Antibiotics	Preparation
Ampicillin	100mg/ml in distilled H_2O stock
	100µg/ml final concentration
Chloramphenicol	34mg/ml in 100% ethanol stock
	34µg/ml or 12.5µg final concentration
Kanamycin	200mg/ml in distilled H_2O stock
	200µg/ml final concentration
Thiostrepton	1mg/ml in distilled H_2O stock
	1µg/ml final concentration

3.3 Bacterial strains and plasmids

The bacterial strains and plasmids used in the study are listed in **Table 3.3**.

Table 3.3: Bacterial strains and plasmids used in the study.

Strain	Standard features	Reference
Escherichia coli BL21(DE3)	F^{-}ompT hsdS(rB^{-} $mB^{-)}$ gal dcm λ(DE3)	Invitrogen, USA
E. coli EPI300™	*F- mcrA D(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (Str^{R}) nupG trfA dhfr*	Epicentre Biotechnology, USA
E. coli DH5α	F- *end*A1 glnV44 thi-1 *rec*A1 *rel*A1 *gyr*A96 *deo*R *nup*G Φ80dlacZΔM15 Δ(*lac*ZYA-*arg*F)U169, *hsd*R17(rK-mK+), λ–	Invitrogen, USA
E. coli JM109	*end*A1 *gln*V44 *thi*-1 *rel*A1 *gyr*A96 *rec*A1 *mcr*B+ Δ(*lac*-*pro*AB) e14- [F' traD36 *pro*AB+ *lac*Iq lacZΔM15] hsdR17(rK-mK^{+})	Promega, USA

Strain	Standard features	Reference
pFos-Cell_11	Fosmid clone from compost metagenomic library, which carries a 3481bp cellulase gene (*Cell*_11). 24,435 bp sequence.	Tshukudu, Hons study (2012), IMBM, UWC, South Africa.
pET_Cell_11	Recombinant pET21a(+) construct, containing *Cell*_11 gene; cloned in-frame with the C-terminal His-tag; Amp^R	IMBM lab construct, UWC, South Africa.
pET-AXE10	Recombinant pET21a(+) construct, containing a 800bp acetyl xylan esterase (*axe*10) gene originally cloned from a *Streptomyces* spp, isolated from ORS10; cloned in-frame with the C-terminal His-tag; Amp^R	Gao, MSc. study (2012), IMBM, UWC, South Africa

Strain	Standard features	Reference
pET-XP12	Recombinant pET21a(+) construct, containing a 993bp cellulase gene (*xp*12) gene; cloned in-frame with the C-terminal His-tag; Amp^R	Hu, MSc. study (2012),IMBM, UWC, South Africa
pJet1.2/blunt	Suicide vector (Eco47IR), blunt DNA ends for Ligation with insert, T7 promoter, Amp^R	Fermentas, USA
pTip-RC1	PtipA Chl^R rep (pRE8424), MCS type 1	AIST, Japan (Nakashima and Tamura, 2004)
pTip-AXE10	pTip-RC1 containing *axe*10; Amp^R	This study
pTip-Cell11	pTip-RC1 containing *cell*11; Amp^R	This study
pTip-XP12	pTip-RC1 containing *xp*12; Amp^R	This study

Strain	Genotype/Standard features	Reference
pTip-W6	pTip-RC1 containing *w*6; Amp^R	This study
pET-W6	Recombinant pET21a(+) construct, containing a 1.9kb β-xylosidase gene (*w*6) gene; cloned in-frame with the C-terminal His-tag; Amp^R	Mr Nevondo W. PhD. study (2014),IMBM, UWC, South Africa
Rhodococcus erythropolis PR4	one linear plasmid, pREL1 (271 577 bp) and two circular plasmids, pREC1 (104 014 bp) and pREC2 (3637 bp)	NITE Genome Analysis Center, Department of Biotechnology, National Institute of Technology and Evaluation (NITE), Tokyo,Japan (Sekina *et al*, 2006).

Streptomyces M1146	*coelicolor*	*act*, *red*, *cpk*, *cda* deletions	Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, UK (Gomez-Escribano and Bibb, 2011).

A plasmid encoded antibiotic resistance is indicated as Amp^R, ampicillin resistance; Chl^R rep, chloramphenicol resistance, rep operon. AIST-National Institute of Advanced Industrial Science and Technology, Japan.

3.4 General Nucleic Acid manipulation

3.4 .1 Genomic DNA extraction

Genomic DNA (gDNA) was extracted from *Rhodococcus erythropolis* using a method by Wang *et al*. (1999), with modifications. Briefly, *R. erythropolis* was cultured in 10 ml TGP medium overnight from a plate culture and centrifuged for 5 minutes at 13000 x g to pellet the cells. The supernatant was removed and the pellet was washed 5 times with 1X TE buffer (10mM Tris-HCl, 1mM EDTA, pH 7.6) containing RNAseA (Fermentas) by centrifugation, and resuspension. After washing, the pellet was resuspended in 567µl TE buffer. A volume of 3µl of 20mg/ml Proteinase K was added to the resuspended cells and the mixture was incubated for 1 hour at 50 °C. After incubation, 30 µl of 10% SDS was added and incubated at 37 °C for 30 minutes. Two volumes of absolute ethanol were added to precipitate the DNA and centrifuged. The pellet was washed 5 times with 70% ethanol, air-dried, and resuspended in 100µl of 1X TE buffer (10Mm Tris-HCl, 1mM EDTA, pH8) and stored at -20 °C.

3.4.2 Fosmid extraction

E. coli Epi300™ clones harboring the recombinant fosmid construct pFos-Cell_11 were streaked out from the glycerol stock on LBA supplemented with 12.5µg/ml CAM and incubated overnight at 37 °C. A single colony was inoculated into 5ml LB medium containing CAM and incubated overnight with shaking at 37 °C. A volume of 1ml of the overnight culture was used as starter culture to inoculate 100 ml fresh LB medium containing 0.02 % (w/v) L-arabinose and 12.5µg/ml CAM, and incubated overnight at 37 °C with 160 rpm agitation. The cells were harvested by centrifugation at 4000 x g for 30 minutes at 4 °C and the fosmid DNA was extracted using the Plasmid Mini Prep Kit (Qiagen), according to the manufacturer's instructions.

3.4.3 RNA extraction

RNA was extracted from the CFPS reaction using the TRIzol® Reagent according to the manufacturer's instructions (Life Technologies). Briefly, a homogenized sample was prepared by harvesting the CFPS reaction by centrifugation at room temperature at 11 000 rpm for 5 minutes and 1 ml of TRIzol® Reagent was directly added. The contents were mixed by pipetting, 200µl of chloroform was added and vigorously mixed. The mixture was allowed to stand for 3 minutes at room temperature and centrifuged at 12 000 rpm for 15 minutes at 4 ˚C. The aqueous phase containing RNA was transferred to a clean tube, mixed with 500µl of 100% isopropanol and centrifuged at 12 000 rpm for 15 minutes at 4 ˚C. The pellet was washed with 75% ethanol and

centrifuged at 7 500 rpm for 5 minutes at 4 ˚C and air-dried for 10 minutes. The pellet was resuspended in 20-50µl of diethylpyrocarbonate (DEPC) H_2O and incubated at 60 ˚C for 10 minutes to dissolve, and stored at -70 until further use.

3.4.4 Plasmid extraction

Cultures containing plasmids were incubated in LB containing 100µg/ml Amp at 37 ˚C for overnight with shaking (150rpm), and harvested by centrifugation at 6000 x g for 10 minutes at 4 ˚C. Plasmid DNA extraction was performed using the Plasmid Mini Prep Kit (Qiagen), according to the manufacturer's instructions.

3.4.5 cDNA synthesis

cDNA synthesis was carried out using the iScript cDNA synthesis Kit (Bio-Rad, South Africa). Briefly, a 20µl reaction consisted of 4µl 5X iScript reaction mix, 1µl of iScript reverse transcriptase, total of 100ng of RNA was used as a template for the first strand synthesis and up to 20µl of nuclease free H_2O. The second strand synthesis was carried out using the first strand synthesis as a template and specific primers were used to perform a standard PCR (**Section 3.6.5**).

3.4.6 Nucleic acid quantification

The extracted gDNA, fosmid, plasmid DNA, and RNA was quantified using a NanoDrop spectrometer (ND-100: NanoDrop Technologies, USA). The OD260/280 ratio was measured and used as an overall measure of DNA sample purity.

3.5 Agarose gel electrophoresis

3.5.1 DNA

Agarose gel electrophoresis was performed according to Sambrook and Russell (2001). 0.8-1% of agarose gel was prepared. A total of 10ng of DNA sample was mixed with 2µl of 6X DNA loading buffer [20% (v/v) glycerol, 60 mM EDTA and 5mg/ml bromophenol blue] and loaded on a gel. The gel was electrophoresed in 1xTAE [0.2% (w/v) Tris-base, 0.5% (v/v) glacial acetic acid and 5M EDTA, pH 8.0] at 90V for 1 hour. The gel was then visualized using a UV transilluminator (AlphaImager 2000, Alpha Innotech, USA). The bands of interest were excised from the gel and purified using the QIAGEN Gel Purification kit according to the manufacturer's instructions.

3.5.2 RNA

Bleach gel preparation was carried out using a protocol by Aranda *et al*. (2012). Agarose gel was prepared by dissolving 1% (w/v) agarose in 1x TAE buffer, 6% of sodium hypochlorite was added and incubated at room temperature with swirling for 10 minutes. After incubation, the

mixture was microwaved until the agarose had dissolved. The molten gel was cooled by cold tap water to 56 °C. Prior to pouring the molten gel into a casting tray with combs, 0.5µg/ml ethidium bromide was added and mixed. The gel was allowed to solidify. RNA samples were mixed with 2µl of 6X DNA loading buffer and loaded on a gel. The gel was electrophoresed in 1XTAE at 100V for 1 hour, and visualized using a UV transilluminator.

3.5. 3 DNA extraction from agarose gels

A 1% (w/v) agarose gel prepared as in **Section 3.5.1** was viewed on long UV (365nm) and the band of interest was excised from the gel. The DNA extraction from agarose gel was carried out by the use of Nucleospin® Gel and PCR Clean-up kit (MACHEREY-NAGEL, USA), according to the manufacturer' s instructions.

3.6 Restriction enzyme digestion

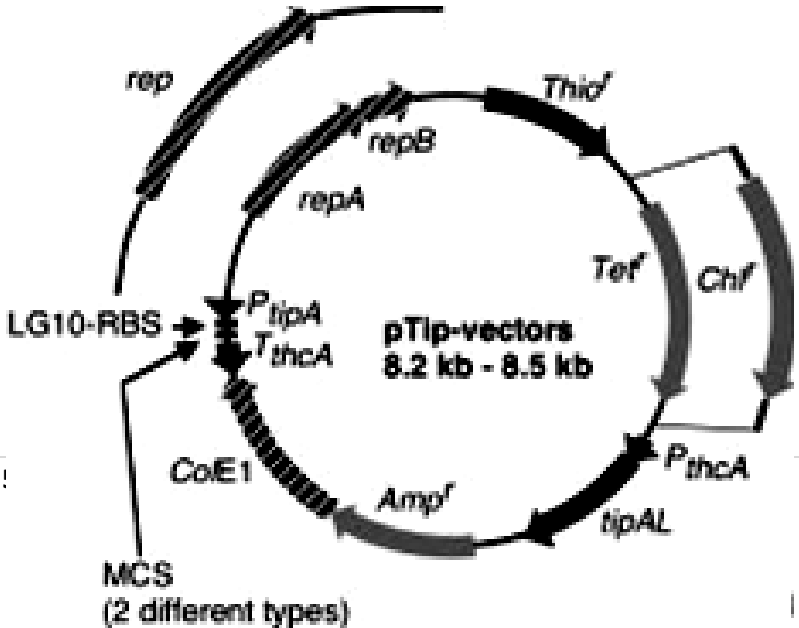
Restriction enzyme digestions were prepared in sterile 1.5 ml eppendorf tubes in 20 µl reaction volumes and were incubated at 37°C for 1 hour. Approximately 1 U of enzyme was used per µg of plasmid or gDNA in the presence of the appropriate buffer as supplied by the manufacturer. Enzymes used were: *Hind*III, *Nde*I, *Bam*HI, and *Xho*I. *Hind*III, *Bam*HI and *Xho*I were heat inactivated at 80 °C for 20 minutes, whereas *Nde*I was inactivated by incubation at 65° C for 20 minutes. The digested products were visualized on agarose gel (**Section 3.5.1**).

3.6.1 Preparation of plasmid vector pTip-RC1

The *E. coli- Rhodococcus* shuttle vector, pTip-RC1 (**Table 3.3** and **Figure 3.1**; Nakashima *et al*., 2005), was chosen for expression of genes *xp*12 and *axe*10. This vector is a fusion of the replicon of a cryptic plasmid isolated from *R. erythropolis* cloned into pRE8424 which replicates *via* a rolling circle type mechanism and contains origins of replication for both *E. coli* and *R. erythropolis.*

pTip-RC1 was transformed into *E. coli* JM109, plasmid DNA was extracted using QIAGEN Midi-prep kit according to the manufacturer's instructions, and linearized using *Nde*I and *Xho*I restriction enzymes (Fermentas). The linearized vector was resolved on a 0.8% (w/v) TAE agarose electrophoresis gel, viewed under long wavelength UV light (365nm), excised from the gel and purified using the QIAGEN Gel Purification kit according to the manufacturer's instructions.

A



B

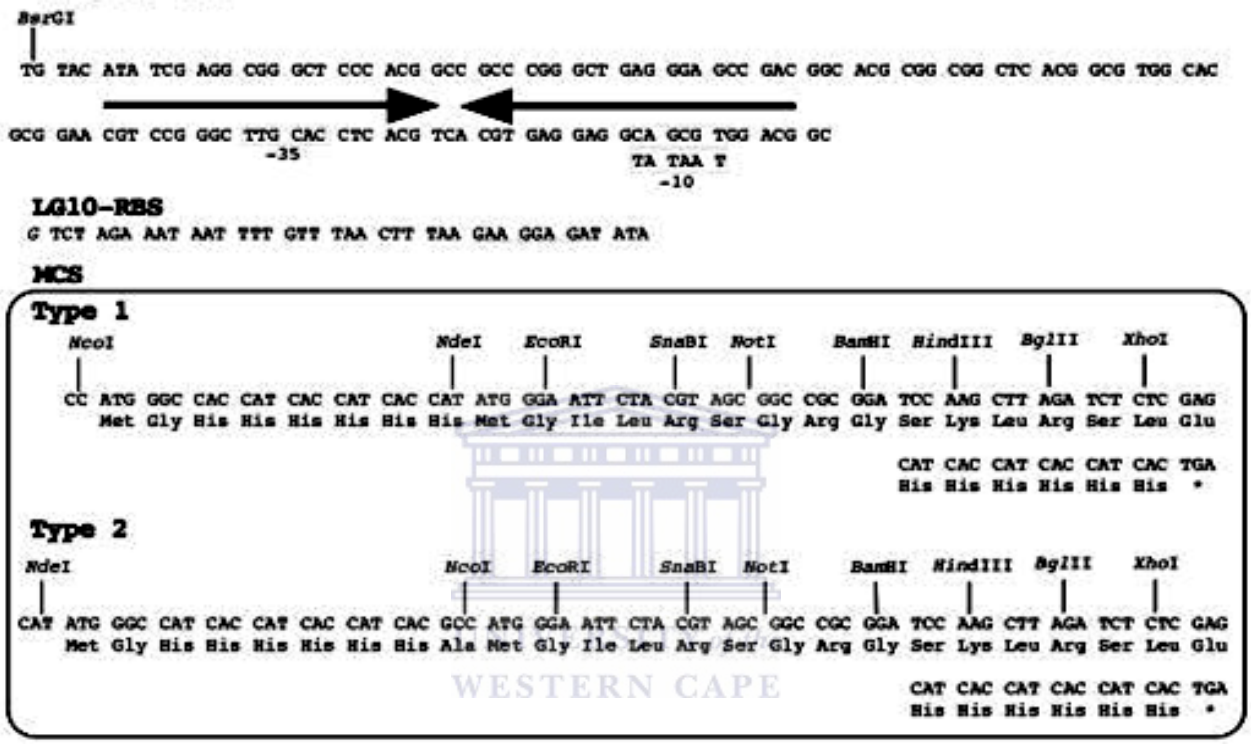


Figure 3.1: The schematic diagram of pTip vector map showing (A) the positions of restriction enzymes, *Tthc*A, *thc*A transcriptional terminator; CoE1, replication origin for *E. coli*; PtipA, *tip*A promoter; *rep*, derived from pRE8424 for autonomous replication in *R. erythropolis*; and antibiotic resistance genes; *Tet*^{r}, tetracycline resistance gene; *Thio*^{r}, thiostrepton resistance gene; *Amp*^{r}, ampicillin and *Chl*^{r}, chloramphenicol. (B) Sequences of multiple cloning site (MCS) 1 or 2 sequences (Each pTip vector either consists of MCS type 1 or type 2); LG10-RBS, the RBS is derived from gene 10 of bacteriophage *T7* and initiates

transcription at the first "*G*" nucleotide; and the -35 and -10 hexamers are boxed in the P_{tipA} promoter (Nakashima *et al*., 2005).

3.6.2 DNA ligation

Ligation reactions were performed in a 20µl volume containing the vector and insert DNA, 1x ligation buffer 1U of T4 DNA ligase, 5% of polyethyleneglycol (PEG) and distilled water. The reactions were incubated overnight at 16 °C, and subsequently transformed into *E. coli* DH5α competent cells using the heat shock method (Sambrook and Maniatis, 1989).

3.6.3 Preparation of competent cells

Chemically- and eletro-competent cells were prepared according to Sambrook and Russell, (2001).

3.6.4 Transformation of competent cells

E. coli DH5α competent cells were transformed by the addition of 5µl ligated DNA to 50 µl of the competent cells in a sterile 2 ml eppendorf tube. The cells were incubated on ice for 30 minutes prior to addition of ligated DNA, followed by heat shock at 42 °C for 2 minutes and then incubated on ice for 2 minutes. After the addition of 950 µl of LB medium, the transformation mixtures were incubated for 1 hour at 37 °C to allow recovery of transformed cells. A volume of

100 µl of cells was plated on LB-agar plates supplemented with 100 µg/µl Amp. Agar plates were incubated overnight at 37 °C.

3.6.5 Polymerase chain reaction

Polymerase chain reaction (PCR) amplification was performed in 0.2ml thin-walled tubes using the thermal cycler Gene Amp PCR system. Each PCR reaction (final volume of 50 µl) contained 1 X Dream*Taq* buffer (Fermentas), 2 mM $MgCl_2$, 0.2U Dream*Taq* DNA polymerase (Fermentas), 200µM of dNTPs (Bioline), 5µM of each primer and 50ng of DNA template. The primers used and the cycling conditions are listed in **Table 3.4**.

Colony PCR was also performed using same condition.

Table 3.4: Primers used in this study.

Primer	Primer sequence (5'-3')	PCR conditions	References
E9F	GAGTTTGATCCTGGCTCAG	94 ˚C/4mins	Hansen *et al*., 1998
U1510R	GGTTACCTTGTTACGACTT	30X (94 ˚C/30s, 52 ˚C/30s, 72 ˚C/10s), 72 ˚C/10mins	Reysenbach and Pace, 1995
pJET 1.2 blunt Forward	CGACTCACTATAGGGAGAG CGGC		Fermentas, USA
pJET 1.2 blunt Reverse	AAGAACATCGATTTTCCAT GGCAG		Fermentas, USA
AXE10_Fwd	GACACATAGTCCACTGCTC TCTGC	96 ˚C/2mins 30X (96˚C/45s, 56 ˚C/30s, 72 ˚C/2min), 72 ˚C/10mins	Gao, MSc. study (2012), IMBM, UWC, South Africa
AXE_Rev	GCTCGAGGGCGCGGACGTC G		Gao, MSc. study (2012), IMBM, UWC, South Africa

Primer	Primer sequence (5'-3')	PCR conditions	References
Gene12FW	CATATGCGAAAACCCGTCTGGC GC	94 ˚C/4mins 30X (94 ˚C/30s, 57 ˚C/30s, 70 ˚C/10s), 70 ˚C/10mins	Hu, MSc. study (2012), IMBM, UWC, South Africa
Gene12XhoIRV	GCTCGAGTTGCTGGGCGATCCACAC CAG		Hu, MSc. study (2012), IMBM, UWC, South Africa

3.6.7 DNA sequence determination

Sequencing was performed with an automated sequencer model (ABI PRISM 377) at the University of Stellenbosch. The sequencing reactions were carried out using 16S rRNA, AXE10 and XP12 primers (**Table 3.4**). The sequences were edited using Chromas Lite version 2.01 (Technelysium, Australia), assembled using DNAMAN version 4.13 (Lynnon Biosoft), and then aligned using nucleotide-nucleotide basic local algorithm search tool (BLASTn; http://blast.be-md.ncbi.nlm.nih.gov/).

3.7 Preparation of cell free extracts

E. coli and *R. erythropolis* cells were cultured at 37 °C and 30 °C, respectively in 1.5 L of TGP medium for *R. erythropolis* and LB medium for *E. coli* with 400 rpm agitation and aeration using an INFORS HT Minifors (bioreactor). The cells were harvested in the mid-log phase (OD_{600}~2.6) by centrifugation at 6000 rpm at 4 °C for 30 minutes and washed three times by resuspending in 20 ml ice-cold buffer A [(10mM Tris-acetate buffer (pH 8.2), 14mM magnesium acetate, 60mM potassium glutamate, and 1mM dithiothreitol (DTT) containing 0.05% (v/v) 2-mercaptoethanol]. The cells were stored at -80 °C for 15 minutes. The thawed cells were re-suspended in 1 ml buffer B (buffer A without 2-mercaptoethanol) prior to cell lysis. Cell suspensions were lysed using the Bandelin Sonopuls sonicator for 6 x 30 sec at 0.5% power using MS 73 Titanium microtip horn, the samples were kept on ice to prevent the sample from heating during sonication. The lysate was centrifuged for 30 min at 6000 rpm at 4 °C and the supernatant was retained. The run-off reaction to stop transcription

was performed by incubating the supernatant at 30 °C with 280 rpm agitation. The extract was flash-frozen in liquid nitrogen and stored at – 80 °C until use (Kim *et al*., 2006).

3.8 Cell free protein synthesis

CFPS using *R. erythropolis* and *E. coli* cell free extracts were carried out using a method adapted from Kim *et al*. (2006). Typically, a 50µl reaction mixture consisted of 57 mM HEPES-KOH (pH 8.2), 1.2 mM ATP, 0.85 mM each of CTP, GTP, and UTP, 2 mM DTT, 0.17 mg/ml total tRNA mixture, 0.64mM cAMP, 90 mM potassium glutamate, 80 mM ammonium acetate, 12 mM magnesium acetate, 34 µg/ml folinic acid, 1.5 mM of each 20 amino acids, 2 % (w/v) PEG (8000), 67 mM creatine phosphate, 3.2 µg/ml creatine kinase, 3 µg/µl DNA (10ng for double emulsion CFPS) and 50 % of cellular extract. The reactions were incubated at 30ºC for two hours. *E. coli* S30 cell free protein synthesis was carried out using the *E. coli* S30 system for linear DNA kit (Promega) according to manufacturer's recommendations.

3.9 Protein analysis techniques

3.9.1 Bradford assay

The Bradford assay (Bradford, 1976) was used to determine the concentration of protein. Bovine serum albumin (Sigma-Aldrich) was used as a standard (**APPENDIX E**). The reactions were incubated at room temperature, and the absorbance was recorded at 595nm.

3.9.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate proteins according to size by using the Laemmli (1970) method. Briefly, 12% separating and 4% stacking gels were first cast using a Hoefer Mighty Small System (Hoefer, USA) (**APPENDIX F**). An equal volume of protein sample was mixed with 2X loading buffer [10% glycerol (v/v), 1% β-mercaptoethanol (v/v), 0.2% bromophenol blue (w/v), and 80 mM Tris-HCl pH 6.8] and denatured at 95 ˚C for 10 minutes. The denatured samples were loaded onto the polyacrylamide gel and separated using constant 100 V until the bromophenol blue loading dye reached the bottom of the gel. The polyacrylamide gel was removed from the plates, stained with Coomassie staining solution [10% glacial acetic acid (v/v), 30% methanol (v/v), 0.25% Coomassie Brilliant Blue R250 (w/v) and 60% distilled H_2O] overnight and destained with destaining solution [30% methanol (v/v), 10% glacial acetic acid (v/v) and 60% distilled H_2O] overnight. The Pre-stained PageRuler (10 to 170kDa) (product # 26616) (Fermentas, Germany) was used to determine the sizes of the proteins.

3.9.3 His-Tag affinity chromatography

AXE10 and XP12 proteins were purified using Ni-chelating His-Bind resin and buffer kit (Novagen) according to the manufacturer's instructions. Briefly, the His-Bind resin was washed with three volumes of sterile deionized H_2O, charged with five volumes of charge buffer (50 mM $NiSO_4$) and equilibrated with three volumes of binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.9). Cell free protein synthesized products using pTip-XP12 and pTip-AXE were loaded on the columns and allowed to flow through under gravity. The columns were washed with ten volumes of 1X Binding buffer and six

volumes of 1X wash buffer (60 mM imidazole, 0.5 NaCl, and 20 mM Tris-HCl, pH 7.9). The bound proteins were eluted with six volumes of 1X elution buffer (62.5 mM imidazole, 0.5 mM NaCl, and 20 mM Tris-HCl, pH 7.9). The eluted proteins were concentrated at 40 ˚C using a speedyvac (Labconco) and separated using a 12% SDS-PAGE gel (**Section 3.9.2**).

3.9.4 MALDI-TOF Mass Spectrometry

AXE10, Cell11 and XP12 protein bands were purified from SDS-PAGE gels and washed with acetonitrile (ACN) until the Coomassie stain was washed off (gel pieces turns white). The ACN was evaporated at 50 ˚C using a speedyvac (Labconco). The washed gel-extracted protein bands were analysed at the Stellenbosch University central analytic facility (CAF) with electronspray ionization (ESI) mass spectrometry.

3.10 Evaluating cell-free expression using enzyme assays

3.10.1 Acetyl xylan esterase activity screening

The CFPS reaction with pTip-AXE10 as a template was assayed for AXE activity using the Fast Garnet liquid assay (Koseki *et al*., 2005). Briefly, a 200µl reaction consisting of 170µl 50mM NaH_2PO_4 buffer (pH 7.0), 20µl µmol α-naphthyl acetate (Fluka) and 10µl CFPS. The reaction was stopped with the addition of 110µl Fast Garnet [6mg/ml dissolved into 10% SDS (w/v); Sigma]. The absorbance was determined at 560nm by mixing 50µl of the assay reaction and 950µl of distilled H_2O. A standard curve was constructed using α-naphthol in the concentration range of 0.25-5mM to generate this equation; $y = 0.273x + 0.026$ (**APPENDIX G**).

3.10.2 Cellulase activity screening

The cellulase activity assay was carried out using the Dinitrosalicylic acid (DNS) reagent [5g DNS (Sigma), 150g potassium sodium tartrate (Rochelle salts), 10g phenol, and 0.5g sodium hydroxide]. An amount of 2mg/ml of glucose monohydrate was prepared as a stock and diluted with H_2O to 0.1mg/ml, 0.25mg/ml, 0.5mg/ml, 1.0mg/ml, and 1.33mg/ml to generate a standard curve. A volume of 500µl of the DNS reagent was added to 500µl of each diluted glucose monohydrate, boiled for 15 minutes and cooled on ice. A volume of 1ml of H_2O was added to the reaction and the tube was inverted to obtain a homogenized solution. The absorbance was measured at 540nm.

The cellulase activity of Cell11 and XP12 in CFPS reactions were carried out as follows; the 1ml assay consisted of 500µl of 2% (w/v) carboxymethyl cellulose (CMC) dissolved in sodium citrate buffer (pH 4.8) and 500µl of either Cell11 or XP12 CFPS, and incubated at 30 ˚C in a waterbath overnight. A volume of 1ml of DNS reagent was added to each mixture and boiled for 15 minutes. The reaction mixtures were cooled down by placing the tubes on ice and addition of 1ml H_2O before recording the absorbance at 540nm. The concentrations of Cell11 and XP12 were measured in mg/ml using the equation $y = 0.600x + 0.019$ (where $y = OD_{540}$ and x= concentration of the unknown) from the glucose monohydrate standard curve in mg/ml.

Rhodococcus erythropolis, Streptomyces coelicolor, and *E. coli* harbouring the following constructs; pFOSCell11, pTip-XP12 and pET-XP12 were streaked out onto CAM LB agar

plates containing CAM (12.5 µg/µl) and incubated overnight or for three days (in the case of *S. coelicolor*). The plates were flooded with Congo red. After discarding the Congo red the plates were flooded with 1 M NaCl for 30 mins and finally flooded with 1 M HCl for 10 second to visualize a zone of clearing (Kasana *et al.*, 2008).

3.10.3 β-xylosidase activity assays

The CFPS products using plasmid constructs encoding β-xylosidase gene (pTip-W6 and pET-W6; Nevondo, 2014) as templates, were used as crude enzymes for β-xylosidase activity using 0.5mM *ρ*-nitrophenyl-β-D-xylopyranosidase (*ρ*NPX) as substrate. Briefly, 10µl of crude enzyme was mixed with 10µl of *ρ*NPX and 125µl of 0.2M $NaP0_4$, pH 8. The reaction mixture was incubated at 30 ˚C overnight and stopped by the addition of 1M sodium carbonate. The *ρ*-nitrophenol product was quantified at 405nm using absorbance coefficient of 18.5ml.µmole^{-1}.cm^{-1} (Krisana *et al*., 2005).

3.10.4 Zymograms

Zymograms were prepared for the determination of acetyl xylan esterase and cellulase activity, according to the method by Watson *et al*. (2009) with modifications. CFPS samples were run in SDS-PAGE gels in duplicate; one portion was stained with coomassie brilliant blue and destained as described in **Section 3.9.2**, and the other portion was washed with 2.5% Triton X-100, stained with 0.8µmol α-naphthyl acetate substrate overnight and then stopped by the addition of 0.1ml Fast Garnet GBC sulphate salt [6mg/ml in 10% SDS; Fluka] for acetyl xylan esterase activity. For cellulase activity, gel contained 1% CMC (w/v) as a

substrate, the gel was stained with 0.1% Congo red for 1 hour and washed with 1M NaCl solution for 1 hour.

CHAPTER 4:
Results and Discussion

Chapter 4 Results and Discussion

4.1 Selection and construction of test plasmid construct

Template DNA is needed for proteins to be expressed using the CFPS system. The CFPS system synthesizes only the proteins which are encoded on the DNA template provided. Therefore, for fosmid and plasmid constructs all genes encoded will be transcribed and translated. In order to determine the breadth of expression level using *R. erythropolis*, a number of different plasmid and fosmid constructs were selected to test the expression of genes from their native promoters and also from promoters provided on the pTip expression system developed for expression in *Rhodococcus*. Expression was also compared using *E. coli* expression vectors and extract. All the plasmid constructs and lignocellulase harbouring fosmids used in this study are summarized in **Table 4.1**.

Three previously identified lignocellulosic genes were selected, *axe*10, *xp*12 and *cell*11. The *axe*10 was originally identified in *Streptomyces achromogenes* ORS10, (Gao H. MSc. thesis, 2012). The *xp*12 cellulase was isolated from a metagenomics library screen of DNA from Mphizi hot spring, Malawi (Hu X. MSc. Thesis, 2010). pFos-Cell_11 was previously identified from a metagenomic library constructed from high-temperature composts (Tshukudu O. Hons. Thesis, 2012).

These genes were chosen for several reasons. The first is that they are potentially useful in lignocellulosic degradation. The second is that each has a distinct activity (acetyl xylan esterase and cellulase), allowing one to assay for an enzyme activity which *Rhodococcus*

might not have. Also, being from different sources, these genes potentially represent different codon profiles and allow one to test the extracts ability to produce these proteins.

Table 4.1: Constructs used in this study.

Construct name and type	Gene	Promoter type	Protein activity	Used for	Source
pFos-cell11 / fosmid	*cell11*	native	Cellulase	Test native promoter and protein folding	IMBM lab construct, previous study, O. Tshukudu
pFos-XP12 / fosmid	*xp12*	native	Cellulase	Test native promoter and protein folding	IMBM lab construct, previous study, X. P Hu
pET-XP12 / expression	*xp12*	*T*7	Cellulase	Protein folding	IMBM lab construct, previous study, X. P Hu
pTip-XP12 / expression	*xp12*	P_{tipA}	Cellulase	Protein folding	This study
pTip_AXE10 / expression	*axe10*	P_{tipA}	Acetyl xylan esterase	Protein folding	This study
pET-Axe10 / expression	*axe10*	*T*7	Acetyl xylan esterase	Protein folding	IMBM lab construct, previous study, Y. Gao
pTip-W6	*w6*	P_{tipA}	β-xylosidase	Protein folding	This study

Construct name and type	**Gene**	**Promoter type**	**Protein activity**	**Used for**	**Source**
pET-W6	*w6*	*T7*	β-xylosidase	Protein folding	IMBM lab construct, previous study, W. Nevondo
W6	*w6*	native	β-xylosidase	Test native promoter and protein folding	IMBM lab construct, previous study, W. Nevondo

pTip-RC1 is a vector which was constructed for the over-expression of proteins in *Rhodococcus*, providing an inducible (thiostrepton) promoter and the ability to attach a poly-histidine tag to the protein of interest (Nakashima and Tamura, 2004). The expression levels of proteins in *R. erythropolis* using this vector are high as compared with *E. coli* because of LG10-RBS (LG10-ribosome binding site), from the bacteriophage *T7* (Nakashima *et al*., 2005).

To prepare the constructs of interest, plasmid pTip-RC1 vector was linearised using *Nde*I and *Xho*I (**Figure 4.1**) into which *axe*10 (**Figure 4.2**) and *xp*12 (**Figure 4.3**), digested with the same enzymes were cloned.

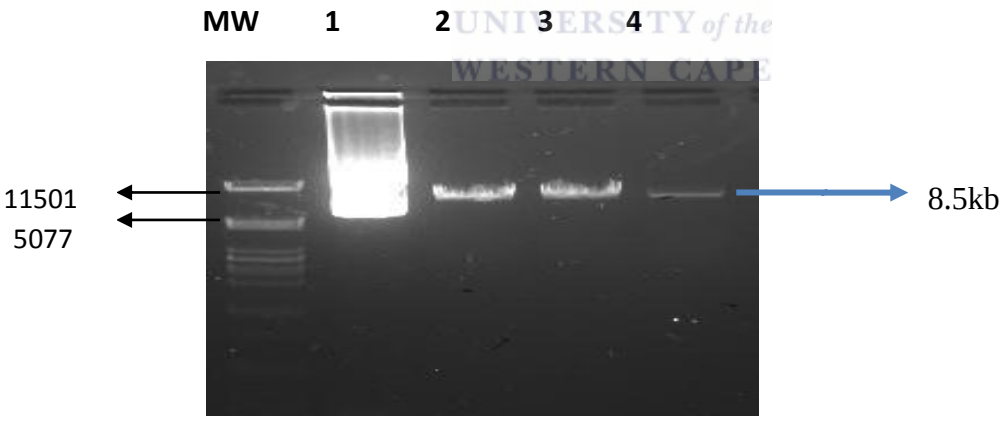


Figure 4.1: pTip-RC1 digested by *Nde*I and *Xho*I. MW-λ-*Pst*I ladder, lane 1-pTip-RC1undigested, lane 2- digested with *Nde*I only, lane 3- digested with *Xho*I only, and lane 4- double digestion with *Nde*I / *Xho*I.

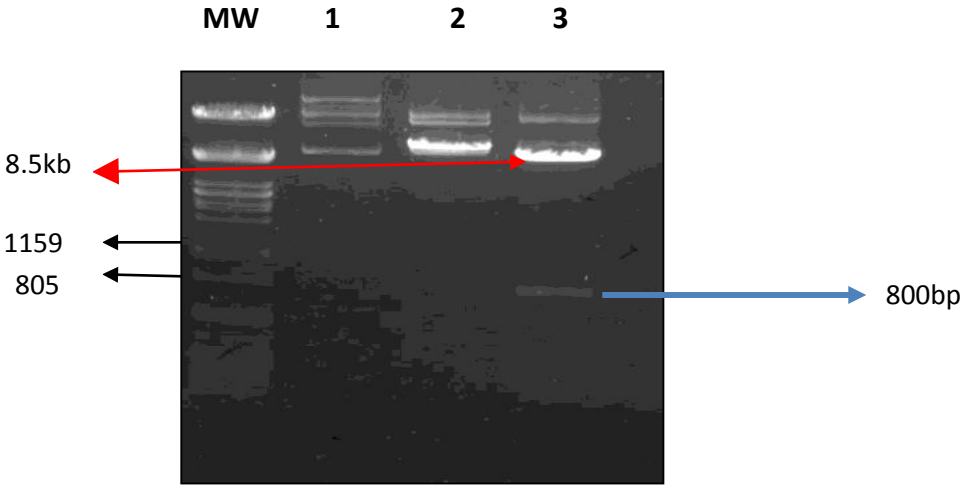


Figure 4.2: pET21_AXE10 was digested with *Nde*I and *Xho*I to release the 800 bp *axe*10 gene. MW- λ-*Pst*I marker, lane 1- undigested pET21_AXE10, lane 2- pET21_AXE10 digested with *Nde*I only, and lane 3- pET21_AXE10 double digested with *Nde*I / *Xho*I.

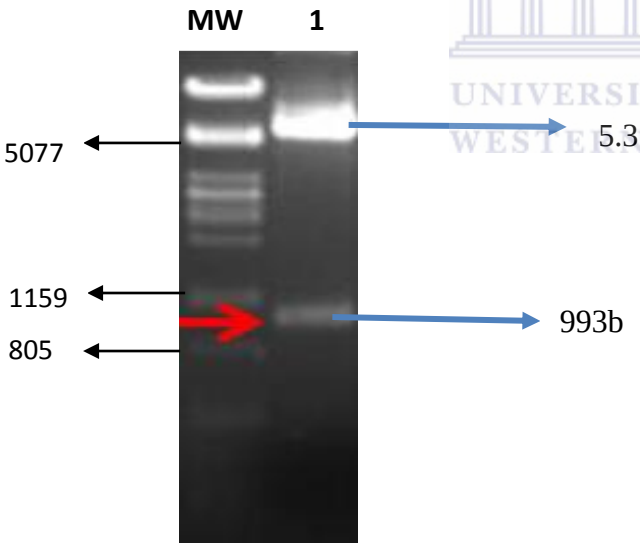


Figure 4.3: pET-XP12 was digested with *Nde*I and *Xho*I to release the 993bp *xp*12 gene (Lane 1).

The concentrations of the pTip-RC1 and AXE10 gene were measured to be both 20ng/µl after gel purification. After the insert and the vector were ligated using T4 DNA Ligase, the ligation mixture was transformed into DH5α, resulting in a 9.3kb pTip_AXE construct. The construct was confirmed by performing a double digest with *Nde*I and *Xho*I (**Figure 4.4**), confirming the sizes of *axe*10 gene to be 800bp (Degrassi *et al*., 1998; Koseki *et al*., 2005; IMBM lab construct, 2012).

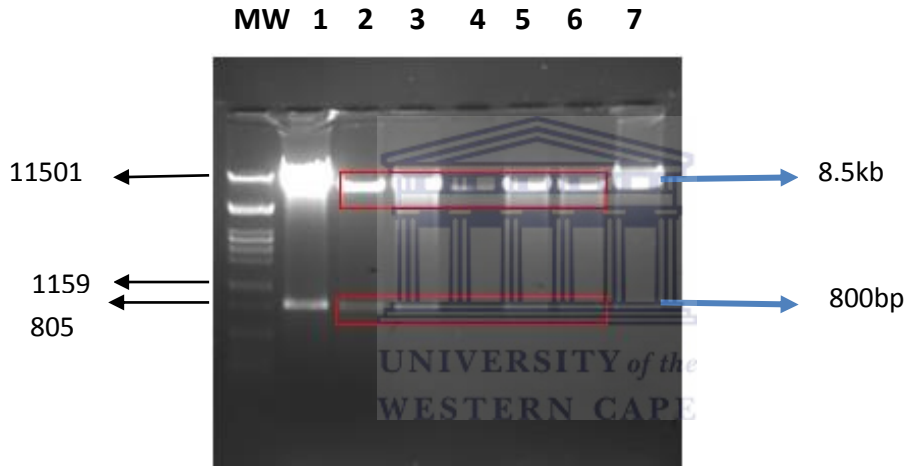


Figure 4.4: Restriction analysis of pTip_AXE10. Transformants were digested with *Nde*I and *Xho*I. MW-λ-*Pst*I DNA molecular weight marker; lanes 1-7, random clones digested with *Nde*I and *Xho*I. The 8.5kb vector backbone and 800bp insert were present.

The pFos-Cell_11 was confirmed through restriction digestion with *Bam*HI / *Hind*III **Figure 4.5**, **lane 2** showing the 3481bp *cell*_11-containing fragment and the 8.1kb pCC1FOS™.

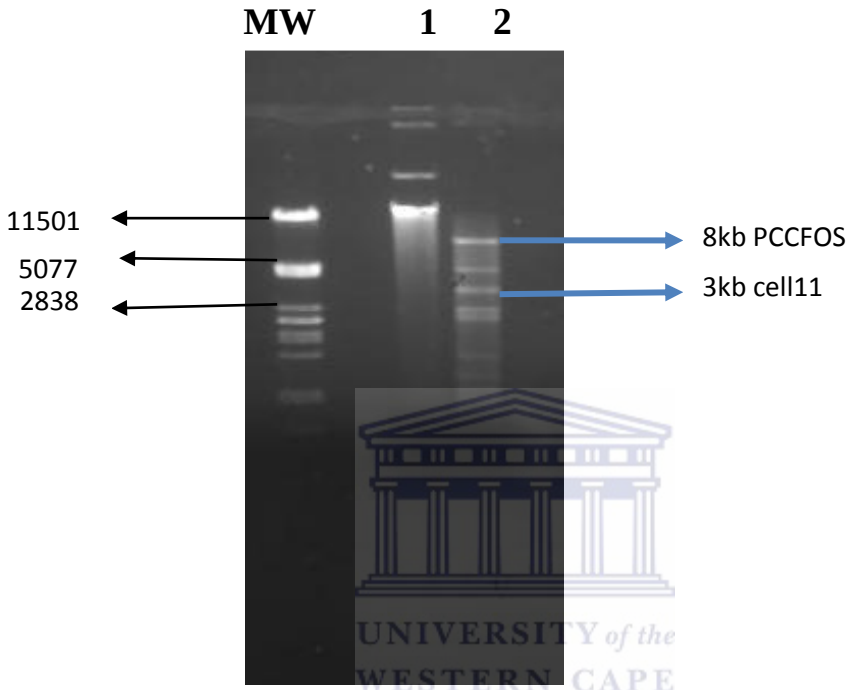


Figure 4.5: Extraction and digestion of pFos-Cell_11 with *Bam*HI / *Hind*III. MW-λ-*Pst*I DNA marker, lane 1-undigested pFos-Cell_11, and lane 2- pFos-Cell_11 digested with *Bam*HI and *Hind*III.

The pFos-XP_12 was confirmed through restriction digestion with *Bam*HI/*Hind*III and PCR amplification using gene12FW and gene12XhoIRV primers (**Table 3.4**), confirming the size of *xp*12 to be 993bp (**Figure 4.6**).

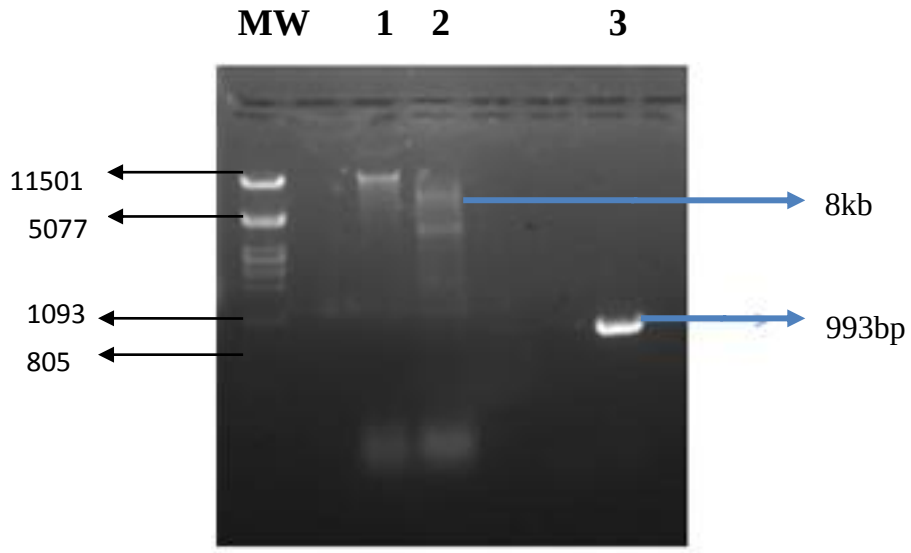


Figure 4.6: Extraction, digestion and amplification of pFos-XP_12. MW-λ-*Pst*I DNA marker, lane 1- undigested pFos-XP_12, lane 2- pFos-XP_12 digested with *Bam*HI / *Hind*III, and lane 3-XP12 PCR amplicon using pFos-XP_12 as a template.

A previously identified DNA fragment encoding β-xylosidase which was cloned behind inducible *E. coli* promoter (pET-W6) was double digested with *Bam*HI/*Xho*I to release 1.9kb *w*6 gene (**Figure 4.7**) (discussed in details in **Section 4.3**). pTip-RC1 vector was linearised using *Nde*I and *Xho*I (**Figure 4.7**) into which *w*6 was cloned. The pTip-W6 was confirmed through restriction digestion with *Bam*HI/*Xho*I, the size of *w*6 to be 1.9kb (**Figure 4.8**).

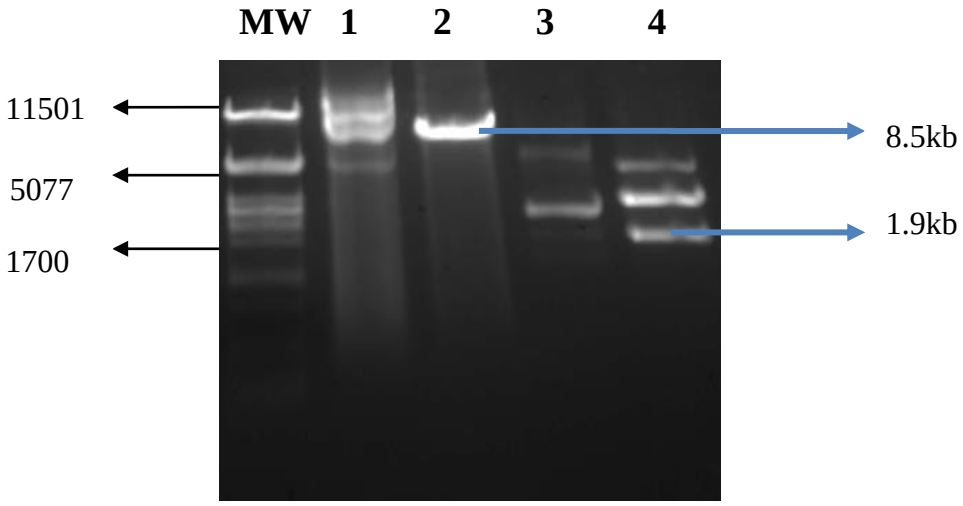


Figure 4.7: pTip-RC1 and pET-W6 digested with *Bam*HI and *Xho*I. MW-λ-PstI ladder, lane 1-pTip-RC1undigested, lane 2- pTip-RC1 digested with *Bam*HI only, lane 3- undigested pET-W6, and lane 4- pET-W6 double digested with *Bam*HI / *Xho*I.

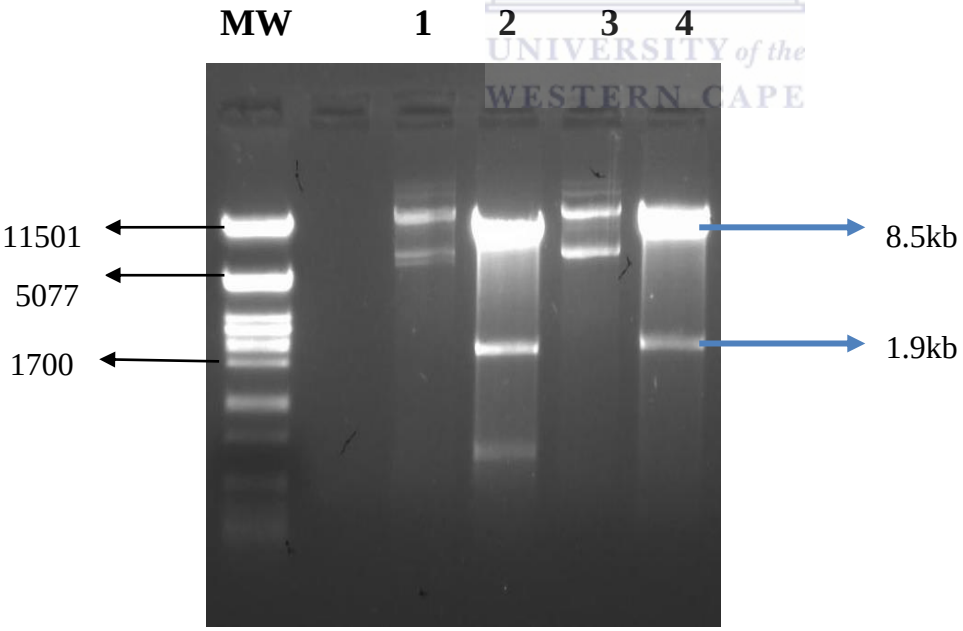


Figure 4.8: Restriction analysis of pTip_W6. Transformants were digested with *Bam*HI and *Xho*I. MW-λ-*Pst*I DNA molecular weight marker, lanes 1&3- undigested random clones, and lanes 2&4 digested pTip-W6 with *Bam*HI and *Xho*I. The 8.5kb vector backbone and 1900bp insert were present.

4.2 Preparation of *R. erythropolis* cell-free extract and *in vitro* expression of heterologous genes

R. erythropolis cell-free extract was used to address some of the problems associated with low yield or no protein yield in heterologous expression systems. For example, when a protein of interest contains codons of which the host cell-free extract contains few tRNAs, the ribosome may struggle to find tRNA to proceed with synthesis, resulting in no or low protein yields (Promega guide, 2007). *R. erythropolis* is known to express proteins over a wide range of temperatures (4-35 ˚C). Lower temperatures slow down translation to give time for the correct tRNA to be recruited (Nakashima and Tamura, 2004), therefore all the CFPS using *R. erythropolis* cell-free extracts were carried out at 30 ˚C. *E. coli* cell-free extract was used in this study for comparison with the *R. erythropolis* cell-free protein synthesis system since it is well developed system (Carlson *et al.*, 2012).

4.2.1 RNA extraction from CFPS products

To confirm that DNA has been transcribed to mRNA, mRNA from the CFPS reactions using pTip-AXE10 and pTip-XP12 as templates was extracted (**Figure 4.9**). RNA concentrations before and after DNase treatment are recorded in **Table 4.2**. The RNA purity ranged from 1.84 to 2.00; indicating that highly pure RNA was extracted.

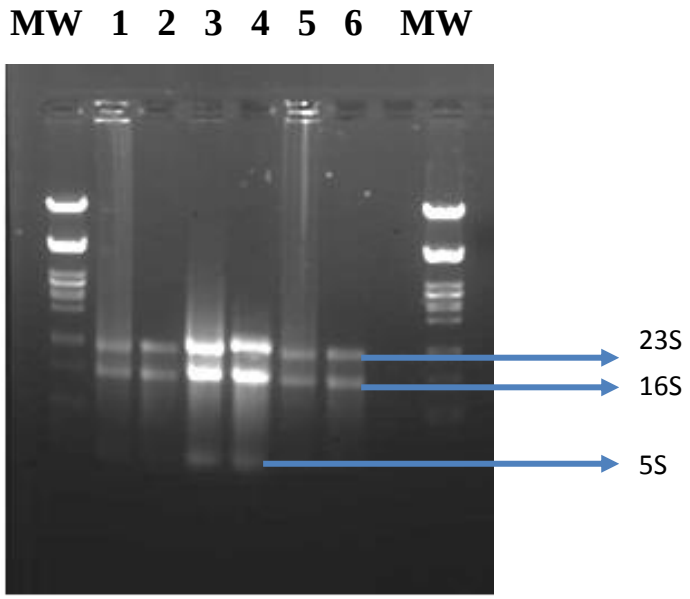


Figure 4.9: A 1% (w/v) agarose gel analysis of extracted RNA samples from CFPS expression using *R. erythropolis* cell-free extract showing traces of gDNA (presence of smear), and treated with DNase (absence of smear). MW-λ-*Pst*1 DNA molecular weight marker, lane 1- untreated pFos-XP_12, lane 2-treated pFos-XP_12, lane 3-untreated pTip-AXE10, lane 4- treated pTip-AXE10, lane 5-untreated pTip-XP12, lane 6- treated pTip-XP12.

Table 4.2: The RNA concentrations from pFos-XP_12, pTip-AXE10, and pTip-XP12, before and after DNAase treatment. The RNA purity was also measured after DNAase treatment.

Sample	Concentration before DNAase treatment	Concentration after DNAase treatment	260/280
pFos-XP_12	689.0 ng/µl	609.1 ng/µl	1.90
pTip-AXE10	760.8 ng/µl	710.9 ng/µl	2.00
pTip-XP12	680.5 ng/µl	600.0 ng/µl	1.84

The 16S rRNA PCR was performed on DNase-treated RNA sample to have assurance of the absence of contaminating DNA. It was shown by the absence of 16S rRNA amplification (1500bp) as shown in **Figure 4.10**. *R. erythropolis* gDNA was used as a positive control. This is the most crucial step as the absence of gDNA eliminates the chance of false positive results in the subsequent Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) analysis (Sheridan *et al.*, 1998).

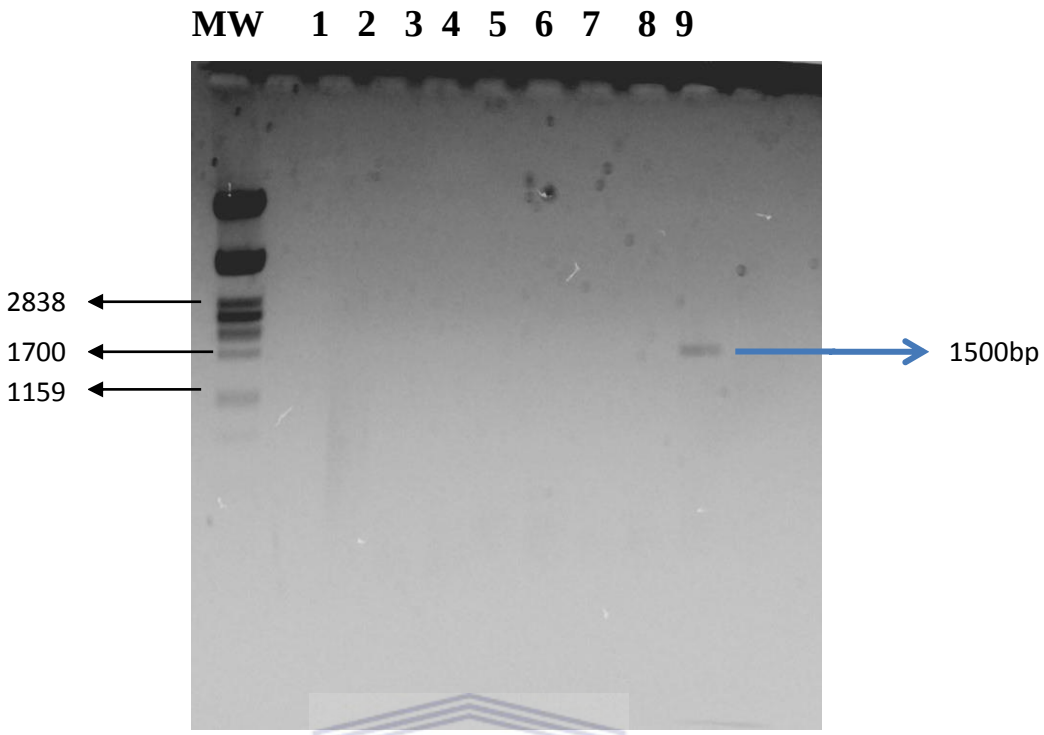


Figure 4.10: PCR test to show the absence of 16S rRNA in DNAase-treated total RNA samples. MW-λ-*Pst*I DNA molecular weight marker; lanes 1-8, DNAase treated samples; lane 9-positive control.

4.2.2.1 cDNA synthesis

The synthesis of the first strand was carried out. The cDNAs produced are not full length, which is why they are generally not visible in a gel. Single stranded cDNAs were used for **RT-PCR** (Reverse Transcriptase-Polymerase Chain Reaction) for further gene expression studies. This determines whether the gene of interest is expressed or not, and the level of its transcription. Gene specific primers were used to amplify *axe*10 and *xp*12. In reactions with reverse transcriptase (R+), *axe*10 and *xp*12 were expressed. This corresponded to the sizes of *axe*10 at 800bp (**Figure 4.2**) and *xp*12 at 993bp (**Figure 4.3**) (**Figure 4.11**).

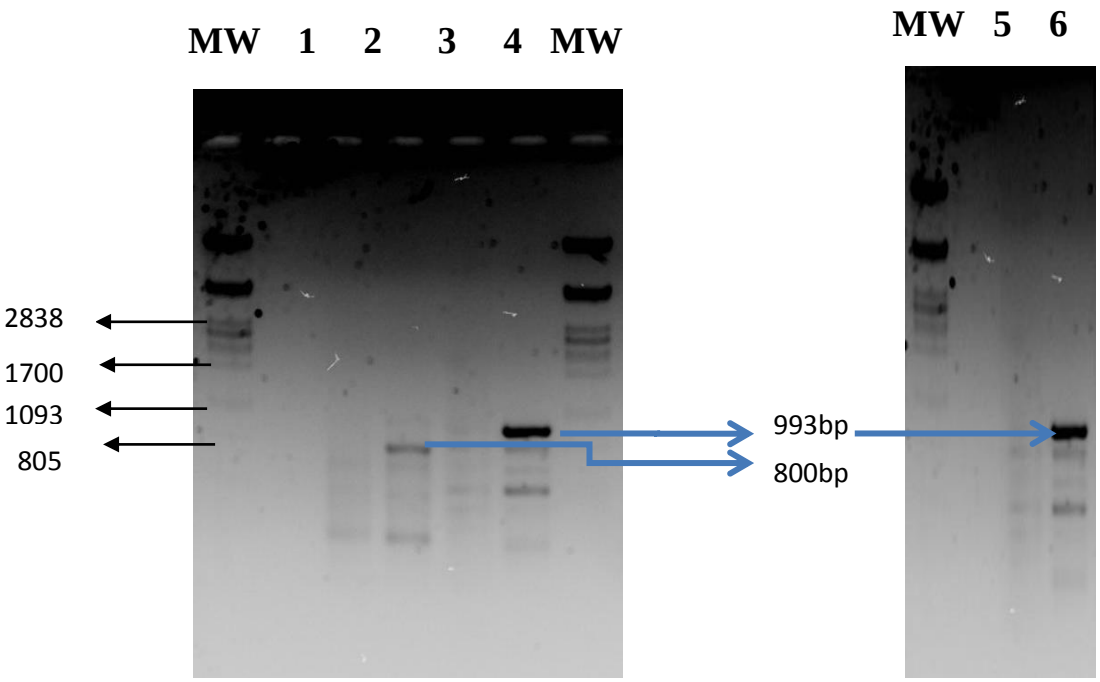


Figure 4.11: Amplification of the *axe*10 and *xp*12 genes by RT-PCR. MW-λ-*Pst*1 DNA molecular weight marker, lane 1-(R-) for pTip-AXE, lane 2- (R+) for pTip-AXE, lane 3-(R-) for pTip-XP12; 4-(R+) for pTip-XP12, lane 5-(R-) for pFos-XP_12, and lane 6 -(R+) for pFos-XP_12.

After confirming transcription using *R. erythropolis*, the level of translation was analysed. The constructs (pTip-AXE10, pTip-XP12, pET-AXE10, pET-XP12, pFos-Cell_11, and pFos-XP_12) that are listed in **Table 4.1,** were used as templates for cell-free protein synthesis for specific aspects of heterologous expression. The Bradford assay was used to measure the amount of protein produced during CFPS (**Table 4.3**). As seen from the table below, for all the constructs selected, proteins were produced using *R. erythropolis* and *E. coli* cell-free extract.

Table 4 .3: Amounts of protein produced during CFPS were measured using Bradford assay.

Sample	***E. coli* in mg/ml**			**Average**	**STD**		***R. erythropolis* in mg/ml**			**Average**	**STD**
Cell-free extract	1.5	1.4	1.3	**1.4**	**0.1**	Cell-free extract	1.1	0.9	1.11	**1.0**	**0.12**
pET21a	1.6	1.5	1.7	**1.6**	**0.1**	pTip-RC1	1.4	1.3	1.5	**1.4**	**0.1**
pET-AXE10	2.2	2.3	2.1	**2.2**	**0.1**	pTip-AXE10	2.8	2.9	2.6	**2.77**	**0.15**
pET-XP12	1.75	1.9	1.6	**1.75**	**0.15**	pTip-XP12	1.5	1.68	1.6	**1.59**	**0.07**
pFos-Cell_11	2.10	2.8	2.9	**2.6**	**0.44**	pFos-Cell_11	2.8	2.7	2.5	**2.67**	**0.15**
pFos-XP_12	2.7	2.5	2.9	**2.7**	**0.2**	pFos-XP_12	2.5	2.3	2.4	**2.4**	**0.1**

4.2.2 CFPS of Acetyl Xylan Esterase (AXE10)

pTip-AXE10 was used as a template for CFPS using *R. erythropolis* cell-free extract, and pET-AXE10 was used with *E. coli* BL21 (DE3) cell-free extract for comparison. The products produced by both systems were examined using SDS-PAGE (**Figure 4.12**), zymogram-based assay and Fast Garnet liquid assay (**Figure 4.13)** (Koseki *et al*., 2005). The α-naphthyl acetate was used as the assay substrate in the zymogram, which is cleaved by the Fast Garnet, to release α-naphthol. The 29 kDa AXE protein stained red (**Figure 4.13**), corresponding to the over-expressed 29kDa band on SDS-PAGE gel run under the same conditions. The two control lanes (*R. erythropolis* extract only and *R. erythropolis* extract + pTip-RC1) did not show over-expressed band, showing that there was no background acetyl xylan esterase activity.

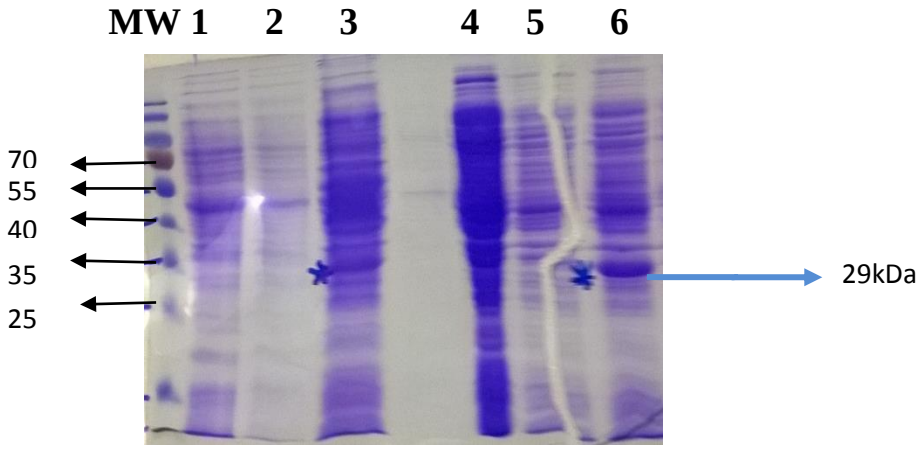


Figure 4.12: SDS-PAGE analysis of cell-free protein synthesis using *R. erythropolis* and *E. coli* BL21 (DE3) cell-free extracts. MW- Pre-stained PAGE ladder, lane 1- *R. erythropolis* cell-free extract only, lane 2- pTip-RC1 + *R. erythropolis* cell-free extract, lane 3- *R. erythropolis* cell-free extract + pTip-AXE10, lane 4- *E. coli* BL21(DE3) cell free extract only, lane 5- *E. coli* BL21(DE3) cell-free extract +pET21a(+), and lane 6- *E. coli* BL21(DE3) cell-free extract + pET-AXE10. The blue arrow indicates AXE protein band at 29kDa.

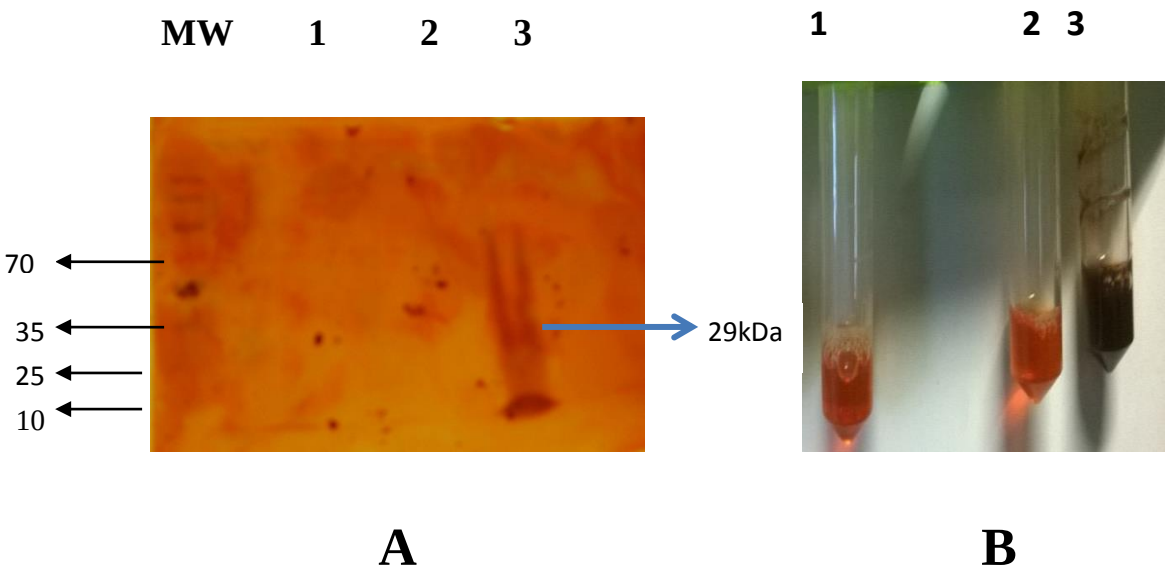


Figure 4.13: Zymogram analysis of CFPS using pTip-AXE10 as a template, showing enzyme activity after staining with Fast Garnet (A). MW- Pre-stained protein ladder, lane 1- *R. erythropolis* extract only, lane 2- *R. erythropolis* extract + pTip-RC1 and lane 3- R. *erythropolis* extract + pTip-AXE10. The size of AXE10 is indicated by the blue arrow at 29kDa. Liquid assay was also performed to measure AXE enzyme activity (B); tube 1 -*R. erythropolis* extract only, tube 2 -*R. erythropolis* extract + pTip-RC1 and tube 3- *R. erythropolis* extract + pTip-AXE10.

A qualitative liquid assay was used to measure AXE activity based on the amount of α-naphthol produced (**Figure 4.13, B**). OD_{560}nm readings were recorded for; tube 1 -0.125, tube 2 -0.33 and tube 3 -0.576. The AXE activity was reported as the amount of naphthol released based on α-naphthol standards. The activity was 0.9mg/ml.

The CFPS product, when pTip-AXE10 was used as a template, was His-Tag purified using the Ni-chelating His-Bind resin (**Section 3.7.4**) and run on a 12% SDS-PAGE gel. The

analysis showed a single band on the eluted fraction at 29kDa, which correlates to the expected size of the AXE10 (**Figure 4.14**). This confirmed that His-tagged AXE10 was expressed using the *R. erythropolis* CFPS.

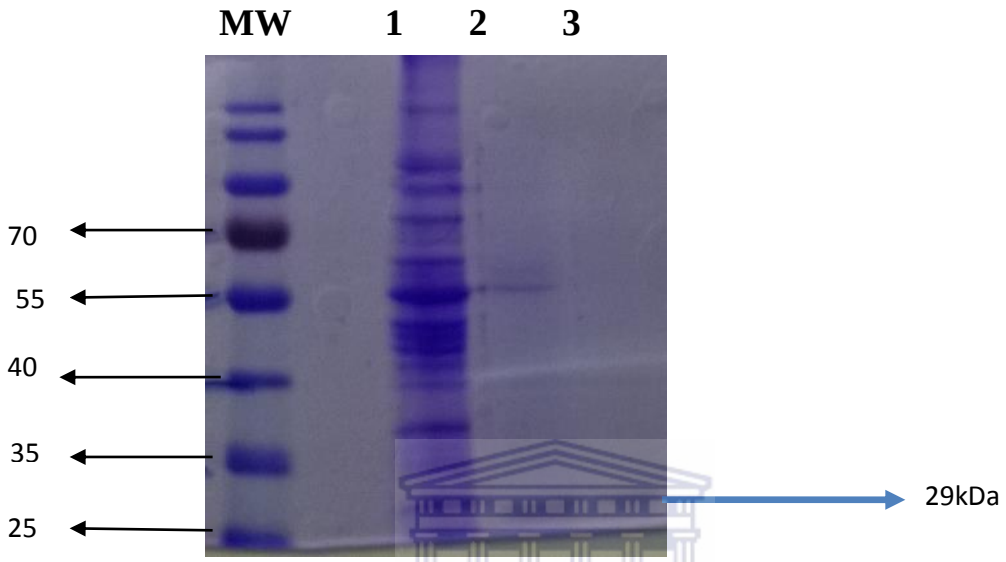


Figure 4.14: The purification of AXE10 from the CFPS reaction, when pTip_AXE10 was used as a template using a Ni-chelating His-Bind resin column. MW, Pre-stained PageRuler protein ladder (Fermentas, product # 26616), lane 1- flow –through fraction, lane 2- wash buffer flow through fraction and lane 3- eluted flow through fraction. The 29kDa AXE10 protein band is indicated by the blue arrow.

4.2.3 CFPS of cellulases XP12 and Cell_11

A cellulase plate based assay was first performed to confirm cellulase activity for the constructed plasmid and fosmid clones. *R. erythropolis* was also checked on CMC media to verify that there was no cellulase activity present. This was confirmed by the absence of hydrolysis around the colonies. *S. coelicolor* strains are known to produce cellulase and it

was used as positive controls. The pTip-XP12, pET-XP12, and pFos-Cell_11 constructs produced a clearing zone around the colonies (**Figure 4.15**).

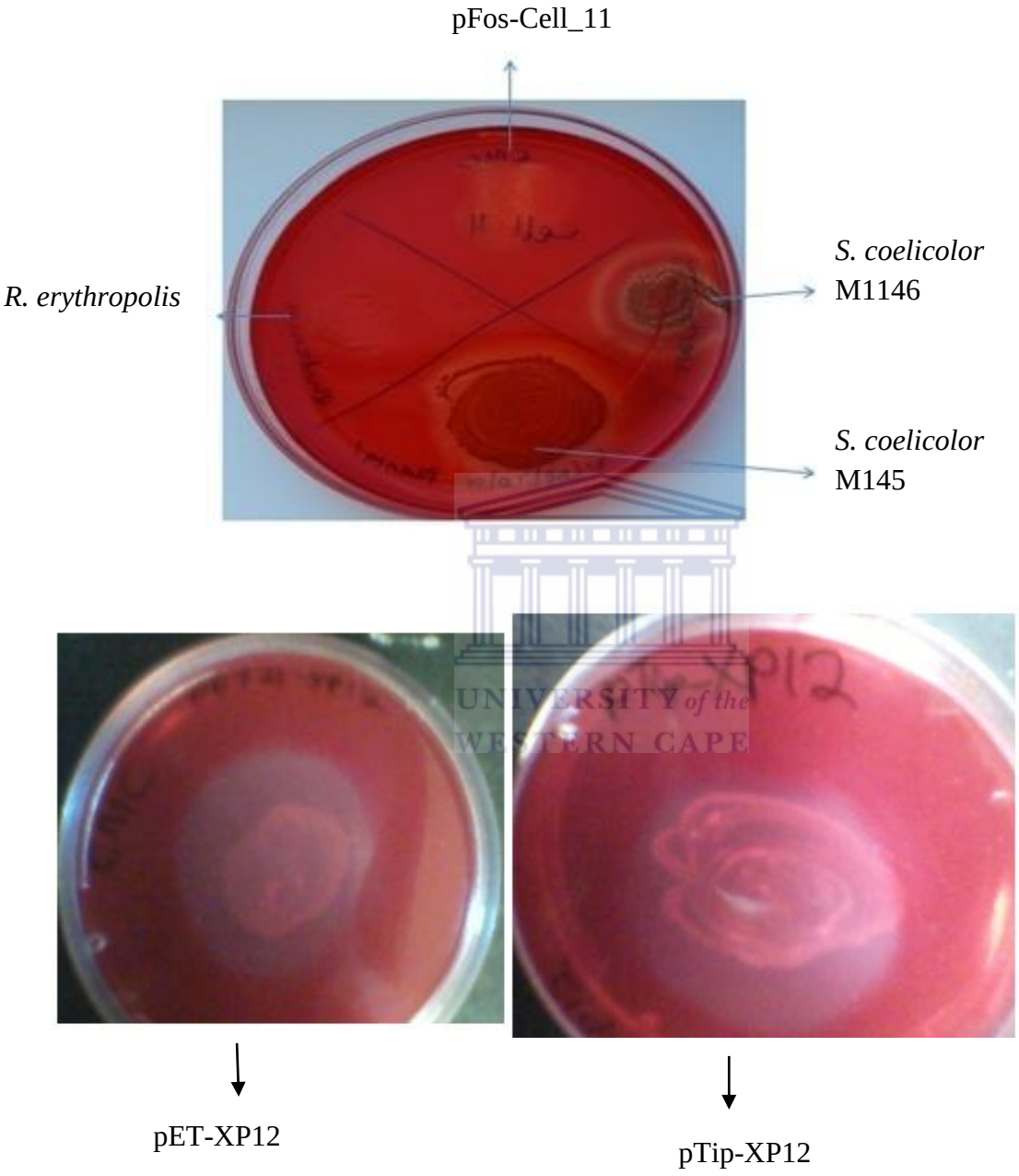


Figure 4.15: Cellulase producing plasmids and fosmid clones on CMC LB agar. The clearing zone around the colonies shows the hydrolysis of CMC. No zone of clearing was detected for *R. erythropolis*.

4.2.3.1 XP12

Expression of the XP12 protein using *R. erythropolis* CFPS was analysed. A protein band at approximately 38kDa was present when using both constructs (pTip-XP12 and pFos-XP_12) as templates (**Figure 4.16**). This shows that XP12 was expressed using *R. erythropolis* machinery from the XP12 native promoter and also by PtipA promoter, which resulted in correctly folded protein. The XP12 produced when pTip-XP12 was used as a template, was His-Tag purified as shown in **Figure 4.17**. As judged by the zymogram, XP12 was active when tested for cellulase activity.

DNS assays were performed to confirm the cellulase activity produced by XP12 after being synthesized through *R. erythropolis* CFPS. The enzyme activity was measured by the amount of reducing sugar released when CMC was used as a substrate (**Appendix C**), the standard curve was generated. Low concentration of XP12 was observed; 0.1mg/ml from pTip-XP12 and 0.3mg/ml from pFOS-XP_12.

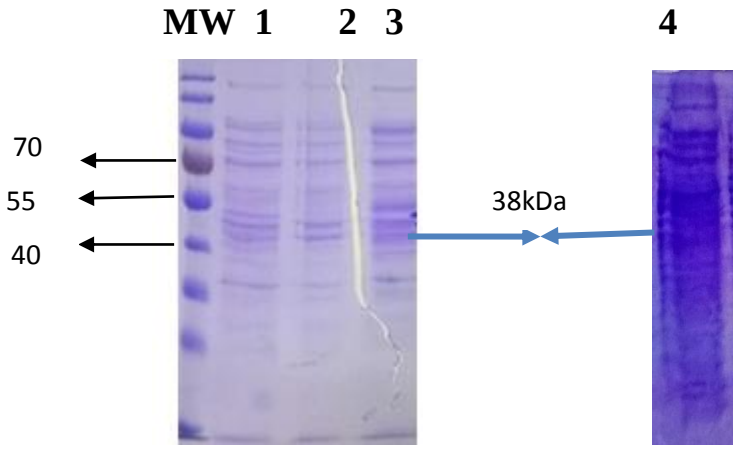
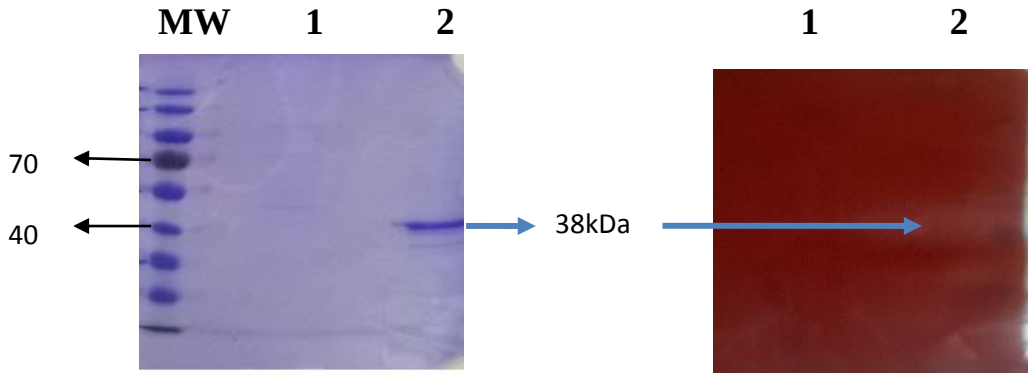


Figure 4.16: SDS-PAGE analysis of cell-free protein synthesis using *R. erythropolis* cell-free extract. MW-Pre stained PAGE ladder, lane 1-*R. erythropolis* cell-free extract only, lane 2-pTip-RC1 + *R. erythropolis* cell-free extract, lane 3- *R. erythropolis* cell-free extract + pTip-XP12, and lane 4- *R. erythropolis* cell-free extract _ pFos-XP_12. The blue arrow indicates XP12 protein band at 38kDa.



Stained with Coomassie blue Stained with Congo red

Figure 4.17: The purification of XP12 from the CFPS reaction, when pTip_XP12 was used as a template using a Ni-chelating His-Bind resin column. MW- Pre-stained PageRuler protein ladder (Fermentas, product # 26616), lane 1- flow –through fraction, and lane 2- eluted flow-through fraction. The gels were stained with Coomassie Blue (left) and Congo red (right). The 38kDa XP12 protein band is indicated by the blue arrow.

4.2.3.2 Cell_11

A recombinant fosmid, pFos-Cell_11 was used as template for CFPS for *R. erythropolis* and *E. coli* BL21 (DE3) cell-free systems. Both CFPS systems yielded Cell_11 (**Figure 4.18**). A zone of clearing was observed at 127kDa on the zymogram, showing that the Cell_11 produced was active.

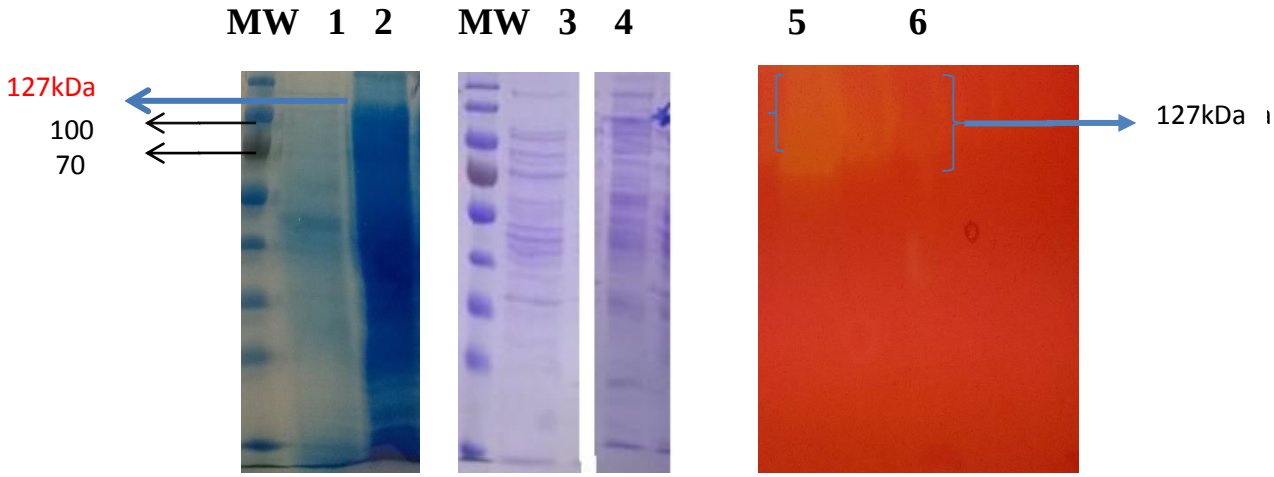


Figure 4.18: SDS-PAGE analysis of CFPS using pFos-Cell_11 with *R. erythropolis* and *E. coli* cell-free extracts. MW- Pre-stained PageRuler protein ladder (Fermentas, product # 26616), lane 1- *E. coli* BL21 (DE3) cell-free extract only, lanes 2 & 6- *E. coli* BL21 (DE3) cell-free extract + pFos-Cell_11, lane 3- *R. erythropolis* extract only, and lanes 4 & 5 *R. erythropolis* cell-free extract + pFos-Cell_11. The gels were stained with Coomassie Blue (left/blue) and Congo red (right/red). The 127kDa Cell_11 protein band is indicated by the blue arrow.

DNS assay was also performed to confirm the cellulase activity produced by Cell_11 synthesized through CFPS. A visible colour change was observed from Cell_11 with a concentration of 1.0mg/ml generated from the CMC standard curve.

Mass Spectrometry was used to confirm that the protein yielding cellulase activity on the zymogram was indeed Cell_11. The band correlating to the Cell_11 protein was sequenced; the nucleotide and translated amino acid sequence are shown in **Appendix A**. A total of 23 peptides identified by the mass spectrometry matched to the Cell_11 protein (**Appendix B**).

4.3 Wider application of the *R. erythropolis* CFPS system in metagenomics screening

In vitro metagenomics have been suggested as a solution to overcome some problems associated with heterologous expression including protein folding, codon usage and promoters (Ferrer *et al*., 2007). *R. erythropolis* or *E. coli* BL21 (DE3) cell-free protein synthesis was combined with *in vitro* compartmentalisation (IVC) and fluorescent activated cell sorter (FACS) to screen for genes encoding active β-xylosidase from uncloned environmental DNA. This was performed by Mr. Nevondo, PhD student (2014). IVC-FACS is ultra-high throughput metagenomic screening platform which does not require cloning (Scanlon *et al*., 2014). The methodology involves preparing double emulsion using *R. erythropolis* or *E. coli* cell free extract to provide transcription-translation machinery for protein synthesis. The positive activity within a double emulsion is dependent on a fluorescent substrate, 4-methylumbelliferyl-β-D xylopyranoside (MUX), which releases methylumbeliferyl. The product is thought to contain the protein and the gene of interest. The events were gated far from background activity. The results showed that there was no significant difference in fluorescence intensity between the experiment (mDNA and substrate added) and control (no mDNA and no substrate) when using *E. coli* BL21 (DE3) IVC extract. With the *R. erythropolis I*VC extract, a 3 fold increase in fluorescence between the experiment (mDNA and substrate added) and controls (no mDNA and no substrate) was observed. The genes (11 genes) encoding β-xylosidase activity were sorted from the uncloned metagenomic DNA. The β-xylosidases identified were amplified in such a way as to include their natural promoter sequences. All the 11 genes showed similarity with *Elizabethkingia anopheles* (Nevondo, 2014), this was not unexpected since metagenomic DNA was extracted

from a from compost farm which is an obvious mosquito habitat, making it highly likely to find these bacteria (Ngwa *et al*., 2013; Kukutla *et al*., 2014).

4.3.1 CFPS using β-xylosidase gene

To answer some of the questions on whether the identified genes do not express well in *E. coli* due to protein folding or if the native promoter not recognized by *E. coli*, one of the genes, *W*6 was used as a template for CFPS. An over-expressed protein corresponding to the 61kDa predicted size from the protein sequence (**Appendix D**) was observed when assessed using SDS-PAGE (**Figure 4.19**).

To check whether protein will be expressed in high amount, *w*6 was cloned into pTip-RC1 and pET21. An amount of 0.5µg of pTip-W6 and pET-W6 DNA was used for CFPS in both *E. coli* and *R. erythropolis* cell-free extracts to test the strength of PtipA and T7 promoter, if the protein can fold correctly in both bacteria. β-xylosidase activity using *ρ*NPX as a substrate was carried out (**Figure 4.20**). OD_{410}nm was measured and the data was recorded in **Table 4.4**.

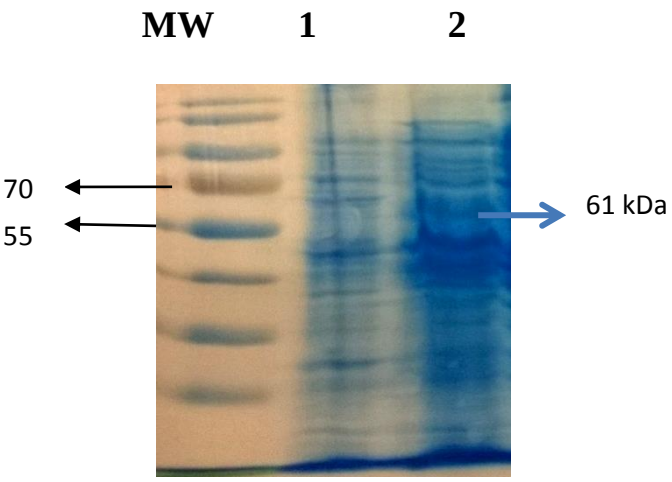


Figure 4.19: SDS-PAGE analysis of cell-free protein synthesis using *R. erythropolis* cell-free extract. MW- Pre-stained PageRuler protein ladder (Fermentas, product # 26616), 1- *R. erythropolis* cell-free extract only, lane 2- *R. erythropolis* cell-free extract + *w*6. The 61kDa W6 protein band is indicated by the blue arrow.

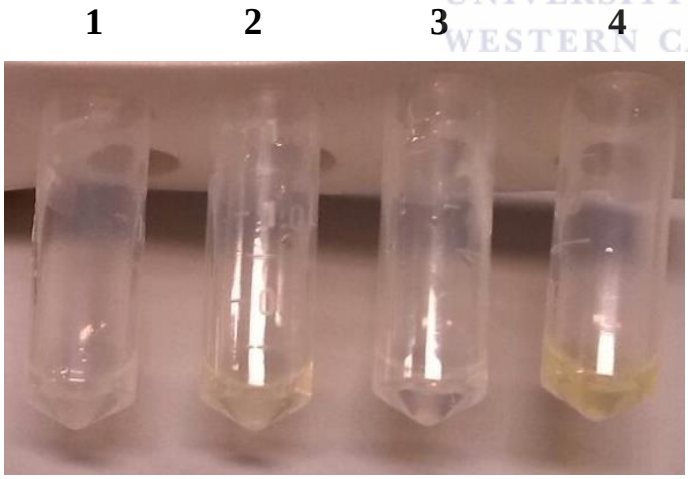


Figure 4.20: β-xylosidase liquid assay using CFPS products when pTip-W6 and pET-W6 were used evaluate if proteins folds correctly. Tube 1- pET-W6 used in *R. erythropolis* cell-free extract, tube 2- pET-W6 used in *E. coli* cell-free extract, tube 3- pTip-W6 used in *E. coli* cell free extract, and tube 4- pTip-W6 used in *R. erythropolis* cell-free extract.

Table 4.4: Evaluating whether the proteins produced are active after *E. coli* and *R. erythropolis* cell-free expression.

Plasmid clones	*E. coli* (OD_{410}nm)	*R. erythropolis* (OD_{410}nm)
pET-W6	0.450	0.230
pTip-W6	0.145	0.510

β-xylosidase liquid assay using CFPS products when pTip-W6 and pET-W6 were used showed that an active protein was produced by both *R. erythropolis* and *E. coli* CFPS systems (**Table 4.4**). These results proved that the genes identified by the IVC-FACS did not express well in *E. coli* because native promoter was not recognized by *E. coli* and there is no problem with protein folding. The capability of the recovered β-xylosidase from the IVC-FACS to be expressed in the *R. erythropolis* CFPS system but not in the *E. coli* validates use of this system for functional screening of metagenomic DNA to recover genes which normally would not be identified in *E. coli*-based screening.

CHAPTER 5:
General Discussion and Conclusion

Chapter 5 General discussion and conclusion

Classic metagenomic gene discovery using a functional screening approach involves isolation of DNA from the environment, cloning the DNA into the suitable vector, transforming into a host organism followed by the screening of the transformants (Takai *et al*., 2000; Gilbert and Dupont, 2011). The scope and application of metagenomics is limited by numerous challenges, most of which are associated with heterologous expression of recombinant DNA and low throughput screening methods that have relied on the construction of DNA libraries. The heterologous host most often employed is *E. coli*; however, *in silico* studies estimated that *E. coli* can only support the expression of 40% of environmental genes (Gabor *et al*., 2004). Other problems associated with heterologous protein expression include: host toxicity of expressed products, microbial codon preferences because the frequency of the codon usage in an organism is directly related to the corresponding tRNAs which can cause foreign genes with rare codons to likely not be expressed or have defective protein folding due to the expression being limited to the available aminoacyl-tRNA (Chen and Inouye, 1994; Stoletzki and Eyre-Walker, 2007; Angov, 2011). These problems lead to the majority of metagenomic DNA being untapped. Attempts to overcome problems associated with heterologous expression includes: heterologous complementation by mutants or host stains (Simon and Daniel, 2009), and development of compatible shuttle vectors to enable screening to be conducted in parallel with multiple hosts (Troeschel *et al*., 2012; Taupp *et al*., 2011). Given some of the reasons for failed heterologous expression of environmental DNA mentioned above, there is a need to develop alternative systems or methods to access a wider functional diversity. Ferrer *et al*. (2007) suggested *in vitro* metagenomics to be a solution for overcoming the heterologous expression, regarding issues with toxicity. Cell-free protein synthesis also known as *in vitro* protein synthesis systems are capable of expressing

environmental DNA to produce functional proteins *in vitro* (Carlson *et al.*, 2012). Currently, the cell-free extracts that are available commercially include; prokaryotic (*E. coli*) and eukaryotic [(wheat germ, rabbit reticulocytes, *Spodoptera frugiperda* (insect)] cells (Sierecki *et al*., 2013).

They enable unparalleled access to the protein production environment as compared to the *in vivo* systems because it allows for control of energy sources, salt concentrations, tRNAs, mRNA and many other factors (Carlson *et al.*, 2012). The fact that metabolism and cell reproduction are eliminated, allows the manipulation of cell metabolic systems to focus on the targeted protein or set of proteins, instead of survival (Ranji *et al*., 2013). When comparing CFPS to cell-based gene expression, it has advantages when it comes to the speed and flexibility for the range of proteins produced because the reaction time can be less than 1hour (Kwon *et al*., 2010). The other advantages of using CFPS include the ability to express and study products which are toxic to a host cell, such as membrane proteins (Li *et al*., 2014) and the direct access to antimicrobial metabolites. *E. coli* based cell-free extracts have been used due to their well-known machinery and ease of use (Carlson *et al*., 2012). However, due to problems associated with misfolded or partially folded proteins (Rosano and Ceccarelli, 2009; Correa and Oppezzo, 2015) that are either free in the cytoplasm or still bound to the ribosome (Li *et al*., 2014), there is a need to develop a cell-free system of another organism other than *E. coli*.

In this study an actinobacterial crude extract was prepared and used as an *in vitro* transcription-translation system, which was later coupled with IVC and flow cytometry to screen for β-xylosidases. An actinobacterial source, *Rhodococcus erythropolis*, was selected

because they might recognize promoters other than *E. coli* and fold proteins differently (Kayser *et al*., 1996), particularly when working with environmental samples and with a focus on lignocellulases. The crude extracts prepared were easy and cheap (using bioreactor for growth and sonicating to reduce cost) to produce.

To investigate the ability of the *R. erythropolis in vitro* transcription-translation system prepared to produce active protein, three genes were selected for synthesis: cellulase (*xp*12 and *cell*_11) and acetyl xylan esterase (*axe*10). All three were previously isolated from metagenomic libraries prepared from thermophilic material (compost or hot spring sediment) and screened in *E. coli*. All three enzymes could be expressed from the native promoters associated with these genes through the *R. erythropolis* CFPS as judged by enzyme activity and zymograms. The proteins that were difficult to express using *E. coli* cell-free system, were produced using the *R. erythropolis* CFPS system. These include XP12, which was produced in low amounts in *E. coli* (0.8 mg/ml) as compared to *R. erythropolis* (1.9 mg/ml) (**Table 4.3**), which could mean that the native promoter does not work well in *E. coli* as compared to *R. erythropolis*.

The essence of CFPS is to produce functional protein in a test tube, rather than in cells (Rosenblum and Cooperman, 2014). However, if no protein is produced, optimization of the CFPS is required. *R. erythropolis* CFPS system was designed in a way that would not require optimization because polyethylene glycol was added to avoid aggregated proteins to form inclusion bodies (Kai *et al*., 2013). The proteins produced did not inhibit transcription, thus there was no need to uncouple transcription from translation; whereby synthetic mRNA would be generated in the presence of modified nucleotides to allow one to study post-

transcriptional modification and use large quantities of mRNA for translation (Rosenblum and Cooperman, 2014). Although the general temperature for bacterial CFPS is 37˚C [(*E. coli* is the only available bacterial cell-free expression system, (Zawada, 2012)], it has been suggested that lowering the temperature to 30˚C might overcome the problem that a large fraction of protein produced is inactive (Iskakova *et al*., 2006; Rosenblum and Cooperman, 2014). In the case of the *R. erythropolis* CFPS system, there was no need to alter the temperature, because proteins are produced optimally at 30˚C.

Due to the high engineering flexilbility, CFPS can be applied with high-throughput technologies (Li *et al*., 2014). To demonstrate the applicability of the *R. erythropolis* cell-free extract, it was used in an IVC/FACS-based metagenomic screen, for β-xylosidases. The recovery of ORF's encoding β-xylosidases, the products which seem to function when placed behind their native promoter in *R. erythropolis in vitro* and not in *E. coli,* validates the system for use in functional metagenomic screening. In particular, to recover genes that would not be identified using *E. coli* based screening. These encouraging results indicate that a *R. erythropolis* extract or that of many microorganisms could be used to tap into a larger portion of the metagenomic sequence space than what is allowed using classic screening techniques. This will contribute to the identification and development of new biocatalysts as it combines un-cloned (less biased) metagenomic DNA and high-throughput technology (FACS), which should allow more access to functional diversity through increasing gene expression repertoire available. Other benefits of using a CFPS system are as follows. i) Screening can be performed at different temperatures without affecting cellular metabolisms *e.g.* under thermophillic conditions. Although, many thermophillic proteins have been functionally expressed in *E. coli*, some proteins are less stable and thermophillic than the native enzyme *e.g Sulfolobus solfataricus* 5′-methylthioadenosine phosphorylase in this

system (Cacciapuoti *et al*., 1999). Cell-free extracts from hyperthermophiles can be produced to allow folding of thermophillic enzymes under native conditions. ii) Classic metagenomic screening for genes of interest is time consuming when compared with the IVC-FACS which allows screening for up to 20 000 events per second. This shows that this technology is more efficient than classical metagenomics (Ma *et al*., 2014). iii) The insert size limitation with fosmid libraries results in the inability to identify large pathways or operons easily. The IVC-FACS system can be used to fully screen uncloned mDNA which will increase the probability of identifying such targets.

The complete hydrolysis of lignocellulosic biomass into simple fermentable sugars requires hydrolysis of hemicellulose and separation of lignin. Lignin is often unused or burned into energy (Ragauskas *et al*., 2014). However, insufficient separation of cellulose and lignin could occur, resulting in the formation of by-products that inhibit fermentation (e.g. acetic acid from hemicellulose, furans from sugars and phenolic compounds from the lignin fraction). The effectiveness of enzymatic processes on lignocellulose hydrolysis could be improved by the isolation of lignocellulases such as, laccases and peroxidases), and which could be incorporated in the enzyme cocktails used. Only a few bacterial ligninases have been reported, and this could be attributed to limitations associated with convectional screening methods employed. IVC-FACS could be employed to screen for ligninases and other lignocellulases for the development of improved enzyme cocktails.

Future work will include improving the *R. erythropolis* CFPS system, in a similar way in which the *E. coli* system has been developed, in order to produce high protein yield. A number of suggestions will follow. i) Developing components for translation specific for *R. erythropolis*: although all the other components are from cell-free extract (ribosomes, RNA polymerase, sigma factors, chaperons; Shin and Noireaux, 2012; Kwon and Jewett, 2015), some of the components for translation have been generated from *E. coli* (*e.g* tRNA synthetases). The quantity and fidelity of the CFPS is dependent upon the concentration of magnesium (Li *et al*., 2014), because high Mg concentration could cause the ribosomes to be stuck in the translocation step (Qin and Sarnow, 2004). An amount of 12mM magnesium was used for *R. erythropolis* cell-free system which could have contributed to the low protein production. Li *et al*. (2014) suggested that lowering the concentration to 4mM increased the amount of protein produced. ii) Even though harvesting cells at mid-log to late log-phase allows the cells to contain an efficient balance of the transcription-translation machinery to maintain cellular division, Smith *et al*. (2014) showed that by delaying harvest time, the yield after the supplemented components are added into the translation system will increase and thereby achieve more favorable ratios between the endogenous and the supplemented components. Cells harvested at higher optical density results in larger cell mass which results in increased quantity of crude extract prepared per fermentation (Kwon and Jewett, 2015). iii) The efficiency of the *R. erythropolis* CFPS, could be increased by using PCR products as templates instead of constructs to avoid cloning (Pedersen *et al*., 2011).

In conclusion, this study has contributed to the development of the first *in vitro* metagenomic screening platform based on actinobacterial cell-free protein synthesis and has opened doors for screening enzymes that would not be detected through screening in *E. coli* systems and for products which would be toxic if produced intracellularly. The development of this

system using existing *E. coli* protocols was straight-forward and it did not require any modification of the bacterium, which suggested that many other screening hosts could be suitable to use in this approach.

Appendices: Appendix

APPENDICES

Appendix A: Nucleotide sequence and the deduced amino acid sequence of cell_11. The ATG start codon is highlighted in red and the TAG stop codon in green. The two cellulose binding domains are highlighted in purple. The conserved GH family 9 catalytic domain is highlighted in grey. The GH 5 endoglucanase is double underlined. The three conserved CBM 3 family carbohydrate binding domains are highlighted in turquoise. A change of amino acid was observed in a peptide highlighted in blue. All other peptides identified are underlined.

1 ATGCTGAGACGACGGGCGCTGTCGATGTTGACGGGCGCGGCGGTCGTTTTTTCGGCTTTC

1 M L R R R A L S M L T G A A V V F S A F

61 GTGCCGGTTGGAAGTCCGGATTCGGCTGTTGTTCGGGCGGCTCCTACCTCTTACAATTAC

21 V P V G S P D S A V V R A A P T S Y N Y

121 GCGGAAGCGCTTCAAAAGGCAATTTATTTCTACGACGCACAACGTTCCGGCAAGCTGCCT

41 A E A L Q K A I Y F Y D A Q R S G K L P

181 CCCGATAACCGCGTCGAATGGCGCGGGGATTCCGGGCTGAACGACGGAGCCGACGTCGGC

61 P D N R V E W R G D S G L N D G A D V G

241 GTCGATTTGACGGGCGGTTGGTACGACGCCGGCGACCACGTCAAGTTCGGGTTGCCGATG

81 V D L T G G W Y D A G D H V K F G L P M

301 GCATATTCCGCCGCCATGCTGGCGTGGGCGGTGTACGAGTACCGCGACGCTTTCGTGCAG

101 A Y S A A M L A W A V Y E Y R D A F V Q

361 ACGGGGCAGCTCGATTATATTTTGAACAACATCAAGTGGGCGACCGACTATTTCATCAAG

121 T G Q L D Y I L N N I K W A T D Y F I K

421 GCGCATTCCGCGCCCAACGTGCTTTGGGGACAGGTCGGCAAGGGGGACGTCGATCATGCC

141 A H S A P N V L W G Q V G K G D V D H A

481 TGGTGGGGACCGGCAGAAGTGATGCAGATGCCGCGTCCGGCCTACAAAATCGACCCGAGC

161 W W G P A E V M Q M P R P A Y K I D P S

541 TGTCCGGGGTCCGATCTCGCTGCCGGAACGGCCGCCGCGATGGCCGCCGCCGCCGCGGTG

181 C P G S D L A A G T A A A M A A A A V

601 TTTAAACCCACTGACCCGTCTTATGCCTCAACGTTGATCGCTCATGCGAAACAATTGTAT

201 F K P T D P S Y A S T I A H A K Q L Y

661 ACGTTTGCAGATACTTATCGGGGGAAATATTCCGATTGCATCACTGACGCGCAAAATTTC

221 T F A D T Y R G K Y S D C I T D A Q N F

721 TATCGTTCGTGGAGCGGTTACGCCGATGAGCTGACGTGGGGCGCCGTCTGGCTTTATCTC

241 Y R S W S G Y A D E L T W G A V W L Y L

781 GCCACCGGCGAGCAGGCCTATCTTGACAAGGCAATCGCCTCAGTCGCGGAATGGGGGCGC

261 A T G E Q A Y L D K A I A S V A E W G R

841 GAAGGTCAGACACCTTATTGGGGTTACAAATGGACGCAAAGCTGGGACGACGTCCATTAC

281 E G Q T P Y W G Y K W T Q S W D D V H Y

901 GGCGCTCAGCTGCTGTTGGCAAGAATTACGGGCGACCAGCGGTTTATCCAGTCGACGGAG

301 G A Q L L L A R I T G D Q R F I Q S T E

961 CGCAACCTGGAATATTGGACGGACGGCACGGACGACACCGGCGAGCGCATCACGTATACG

321 R N L E Y W T D G T D D T G E R I T Y T

1021 CCCGGGGGGCTTGCTTGGCTAGATTCTTGGGGTTCACTCCGCTATGCGATGAACGCGTCG

341 P G G L A W L D S W G S L R Y A M N A S

1081 TTCTTGGCGTTCGTCTATTCCGACTGGCTGCAAAGCCGCGATCCCGCCAAAGCGGAAAAG

361 F L A F V Y S D W L Q S R D P A K A E K

1141 TACAGGAACTTCGCCGTTCGCCAGGTTCTGTATGCATTGGGCGACAACCCGCGCAATTCT

381 Y R N F A V R Q V L Y A L G D N P R N S

1201 AGCTATGTCGTCGGATTCGGGCGCAATCCGCCGCAGCGGCCGCATCACCGGACGGCGCAC

401 S Y V V G F G R N P P Q R P H H R T A H

1261 GGGTCGTGGGCTGACAGCCAGAACGTTCCCGCTTACCATCGGCACATTTTATATGGCGCT

421 G S W A D S Q N V P A Y H R H I L Y G A

1321 CTTGTGGGCGGTCCGAACCAGTCCGACGCCTATACGGATTCGATCAGCGACTACGTCGGC

441 L V G G P N Q S D A Y T D S I S D Y V G

1381 AATGAGGTTGCGACGGATTACAACGCGGCGTTTACGGGAAATCTCGCAAAAATGTATCTG

461 N E V A T D Y N A A F T G N L A K M Y L

1441 CTGTTCGGCGCCAGCGCCGGACAGCGGCCGCTTGCCAATTTTCCCGAGCCGGAAGTACGC

481 L F G A S A G Q R P L A N F P E P E V R

1501 GAGGACGAGTTTTTCGTCGAAGCCGGCGTGAACAGCTCCGGACCGAACTACACAGAGATC

501 E D E F F V E A G V N S S G P N Y T E I

1561 AAGGCCCTGATCAACAACCGGTCCGGCTGGCCCGCGCGGATGGGCGACAAGCTTTCGTTC

521 K A L I N N R S G W P A R M G D K L S F

1621 AAGTATTTCGTCGATTTGTCTGAAGTTTACGCCGCCGGTTATACGGTCAACGACATTAAG

541 K Y F V D L S E V Y A A G Y T V N D I K

1681 GTGACGACGAACTACAACGAAGGCGCGAAAGTATCCGGTCTGCTTCCGTATGACGAAAGC

561 V T T N Y N E G A K V S G L L P Y D E S

1741 CGTCGTCTTTACTATGTGCTCGTCGATTTTACGGGTACGAAGATTTATCCCGGCGGTCAG

581 R R L Y Y V L V D F T G T K I Y P G G Q

1801 TCCGCCTACAAGAAAGAGGTTCAGTTCAGACTGAGCGCTCCGAGCGGGACATCGTTCTGG

601 S A Y K K E V Q F R L S A P S G T S F W

1861 GATCCGAACAACGATTTCTCGTACCAGTTGATGTCCGGCACGTCCAACAGCAGCCTGGTC

621 D P N N D F S Y Q L M S G T S N S S L V

1921 AAGACGCCGTATATGCCGGTTTATGACGCCGGGGTGAAAATTTTCGGCGTGGAGCCGTCG

641 K T P Y M P V Y D A G V K I F G V E P S

1981 TCCGGAAGCGGGTCGAGCCCGACGCCGCCACCCACGTCGACGCCGACACCGACACCGATG

661 S G S G S S P T P P P T S T P T P T P M

2041 CCGACACCTGCACCGACACCGACGCCCACACCGAGTGTGACACCAACCGTGACGCCAACG

681 P T P A P T P T P T P S V T P T V T P T

2101 CCGACACCGACGCCGACACCGACGCCGACGCCTAGCGCGAGCGGTACCCTGCGCGTCGAG

701 P T P T P T P T P T P S A S G T L R V E

2161 TATCGCGTGGGCGACACGAGCGCTACCGACAACCAGATGAAACCGCAGCTGCGCATCGTC

721 Y R V G D T S A T D N Q M K P Q L R I V

2221 AACACCGGCTCGCAAGCCGTGCCGCTGACCGAGCTGAAGGTGCGCTACTGGTACACGAAG

741 N T G S Q A V P L T E L K V R Y W Y T K

2281 AACTCGACGCAGGCCGAACAGTACTTCTGCGACTGGGCGCAAATCGGCTGCTCGAACATC

761 N S T Q A E Q Y F C D W A Q I G C S N I

2341 CGGGCGCAGTTCGTGTCGCTGGCGCAGCCGGTCAGCGGGGCGGACAGCTACATCGAGCTG

781 R A Q F V S L A Q P V S G A D S Y I E L

2401 AGCTTCACGGGCGGAAGCATTCCGGCGGGAGGCAACACGGGCGAGATACAGAACCGGATT

801 S F T G G S I P A G G N T G E I Q N R I

2461 CACTTCACGAACTGGATGAACTACAACGAAGCGGACGACTGGTCGTACAACGGGACGCAG

821 H F T N W M N Y N E A D D W S Y N G T Q

2521 ACGACGTGGGGGCCGTCGACGCGGATTACGCTTTATCGCAACGGAGTGCTGGTATGGGGG

841 T T W G P S T R I T L Y R N G V L V W G

2581 ACGGAGCCGGGCGGCGGATCATCGACGCCGACGCCGACGGTGACACCGACGCCGACACCG

861 T E P G G G S S T P T P T V T P T P T P

2641 ACGCCGACGCCGACACCTACGCCTACGCCGAGCGCGGCGCCCACACCGACGCCGACGGCC

881 T P T P T P T P T P S A A P T P T P T A

2701 GGCGGCAGCCTGGTCGTGCAGTATCGCGCGGCGGACACGAACGCGGGCGACAACCAGCTG

901 G G S L V V Q Y R A A D T N A G D N Q L

2761 AAGCCGCACTTTAGGATTGTGAACCGCGGGACGACGAGCGTGCCGCTGTCGGAGCTTTCG

921 K P H F R I V N R G T T S V P L S E L S

2821 ATCCGGTACTGGTACACGGTGGACGGGGACAAGCCGCAGGTGTTCAACTGCGACTGGGCG

941 I R Y W Y T V D G D K P Q V F N C D W A

2881 CAGGTGGGTTGTTCGAACGTGCGCGGCAGCTTTGTGAAGCTTTCGACGGGCCGGACGGGG

961 Q V G C S N V R G S F V K L S T G R T G

2941 GCGGACTACTACATCGAGATCACGTTCACGTCGGGCGCGGGCAGCTTGGCGGCTGGGGGA

981 A D Y Y I E I T F T S G A G S L A A G G

3001 AGCAGCGGGGACATTCAGGTGCGGATCAACAAGAACGACTGGACGAACTACAACGAGGCG

1001 S S G D I Q V R I N K N D W T N Y N E A

3061 AACGATTACTCGTATGATCCGACGAAGACGAGTTTTGCGGATTGGAACCGGGTGACGCTG

1021 N D Y S Y D P T K T S F A D W N R V T L

3121 TATCGGAATGGTCAGCTCATCTGGGGCGTCGAACCATAG

1041 Y R N G Q L I W G V E P *

Appendix B: Mass spectrometry data analysis for cell_11

Confidence Level	Sequence	Protein Descriptions	Modifications	DeltaScore	XCorr	m/z [Da]	MH+ [Da]	Delta Mass [Da]
High	VGDTSATDNQmKPQLR	Cell11	M11(Oxidation)	1	2.78	592.9569	1776.856	0
High	AADTNAGDNQLKPHFR	Cell11		1	2.96	585.6249	1754.86	0
High	TAHGSWADSQNVPAYHR	Cell11		1	3.46	474.9737	1896.873	0
High	NPPQRPHHR	Cell11		1	1.15	569.8073	1138.607	0
High	VSGLLPYDESRR	Cell11		1	2.33	464.5808	1391.728	0
High	TPYmPVYDAGVK	Cell11	M4(Oxidation)	1	3.08	678.8286	1356.65	0
High	AHSAPNVLWGQVGK	Cell11		1	3.07	732.3914	1463.775	0
High	TSFADWNR	Cell11		1	2.31	498.7306	996.4538	0
High	EGQTPYWGYK	Cell11		1	2.66	614.7859	1228.565	0
High	NSSYVVGFGR	Cell11		1	2.82	543.272	1085.537	0
High	AAPTSYNYAEALQK	Cell11		1	3.3	763.8778	1526.748	0
High	QVLYALGDNPR	Cell11		1	2.73	623.3323	1245.657	0
High	AIASVAEWGR	Cell11		1	2.95	530.2825	1059.558	0
High	AIYFYDAQR	Cell11		1	2.32	573.7823	1146.557	0
High	VSGLLPYDESR	Cell11		1	2.99	618.317	1235.627	0
High	IVNTGSQAVPLTELK	Cell11		1	4.77	785.4465	1569.886	0
High	QLYTFADTYR	Cell11		1	2.57	639.3113	1277.615	0
High	NLEYWTDGTDDTGER	Cell11		1	3.91	886.375	1771.743	0
High	SGWPARMGDKLSFK	Cell11		1	0.95	790.4003	1579.793	-0.01
High	NDWTNYNEANDYSYDPTK	Cell11		1	5.76	1105.451	2209.894	0
High	GTTSVPLSELSIR	Cell11		1	2.12	680.377	1359.747	0

High	WATDYFIK	Cell11		1	2.22	522.2636	1043.52	0
High	IYPGGqSAYKKEVqFR	Cell11	Q6(Deamidated); Q14(Deamidated)	1	0.93	936.9842	1872.961	0.01
High	NGQLIWGVEP	Cell11		1	1.17	556.7911	1112.575	0

1 2 3 4 5 6 7 8

Appendix C: DNS assay performed on glucose monohydrate to generate a standard curve. Tubes 1-5 glucose standards: tube 1-1.330mg/ml, tube 2-1.0mg/ml, tube 3-0.5mg/ml, tube 4-0.25mg/ml, tube 5-0.1mg/ml; tube 6, negative control (*R. erythropolis* extract); tube 7, pTip-XP12; and tube 8, pFos-cell_11.

Appendix D: W6 protein sequence to determine the protein size. The protein size as calculated by DNAMAN was found to be 61kDa.

>Protein sequence translated from Unknown(1-1593) in RF1

MLQEGYMLKTADFAGPQYQIKNKIAELAGEDGMQAFYKKYLENGITKKDIDALKS
WGFNSVRLPMHYNLYTLPIEKEEVKGKDTWLEEGFRMTDNLLKWCAENKMYLFLD
MHALPGGQGNDVNISDNDKSKPSLWESEENQRKSVALWKKLAERYKDSPWIGGYD
IINEPNYGFTGKNLNGCDEESNAPLRKFMVDVTKAIREVDQKHLIIIEGNCWGNNYK
GIFPLWDNNLVLSFHKYWNKNDQNSIKQMLEYRNQYNVPIWLGESGENSNVWFTE
AISLMENNNIGWAFWPMKKIDNIAGVANVKITPEYEKLLNYWKNGGEKPSKEFAYK
TMMQIADNYKFENTEVKRDVIDAMFRQIKSNEVLPYTSHTIPGRIFATEYDLGRIGAA
YYDKDAINYRIDTGEQVNWNSGDKMRNDGVDIYSNKDKISNGYYVGKIEDGEWLN
FTLKSVKSGKYTLEIRYANANSAGQLSVTNSKGQQIVQTELPSTGGDQIWKTITVKN
VNIAKGTDKIKLQFDKGSFNLNYIEFK*

APPENDIX E: Preparation of Bradford assay using [BSA] standards between 0-0.05mg/ml concentrations to find the concentration of the unknown protein sample.

[BSA](mg/ml)(final concentration)	**0**	**0.005**	**0.01**	**0.02**	**0.04**	**0.05**	**Protein sample (triplicates)**
BSA 5mg/ml (stock)	0µl	1µl	2 µl	4 µl	8µl	10µl	5µl
Extraction buffer	10µl	9µl	8 µl	6 µl	2µl	0µl	5µl
HCl (0.1M)	10µl	10µl	10 µl	10 µl	10µl	10µl	10µl
Distilled H_2O	80µl	80µl	80 µl	80 µl	80µl	80µl	80µl

Bio-Rad Reagent	900µl	900µl	900µl	900µl	900µl	900µl	900µl
Final volume	1ml	1ml	1ml	1ml	1ml	1ml	1ml

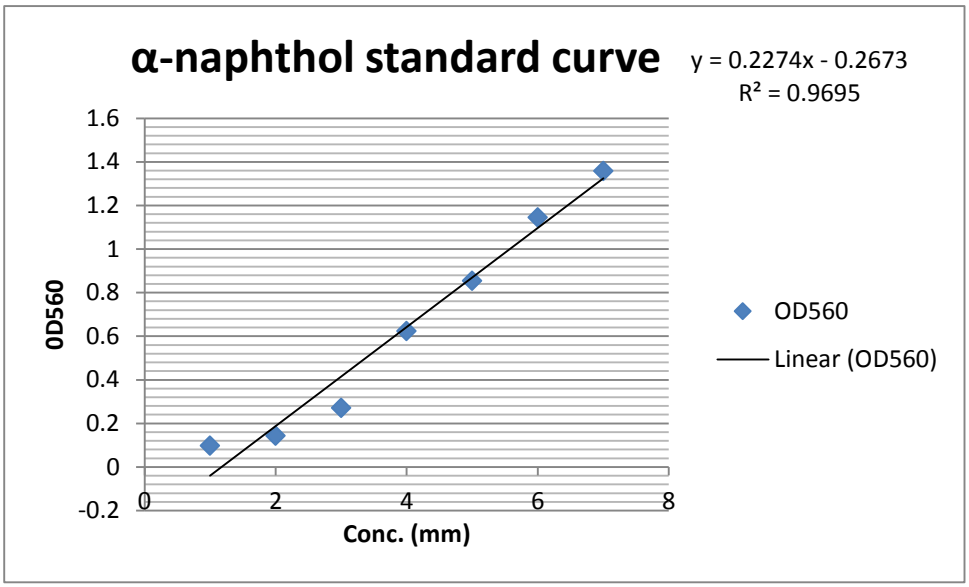
APPENDIX F: Preparation of 7%- 15% separating gel (A) and 4% stacking gel (B) for SDS-PAGE.

A

Separating gels (10ml)				
	7%	10%	12%	15%
Distilled H_2O	5.1 ml	4.1 ml	3.4 ml	2.4 ml
1.5M Tris-HCl, pH 8.8	2.5 ml	2.5 ml	2.5 ml	2.5 ml
20% (w/v) SDS	50 µl	50 µl	50 µl	50 µl
Bis-acrylamide (30%)	2.3 ml	3.3 ml	4.0 ml	5.0 ml
10% (w/v) ammonium persulfate	50 µl	50 µl	50 µl	50 µl
TEMED	10 µl	10 µl	10 µl	10 µl

B

4% Stacking gels (5.05 ml)	
Distilled H_2O	3.075 ml
1.5M Tris-HCl, pH 8.8	1.250 ml
20% (w/v) SDS	25 µl
Bis-acrylamide (30%)	670 µl
10% (w/v) ammonium persulfate	25 µl
TEMED	10 µl



APPENDIX G: A standard curve of α-naphthol using the concentration range of 0.25-5mm.

REFERENCES:
Literature Cited

References

Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology, 6*(6), 431-440.

Agresti, J. J., Kelly, B. T., Jaschke, A., & Griffiths, A. D. (2005). Selection of ribozymes that catalyse multiple-turnover diels-alder cycloadditions by using *in vitro* compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16170-16175.

Amore, A. (2013). Development of new biocatalysts to improve production of second generation bioethanol. P.H.D dissertation, Università di Napoli Federico II.

Angov, E. (2011). Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology Journal, 6*(6), 650-659.

Aranda, P. S., LaJoie, D. M., & Jorcyk, C. L. (2012). Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis*, *33*(2), 366-369.

Avenaud, P., Castroviejo, M., Claret, S., Rosenbaum, J., Mégraud, F., & Ménard, A. (2004). Expression and activity of the cytolethal distending toxin of *Helicobacter hepaticus*. *Biochemical and Biophysical Research Communications, 318*(3), 739-745.

Béguin, P., & Aubert, J. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews*, *13*(1), 25-58.

Béjà, O. (2004). To BAC or not to BAC: Marine ecogenomics. *Current Opinion in Biotechnology, 15*(3), 187-190.

Benizri, E., Baudoin, E., & Guckert, A. (2001). Root colonization by inoculated plant growth-promoting rhizobacteria. *Biocontrol Science and Technology, 11*(5), 557-574.

Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A. D., Magdassi, S., & Tawfik, D. S. (2004). *In vitro* compartmentalization by double emulsions: Sorting and gene enrichment by fluorescence activated cell sorting. *Analytical Biochemistry, 325*(1), 151-157.

Bertrand, H., Poly, F., Lombard, N., Nalin, R., Vogel, T. M., & Simonet, P. (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *Journal of Microbiological Methods, 6* (1), 1-11.

Bhattarai, K., Tiwari, K., & Agrawal, V. (2007). Enhanced antibacterial activity of sodium azide treated mutant streptomyces strain. *Journal Nepal medical association, 8* (1), 67-68.

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*(1), 248-254.

Brady, S. F., Chao, C. J., Handelsman, J., & Clardy, J. (2001). Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic Letters, 3*(13), 1981-1984.

Brandt, A., Erickson, J. K., Hallett, J. P., Murphy, R. J., Potthast, A., Ray, M. J., & Welton, T. (2012). Soaking of pine wood chips with ionic liquids for reduced energy input during grinding. *Green Chemistry, 14*(4), 1079-1085.

Bundy, B. C., & Swartz, J. R. (2010). Site-specific incorporation of *p*-propargyloxyphenylalanine in a cell-free environment for direct protein−protein click conjugation. *Bioconjugate chemistry*, *21*(2), 255-263.

Bundy, B. C., & Swartz, J. R. (2011). Efficient disulfide bond formation in virus-like particles. *Journal of Biotechnology*, *154*(4), 230-239.

Cacciapuoti, G., Fusco, S., Caiazzo, N., Zappia, V., & Porcelli, M. (1999). Heterologous Expression of 5′-Methylthioadenosine Phosphorylase from the Archaeon *Sulfolobus solfataricus*: Characterization of the Recombinant Protein and Involvement of Disulfide Bonds in Thermophilicity and Thermostability. *Protein expression and purification*, *16*(1), 125-135.

Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2012). Cell-free protein synthesis: Applications come of age. *Biotechnology Advances*, *30*(5), 1185-1194.

Cesarino, I., Araújo, P., Domingues Júnior, A. P., & Mazzafera, P. (2012). An overview of lignin metabolism and its effect on biomass recalcitrance. *Brazilian Journal of Botany*, *35*(4), 303-311.

Chen, G. T., & Inouye, M. (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes & Development*, *8*(21), 2641-2652.

Collins, T., Gerday, C., & Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. *FEMS microbiology reviews*, *29*(1), 3-23.

Correa, A., & Oppezzo, P. (2015). Overcoming the solubility problem in *E. coli*: Available approaches for recombinant protein production. *Insoluble proteins* (pp. 27-44).

Cottrell, M. T., Moore, J. A., & Kirchman, D. L. (1999). Chitinases from uncultured marine microorganisms. *Applied and Environmental Microbiology, 65*(6), 2553-2557.

Coutinho, P. M., Stam, M., Blanc, E., & Henrissat, B. (2003). Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends in Plant Science, 8*(12), 563-565.

Craig, J. W., Chang, F. Y., Kim, J. H., Obiajulu, S. C., & Brady, S. F. (2010). Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and Environmental Microbiology, 76*(5), 1633-1641.

Cybinski, D., Layton, I., Lowry, J., & Dalrymple, B. (1999). An acetylxylan esterase and a xylanase expressed from genes cloned from the ruminal fungus *Neocallimastix patriciarum* act synergistically to degrade acetylated xylans. *Applied Microbiology and Biotechnology, 52*(2), 221-225.

da Costa Lopes, André M, João, K. G., Rubik, D. F., Bogel-Łukasik, E., Duarte, L. C., Andreaus, J., & Bogel-Łukasik, R. (2013). Pre-treatment of lignocellulosic biomass using ionic liquids: Wheat straw fractionation. *Bioresource Technology, 142*, 198-208.

Degrassi, G., Okeke, B. C., Bruschi, C. V., & Venturi, V. (1998). Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus*. *Applied and environmental microbiology*, *64*(2), 789-792.

Degrassi, G., Kojic, M., Ljubijankic, G., & Venturi, V. (2000). The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. *Microbiology (Reading, England), 146* (7), 1585-1591.

Demain, A. L. (2009). Biosolutions to the energy problem. *Journal of Industrial Microbiology & Biotechnology*, *36*(3), 319-332.

Doherty, T. V., Mora-Pale, M., Foley, S. E., Linhardt, R. J., & Dordick, J. S. (2010). Ionic liquid solvent properties as predictors of lignocellulose pretreatment efficacy. *Green Chemistry*, *12*(11), 1967-1975.

Duan, C., and Feng, J. (2010). Mining metagenomes for novel cellulase genes. *Biotechnology letters*, 32, 1967-1975.

Edwards, B. S., Oprea, T., Prossnitz, E. R., & Sklar, L. A. (2004). Flow cytometry for high-throughput, high-content screening. *Current Opinion in Chemical Biology*, *8*(4), 392-398.

Faulds, C. B., Zanichelli, D., Crepin, V. F., Connerton, I. F., Juge, N., Bhat, M. K., & Waldron, K. W. (2003). Specificity of feruloyl esterases for water-extractable and water-unextractable feruloylated polysaccharides: Influence of xylanase. *Journal of Cereal Science*, *38*(3), 281-288.

Ferrer, M., Golyshina, O., Beloqui, A., & Golyshin, P. N. (2007). Mining enzymes from extreme environments. *Current opinion in microbiology*, *10*(3), 207-214.

Fernandez, M. G. S., Becraft, P. W., Yin, Y., & Lübberstedt, T. (2009). From dwarves to giants? plant height manipulation for biomass yield. *Trends in Plant Science*, *14*(8), 454-461.

Gabor, E. M., Alkema, W. B., & Janssen, D. B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, *6*(9), 879-886.

Gao, Y. (2012). Isolation, expression, purification and characterisation of novel acetyl xylan esterase from *Streptomyces* ORS10. Masters dissertation, University of the Western Cape.

Ge, Z., Schauer, D. B., & Fox, J. G. (2008). *In vivo* virulence properties of bacterial cytolethal-distending toxin. *Cellular Microbiology*, *10*(8), 1599-1607.

Ghatora, S. K., Chadha, B. S., Saini, H. S., Bhat, M. K., & Faulds, C. B. (2006). Diversity of plant cell wall esterases in thermophilic and thermotolerant fungi. *Journal of Biotechnology*, *125*(3), 434-445.

Gibson, W.R., & Hughes, S.R. (2009). Rated biorefineries with engineered microbes and high-value co-products for profitable biofuels production. *In vitro cell Development plant*, 45: 218-228.

Gilbert, J. A., & Dupont, C. L. (2011). Microbial metagenomics: Beyond the genome. *Annual Review of Marine Science*, *3*(5), 347-371.

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., & Handelsman, J. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and environmental microbiology*, *68*(9), 4301-4306.

Gomez-Escribano, J. P., & Bibb, M. J. (2011). Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters.*Microbial Biotechnology*, *4*(2), 207-215.

Gomez, L. D., Steele-King, C. G., & McQueen-Mason, S. J. (2008). Sustainable liquid biofuels from biomass: The writing's on the walls. *New Phytologist*, *178*(3), 473-485.

Graham-Rowe, D. (2011). Agriculture: Beyond food versus fuel. *Nature, 474*(7352), 66-83.

Gray, K. A., Zhao, L., & Emptage, M. (2006). Bioethanol. *Current Opinion in Chemical Biology, 10*(2), 141-146.

Griffiths, A. D., & Tawfik, D. S. (2000). Man-made enzymes—from design to *in vitro* compartmentalisation. *Current Opinion in Biotechnology, 11*(4), 338-353.

Guazzaroni, M., Silva-Rocha, R., & Ward, R. J. (2014). Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microbial Biotechnology, 8(1),* 52-64.

Hahn-Hägerdal, B., Galbe, M., Gorwa-Grauslund, M., Lidén, G., & Zacchi, G. (2006). Bio-ethanol–the fuel of tomorrow from the residues of today. *Trends in Biotechnology, 24*(12), 549-556.

Hain, T., Otten, S., von Both, U., Chatterjee, S. S., Technow, U., Billion, A., & Chakraborty, T. (2008). Novel bacterial artificial chromosome vector pUvBBAC for use in studies of the functional genomics of *Listeria* spp. *Applied and Environmental Microbiology*, *74*(6), 1892-1901.

Hamedo, H. A., & Makhlouf, A. H. (2013). Identification and characterization of actinomycetes for biological control of bacterial scab of *Streptomyces scabies* isolated from potato. *Journal of Biology, Agriculture and Healthcare, 3*(13), 142-153.

Hamelinck, C. N., Van Hooijdonk, G., & Faaij, A. P. (2005). Ethanol from lignocellulosic biomass: Techno-economic performance in short-, middle-and long-term. *Biomass and Bioenergy*, *28*(4), 384-410.

Handelsman, J. (2005). Sorting out metagenomes. *Nature Biotechnology, 23*(1), 38-39.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry & Biology, 5*(10), 245-249.

Hansen, MC., Tolker-Nielsen, T., Givskov M & Molin., S. (1998). Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. FEMS *Microbiol Ecology*, 26, 141–149.

Himmel, M. E., Xu, Q., Luo, Y., Ding, S., Lamed, R., & Bayer, E. A. (2010). Microbial enzyme systems for biomass conversion: Emerging paradigms. *Biofuels, 1*(2), 323-341.

Hong, S. H., Kwon, Y., & Jewett, M. C. (2014). Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. *Frontiers in Chemistry, 2(34),* 844-853.

Howard, R., Abotsi, E., Van Rensburg, E. J., & Howard, S. (2004). Lignocellulose biotechnology: Issues of bioconversion and enzyme production. *African Journal of Biotechnology, 2*(12), 602-619.

Hu, X. P. (2010). Identification and characterisation of novel cellulolytic genes using metagenomics. *Masters dissertation*, University of the Western Cape.

Huber, G. W., Iborra, S., & Corma, A. (2006). Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. *Chemical Reviews, 106*(9), 4044-4098.

Isaksson, L., Enberg, J., Neutze, R., Karlsson, B. G., & Pedersen, A. (2012). Expression screening of membrane proteins with cell-free protein synthesis. *Protein Expression and Purification, 82*(1), 218-225.

Iskakova, M. B., Szaflarski, W., Dreyfus, M., Remme, J., & Nierhaus, K. H. (2006). Troubleshooting coupled *in vitro* transcription-translation system derived from *Escherichia coli* cells: Synthesis of high-yield fully active proteins. *Nucleic Acids Research, 34*(19), 135-146.

Jewett, M. C., Fritz, B. R., Timmerman, L. E., & Church, G. M. (2013). *In vitro* integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Molecular Systems Biology, 9*(1), 641-651.

Kai, L., Dötsch, V., Kaldenhoff, R., & Bernhard, F. (2013). Artificial environments for the co-translational stabilization of cell-free expressed proteins. *PloS One, 8*(2), 566-573.

Kasana, R. C., Salwan, R., Dhar, H., Dutt, S. & Gulati, A. (2008). A rapid and easy method for the detection of microbial cellulases on agar plates using gram's iodine. *Curre Microbiology* 57, 503-507.

Kayser, K., Yun, C., & Kilbane, J. (1996). The construction and use of promoter probe vectors for *Rhodococcus* sp. *Actinomycetes, 7*, 55-65.

Keshwani, D. R., & Cheng, J. J. (2009). Switchgrass for bioethanol and other value-added applications: A review. *Bioresource Technology, 100*(4), 1515-1523.

Kim, D. M., Kigawa, T., Choi, C. Y., & Yokoyama, S. (1996). A highly efficient cell-free protein synthesis system from *Escherichia coli*. *European Journal of Biochemistry, 239*(3), 881-886.

Kim, H., & Kim, D. (2009). Methods for energizing cell-free protein synthesis. *Journal of Bioscience and Bioengineering, 108*(1), 1-4.

Kim, T., Keum, J., Oh, I., Choi, C., Park, C., & Kim, D. (2006). Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of Biotechnology, 126*(4), 554-561.

Kirk, T. K., Connors, W., & Zeikus, J. (1977). Advances in understanding the microbiological degradation of lignin. *The structure, biosynthesis, and degradation of wood*, 11, 369-394.

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

Koseki, T., Miwa, Y., Fushinobu, S., & Hashizume, K. (2005). Biochemical characterization of recombinant acetyl xylan esterase from *Aspergillus awamori* expressed in *Pichia pastoris*: Mutational analysis of catalytic residues. *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics, 1749*(1), 7-13.

Krastanova, I., Guarnaccia, C., Zahariev, S., Degrassi, G., & Lamba, D. (2005). Heterologous expression, purification, crystallization, X-ray analysis and phasing of the acetyl xylan esterase from *Bacillus pumilus*. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1748*(2), 222-230.

Krisana, A., Rutchadaporn, S., Jarupan, G., Lily, E., Sutipa, T., & Kanyawim, K. (2005). Endo-1, 4-beta-xylanase B from *Aspergillus cf. niger* BCC14405 isolated in Thailand: Purification, characterization and gene isolation. *Journal of Biochemistry and Molecular Biology, 38*(1), 17-30.

Kristensen, J. B., Felby, C., & Jorgensen, H. (2009). Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose. *Biotechnology for Biofuels, 2*(1), 1168-1181.

Kukutla, P., Lindberg, B. G., Pei, D., Rayl, M., Yu, W., Steritz, M., & Xu, J. (2014). Insights from the genome annotation of *Elizabethkingia anophelis* from the malaria vector *Anopheles gambiae*. *PloS One, 9*(5), 97715-97721.

Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives. *Journal of Industrial Microbiology & Biotechnology, 35*(5), 377-391.

Kwon, Y., & Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports, 5*(8663), 163-174.

Kwon, Y. C., Lee, K. H., Kim, H. C., Han, K., Seo, J. H., Kim, B. G., & Kim, D. M. (2010). Cloning-independent expression and analysis of ω-transaminases by use of a cell-free protein synthesis system. *Applied and environmental microbiology, 76*(18), 6295-6298.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680-685.

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

Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. *Journal of biotechnology, 56*(1), 1-24.

Leggewie, C., Henning, H., Schmeisser, C., Streit, W. R., & Jaeger, K. (2006). A novel transposon for functional expression of DNA libraries. *Journal of Biotechnology, 123*(3), 281-287.

Leis, B., Angelov, A., & Liebl, W. (2013). Screening and expression of genes from metagenomes. *Advances in Applied Microbiology, 83*, 1-68.

Levy, M., Griswold, K. E., & Ellington, A. D. (2005). Direct selection of trans-acting ligase ribozymes by *in vitro* compartmentalization. *RNA (New York, N.Y.), 11*(10), 1555-1562.

Li, J., Gu, L., Aach, J., & Church, G. M. (2014). Improved cell-free RNA and protein synthesis system. *PloS One, 9*(9), 1062-1072.

Liles, M. R., Williamson, L. L., Rodbumrer, J., Torsvik, V., Goodman, R. M., & Handelsman, J. (2008). Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Applied and Environmental Microbiology, 74*(10), 3302-3305.

Liu, H., Cheng, S., & Logan, B. E. (2005). Production of electricity from acetate or butyrate using a single-chamber microbial fuel cell. *Environmental science & technology, 39*(2), 658-662.

Liu, H., Feng, Y., Zhao, D., & Jiang, J. (2012). Evaluation of cellulases produced from four fungi cultured on furfural residues and microcrystalline cellulose. *Biodegradation, 23*(3), 465-472.

López-Barrón, C. R., Li, D., Wagner, N. J., & Caplan, J. L. (2014). Triblock Copolymer Self-Assembly in Ionic Liquids: Effect of PEO Block Length on the Self-Assembly of PEO–PPO–PEO in Ethylammonium Nitrate. *Macromolecules, 47*(21), 7484-7495.

Ma, F., Xie, Y., Huang, C., Feng, Y., & Yang, G. (2014). An improved single cell ultrahigh throughput screening method based on *in vitro* compartmentalization. *PloS one*, *9*(2), 1-10.

Malherbe, S., & Cloete, T. E. (2002). Lignocellulose biodegradation: Fundamentals and applications. *Reviews in Environmental Science and Biotechnology*, *1*(2), 105-114.

Martinez, A., Kolvek, S. J., Yip, C. L., Hopke, J., Brown, K. A., MacNeil, I. A., & Osburne, M. S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology*, *70*(4), 2452-2463.

Martins, L. F., Antunes, L. P., Pascon, R. C., de Oliveira, J. C. F., Digiampietri, L. A., Barbosa, D., Peixoto, B.M., Vallim M. A., Viana-Niero, C., Ostroski, E.H., & Setubal, J. C. (2013). Metagenomic analysis of a tropical composting operation at the Sao Paulo zoo park reveals diversity of biomass degradation functions and organisms.*Plos One*, 8 (6), 71-81.

Mielenz, J. R. (2001). Ethanol production from biomass: Technology and commercialization status. *Current Opinion in Microbiology*, *4*(3), 324-329.

Miller, O. J., Bernath, K., Agresti, J. J., Amitai, G., Kelly, B. T., Mastrobattista, E.,Taly, V., Magdassi, S., Tawfik, D.S., & Griffiths, A. D. (2006). Directed evolution by *in vitro* compartmentalization. *Nature Methods*, *3*(7), 561-570.

Miller, D. N., Bryant, J. E., Madsen, E. L., & Ghiorse, W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology, 65*(11), 4715-4724.

Mitani, Y., Meng, X., Kamagata, Y., & Tamura, T. (2005). Characterization of LtsA from *Rhodococcus erythropolis*, an enzyme with glutamine amidotransferase activity. *Journal of Bacteriology, 187*(8), 2582-2591.

Moore, T. C. (1967). Kinetics of growth retardant and hormone interactions in affecting cucumber hypocotyl elongation. *Plant Physiology, 42*(5), 677-684.

Naik, S., 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.

Nakashima, N., & Tamura, T. (2004). A novel system for expressing recombinant proteins over a wide temperature range from 4 to 35˚C. *Biotechnology and Bioengineering, 86*(2), 136-148.

Nakashima, N., Mitani, Y., & Tamura, T. (2005). Actinomycetes as host cells for production of recombinant proteins. *Microbial Cell Factories, 4*(1), 5557-5566.

Nevondo, N. (2014). Development of a high throughput cell-free metagenomic screening platform.*PhD dissertation*, University of the Western Cape.

Ngwa, C. J., Glöckner, V., Abdelmohsen, U. R., Scheuermayer, M., Fischer, R., Hentschel, U., & Pradel, G. (2013). 16S rRNA gene-based identification of *Elizabethkingia meningoseptica* (Flavobacteriales: Flavobacteriaceae) as a dominant midgut bacterium

of the Asian malaria vector *Anopheles stephensi* (Dipteria: Culicidae) with antimicrobial activities. *Journal of medical entomology*, *50*(2), 404-414.

Nisole, A., Lussier, F., Morley, K. L., Shareck, F., Kazlauskas, R. J., Dupont, C., & Pelletier, J. N. (2006). Extracellular production of *Streptomyces lividans* acetyl xylan esterase A in *Escherichia coli* for rapid detection of activity. *Protein Expression and Purification*, *46*(2), 274-284.

Olofsson, K., Bertilsson, M., & Lidén, G. (2008). A short review on SSF-an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol Biofuels*, *1*(7), 1-14.

Ou, S., Zhang, J., Wang, Y., & Zhang, N. (2011). Production of feruloyl esterase from *Aspergillus niger* by solid-state fermentation on different carbon sources. *Enzyme Research*, *20*(11), 848-869.

Pang, M., Abdullah, N., Lee, C., & Ng, C. (2008). Isolation of high molecular weight DNA from forest topsoil for metagenomic analysis. *Asia Pac J Mol Biol Biotechnol*, *16*(2), 35-41.

Park, H., Kayser, K. J., Kwak, J., & Kilbane II, J. J. (2004). Heterologous gene expression in *Thermus thermophilus*: β-galactosidase, dibenzothiophene monooxygenase, PNB carboxy esterase, 2-aminobiphenyl-2, 3-diol dioxygenase, and chloramphenicol acetyl transferase. *Journal of Industrial Microbiology and Biotechnology*, *31*(4), 189-197.

Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A., & Johnson, D. K. (2010). Research cellulose crystallinity index: Measurement techniques and their impact on interpreting cellulase performance.*Biotechnol Biofuels*, *3*(10), 1-10.

Pedersen, A., Hellberg, K., Enberg, J., & Karlsson, B. G. (2011). Rational improvement of cell-free protein synthesis. *New Biotechnology, 28*(3), 218-224.

Penttilä, M. (2010). NEMO: Novel high performance enzymes and micro-organisms for conversion of lignocellulosic biomass to bioethanol. *Journal of Biotechnology*, *150*, 11-31.

Pham, V. H., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology, 30*(9), 475-484.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of paederus beetles. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14002-14007.

Promega Guide. (2007). *Promega Protocols and Applications Guide*. Promega Corporation.

Qin, X., & Sarnow, P. (2004). Preferential translation of internal ribosome entry site-containing mRNAs during the mitotic cycle in mammalian cells. *The Journal of Biological Chemistry*, *279*(14), 13721-13728.

Ragauskas, A.J, *et al*., (2014). Lignin valorization: Improving lignin processing in the biorefinery. *Science,* 6185 (344), 16-42.

Rajan, B. M., & Kannabiran, K. (2014). Extraction and identification of antibacterial secondary metabolites from marine *Streptomyces* sp. VITBRK2. *International Journal of Molecular and Cellular Medicine, 3*(3), 130-141.

Ranji, A., Wu, J. C., Bundy, B. C., & Jewett, M. C. (2013). Transforming synthetic biology with cell-free systems. *Synthetic Biology: Tools and Applications*, Elsevier (pp. 277-301).

Rapa, R. A., Islam, A., Monahan, L. G., Mutreja, A., Thomson, N., Charles, I. G., & Labbate, M. (2014). A genomic island integrated into *rec*A of *Vibrio cholerae* contains a divergent *rec*A and provides multi-pathway protection from DNA damage. *Environmental Microbiology, 17(4),* 1090-1102.

Reysenbach, A. L., Hamamura, N., Podar, M., Griffiths, E., Ferreira, S., Hochstein, R,, Heidelberg, J., Johnson, J., Mead, D., Pohorille, A., Sarmiento, M., Schweighofer, K., Seshadri, R., Voytek, M.A., (2009). Complete and draft genome sequences of six members of the *Aquificales. J Bacteriology* ,191, 1992-1993.

Rosano, G. L., & Ceccarelli, E. A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microbial Cell Factories*, 41(*8)*, 28-41.

Rosenblum, G., & Cooperman, B. S. (2014). Engine out of the chassis: Cell-free protein synthesis and its uses. *FEBS Letters, 588*(2), 261-268.

Rothe, A., Surjadi, R. N., & Power, B. E. (2006). Novel proteins in emulsions using *in vitro* compartmentalization. *Trends in Biotechnology*, *24*(12), 587-592.

Sabree, Z. L., Kambhampati, S., & Moran, N. A. (2009). Nitrogen recycling and nutritional provisioning by blattabacterium, the cockroach endosymbiont. *Proceedings of the National Academy of Sciences of the United States of America, 106*(46), 19521-19526.

Saha, B. C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*, *30*(5), 279-291.

Sambrook, J., & Maniatis, T. (1989). Molecular cloning: A laboratory manual, Cold Spring Harbor.

Sambrook, J. R., & Russel, D. W. (2001). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor.

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

Saulnier, L., Marot, C., Chanliaud, E., & Thibault, J. (1995). Cell wall polysaccharide interactions in maize bran. *Carbohydrate Polymers*, *26*(4), 279-287.

Scanlon, T. C., Dostal, S. M., & Griswold, K. E. (2014). A high-throughput screen for antibiotic drug discovery. *Biotechnology and bioengineering*, *111*(2), 232-243.

Schoborg, J. A., Hodgman, C. E., Anderson, M. J., & Jewett, M. C. (2014). Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. *Biotechnology journal*, *9*(5), 630-640.

Sekine, M., Tanikawa, S., Omata, S., Saito, M., Fujisawa, T., Tsukatani, N., ... & Harayama, S. (2006). Sequence analysis of three plasmids harboured in Rhodococcus erythropolis strain PR4. *Environmental microbiology*, *8*(2), 334-346.

Sheridan, G. E. C., Masters, C. I., Shallcross, J. A., & Mackey, B. M. (1998). Detection of mRNA by Reverse Transcription-PCR as an Indicator of Viability in *Escherichia coli* cells. *Applied and Environmental Microbiology*, *64*(4), 1313-1318.

Shin, J., & Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. *ACS Synthetic Biology, 1*(1), 29-41.

Shrestha, P., Michaele Holland, T., & Charles Bundy, B. (2012). Streamlined extract preparation for *Escherichia coli*-based cell-free protein synthesis by sonication or bead vortex mixing. *Biotechniques, 53*(3), 163-175.

Sierecki, E., Giles, N., Polinkovsky, M., Moustaqil, M., Alexandrov, K., & Gambin, Y. (2013). A cell-free approach to accelerate the study of protein–protein interactions *in vitro*. *Interface focus, 3*(5), 88-98.

Simon, C., & Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology, 85*(2), 265-276.

Simon, C., & Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology, 77*(4), 1153-1161.

Singh, B. K., & Macdonald, C. A. (2010). Drug discovery from uncultivable microorganisms. *Drug Discovery Today, 15*(17), 792-799.

Smith, M. T., Hawes, A. K., Shrestha, P., Rainsdon, J. M., Wu, J. C., & Bundy, B. C. (2014). Alternative fermentation conditions for improved *Escherichia coli*-based cell-free protein synthesis for proteins requiring supplemental components for proper synthesis. *Process Biochemistry, 49*(2), 217-222.

Stoletzki, N., & Eyre-Walker, A. (2007). Synonymous codon usage in *Escherichia coli*: Selection for translational accuracy. *Molecular Biology and Evolution, 24*(2), 374-381.

Subramani, R., & Aalbersberg, W. (2012). Marine actinomycetes: An ongoing source of novel bioactive metabolites. *Microbiological research*, *167*(10), 571-580.

Sukumaran, R. K., Singhania, R. R., Mathew, G. M., & Pandey, A. (2009). Cellulase production using biomass feed stock and its application in lignocellulose saccharification for bio-ethanol production.*Renewable Energy*, *34*(2), 421-424.

Sun, N., Rahman, M., Qin, Y., Maxim, M. L., Rodríguez, H., & Rogers, R. D. (2009). Complete dissolution and partial delignification of wood in the ionic liquid 1-ethyl-3-methylimidazolium acetate. *Green Chemistry*, *11*(5), 646-655.

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

Sunna, A., Moracci, M., Rossi, M., & Antranikian, G. (1997). Glycosyl hydrolases from hyperthermophiles. *Extremophiles*, *1*(1), 2-13.

Taherzadeh, M. J., & Karimi, K. (2007). Enzymatic-based hydrolysis processes for ethanol from lignocellulosic material. *Bioresources*, *2*(4), 707-738.

Takai, S., Hines, S. A., Sekizaki, T., Nicholson, V. M., Alperin, D. A., Osaki, M., & Prescott, J. F. (2000). DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. *Infection and Immunity*, *68*(12), 6840-6847.

Taupp, M., Mewis, K., & Hallam, S. J. (2011). The art and design of functional metagenomic screens. *Current opinion in biotechnology*, *22*(3), 465-472.

Tawfik, D. S., & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnology*, *16*(7), 652-656.

Tebbe, C. C., & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and Environmental Microbiology, 59*(8), 2657-2665.

Teeling, H., & Glockner, F. O. (2012). Current opportunities and challenges in microbial metagenome analysis-a bioinformatic perspective. *Briefings in Bioinformatics, 13*(6), 728-742.

Terrón-González, L., Medina, C., Limón-Mortés, M., & Santero, E. (2013). Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Scientific Reports,* 1107 (3), 1-8.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentations, 2*(3), 1-12.

Topakas, E., Vafiadi, C., & Christakopoulos, P. (2007). Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry, 42*(4), 497-509.

Trincone, A., & Schwegmann-Weßels, C. (2014). Looking for a needle in a haystack: Cellular proteins that may interact with the tyrosine-based sorting signal of the TGEV S protein. *Virus Research, 208(4),* 33-40.

Troeschel, S. C., Thies, S., Link, O., Real, C. I., Knops, K., Wilhelm, S., & Jaeger, K. (2012). Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *Journal of Biotechnology*, *161*(2), 71-79.

Tshukudu, O. (2012). Cloning and characterization of a cellulase from a metagenomic library. Honours dissertation, University of the Western Cape.

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.

Uchiyama, T., & Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: Challenges to efficient screening. *Current Opinion in Biotechnology, 20*(6), 616-622.

Van Dyk, J., & Pletschke, B. (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.

Vanholme, R., Morreel, K., Ralph, J., & Boerjan, W. (2008). Lignin engineering. *Current opinion in plant biology, 11*(3), 278-285.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of actinobacteria: Tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews , 71*(3), 495-548.

Vieille, C., & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews, 65*(1), 1-43.

Viell, J., Wulfhorst, H., Schmidt, T., Commandeur, U., Fischer, R., Spiess, A., & Marquardt, W. (2013). An efficient process for the saccharification of wood chips by combined ionic liquid pretreatment and enzymatic hydrolysis. *Bioresource Technology, 146*(3), 144-151.

Voget, S., Steele, H., & Streit, W. (2006). Characterization of a metagenome-derived halotolerant cellulase. *Journal of Biotechnology, 126*(1), 26-36.

Wang, J. Y., Chou, L. S., & White, B. N. (1999). Mitochondrial DNA analysis of sympatric morphotypes of bottlenose dolphins (genus: *Tursiops*) in Chinese waters. *Molecular Ecology*, *8*(10), 1603-1612

Wang, Y., Hayatsu, M., & Fujii, T. (2012). Extraction of bacterial RNA from soil: challenges and solutions. *Microbes and Environments*, *27*(2), 111-121.

Watson, B. J., Zhang, H., Longmire, A. G., Moon, Y. H., & Hutcheson, S. W. (2009). Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans*. *Journal of Bacteriology*, *191*(18), 5697-5705.

Westenberg, M., Bamps, S., Soedling, H., Hope, I. A., & Dolphin, C. T. (2010). *Escherichia coli* MW005: Lambda red-mediated recombineering and copy-number induction of oriV-equipped constructs in a single host. *BMC Biotechnology, 10 (1),* 27-35.

Willey, J., Sherwood, L., & Woolverton, C. (2008). Actinomycetes in: Prescott, harley and Klein's microbiology. McGrawhill, New York (pp.568-580).

Willey, J. M., Willems, A., Kodani, S., & Nodwell, J. R. (2006). Morphogenetic surfactants and their role in the formation of aerial hyphae in *Streptomyces coelicolor*. *Molecular Microbiology*, *59*(3), 731-742.

Wilson, M. C., & Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry & Biology*, *20*(5), 636-647.

Wyman, C. (1996). Ethanol production from biomass; overview in C.E, Wynman (Ed.). Handbook on bioethanol. Production and utilization. Taylor and Francis group, Washington, D.C, CRC (pp.398-400).

Yeates, C., Gillings, M., Davison, A., Altavilla, N., & Veal, D. (1998). Methods for microbial DNA extraction from soil for PCR amplification. *Biological Procedures Online, 1*(1), 40-47.

Yun, J., & Ryu, S. (2005). Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell Factories, 4*(1), 8.

Zaldivar, J., Nielsen, J., & Olsson, L. (2001). Fuel ethanol production from lignocellulose: A challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology, 56*(1-2), 17-34.

Zawada, J. F. (2012). Preparation and testing of *E. coli* S30 *in vitro* transcription translation extracts. *Ribosome display and related technologies* (pp. 31-41).

Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., & Murray, C. J. (2011). Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. *Biotechnology and Bioengineering, 108*(7), 1570-1578.

Zeng, J., Helms, G. L., Gao, X., & Chen, S. (2013). Quantification of wheat straw lignin structure by comprehensive NMR analysis. *Journal of Agricultural and Food Chemistry, 61*(46), 10848-10857.

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology, 62*(2), 316-322.