Screening of Fungal Isolates for Endoglucanase Activity and Cloning of a Thermostable Endoglucanase from *Phialophora alba*

BY

ASHLEY ISAAC

Submitted in fulfilment of the academic requirements for the degree of Master of Science (MSc) in the Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science at the University of KwaZulu-Natal (Westville Campus).

As the candidate's supervisor, I have approved this dissertation for submission.

Signed:_____ Name:_____ Date:_____

PREFACE

The experimental work described in this dissertation was carried out in the School of Life Science; University of KwaZulu-Natal (Westville Campus), Durban, South Africa between August 2011 to July 2013 and July 2015 to November 2015 under the supervision of Dr. R. Govinden.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

DECLARATION 1- PLAGIARISM

I,……., declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.

2. This dissertation has not been submitted for any degree or examination at any other University.

3. This dissertation does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced or adapted from other persons.

4. This dissertation does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   a. Their words have been re-written but the general information attributed to them has been referenced.
   b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the reference sections.

Signed:

…….

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the following individuals from the Discipline of Microbiology, UKZN (Westville campus):

Dr. Roshini Govinden for her patience, understanding and guidance throughout the course of this study.

The academic and support staff as well as fellow postgraduate students for their valuable input while troubleshooting.

My lab mates and dear friends: Lucretia, Renee, Chika, Tom, Avishka, Philisiwe, Ruqsar and Sarisha for endless hours of conversation and fun in the lab. That kept me sane.

Special thanks to my family and friends for their support, motivation and encouragement.

Lastly I would like to thank The National Research Foundation for financial assistance which allowed me the opportunity to pursue this degree.

TABLE OF CONTENTS

	PAGE
Abstract	**i**
List of Figures	**iii**
List of Tables	**vii**
Chapter 1: Literature Review	**1**
1.1 Introduction	**1**
1.2 Structure of lignocellulose	**3**
1.2.1 Cellulose	**4**
1.2.2 Hemicellulose	**6**
1.2.3 Lignin	**7**
1.3 Enzymatic degradation of cellulose	**7**
1.3.1 Cellulases	**7**
1.3.1.1 Modular nature of cellulases	**8**
1.3.1.2 Catalytic mechanism of action	**9**
1.3.1.3 Classification of glycoside hydrolases	**11**
1.3.1.4 Occurrence in nature	**12**
1.3.2 Hemicellulases	**14**
1.3.3 Ligninases	**15**
1.4 Industrial application of cellulolytic enzymes	**15**
1.4.1 Fuel ethanol	**16**
1.4.1.1 Pretreatment strategies	**18**
1.4.1.1.1 Physical pretreatment	**18**
1.4.1.1.2 Chemical pretreatment	**19**
1.4.1.1.3 Biological pretreatment	**21**
1.4.1.2 Cellulose hydrolysis and fermentation	**22**
1.5 Heterologous expression of cellulases	**23**
1.5.1 *Escherichia coli*	**24**
1.5.2 *Saccharomyces cerevisiae*	**25**
1.5.3 *Pichia pastoris* and other potential hosts	**26**
1.6 Scope of the study	**28**

1.7 Hypothesis **28**
1.8 Objectives **29**

Chapter 2: Screening of fungal isolates from *Eucalyptus* spp. woodchips and the characterization and partial purification of a thermostable endoglucanase **30**

2.1 Introduction **30**
2.2 Materials and methods **33**
2.2.1 Screening for endoglucanase production **33**
2.2.2 Time course analysis for optimal production of endoglucanases **33**
2.2.3 Determination of endoglucanase activity **34**
2.2.4 Determination of pH and temperature optima of crude extracts **35**
2.2.5 Determination of pH and temperature stability of crude extracts **35**
2.2.6 Determination of molecular weight of the endoglucanase **36**
2.2.7 Partial purification of the endoglucanase **37**
2.3 Results and Discussion **37**
2.3.1 Screening for endoglucanase production **37**
2.3.2 Time course analysis for optimal production of endoglucanases **39**
2.3.3 Characterization of crude enzyme extract **41**
2.3.4 SDS-PAGE and zymogram analysis **48**
2.3.5 Purification **50**
2.4 Conclusion and way forward **52**

Chapter 3: Carbohydrate-active Enzyme Analysis of *Phialophora alba* isolated from *Eucalyptus* spp. woodchips **54**

3.1 Introduction **54**
3.2 Materials and methods **57**
3.2.1 DNA extraction and sequencing **57**
3.2.2 Data-mining for CAZymes **59**
3.3 Results and Discussion **59**
3.3.1 Isolation of genomic DNA from *P. alba* **59**
3.3.2 Genome summary **60**
3.3.3 The carbohydrate-active enzymes of *P. alba* **61**
3.3.4 Cellulose degrading capacity **67**
3.3.5 Hemicellulose degrading capacity **69**

3.3.6 Other plant polysaccharide degrading capabilities 70
3.3.7 Chitin degrading capacity 71
3.4 Conclusion 72

Chapter 4: Cloning and expression of the endoglucanase gene from *Phialophora alba* isolated from *Eucalyptus* spp. woodchips 73
4.1 Introduction 73
4.2 Materials and methods 75
4.2.1 Strains, vectors and media used 75
4.2.2 DNA isolation 76
4.2.3 RNA isolation and cDNA synthesis 77
4.2.4 PCR amplification of endoglucanase gene 77
4.2.5 Sequence and structure analysis 79
4.2.6 Cloning and expression of endoglucanase gene 80
4.2.6.1 Transformation of *E. coli* DH5α and *E. coli* HB101 80
4.2.6.2 Expression in *Pichia pastoris* GS115 80
4.2.6.3 Expression in *E. coli* BL21(DE3) 81
4.2.7 Enzyme activity assays 82
4.2.8 SDS-PAGE and zymogram analysis 83
4.3 Results and Discussion 83
4.3.1 DNA, RNA and poly(A) mRNA isolation 83
4.3.2 PCR amplification of the endoglucanase gene from *P. alba* 84
4.3.3 Sequence analysis of endoglucanase gene from *P. alba* 91
4.3.4 Expression in *P. pastoris* 92
4.3.5 Expression in *E. coli* BL21(DE3) 100
4.4 Conclusion 102

Chapter 5: Concluding remarks 103

References 106

Appendix 137

ABSTRACT

Microbes are fast becoming the solution to the energy crisis. Alternative fuels that can be produced by microbes include bioethanol, biodiesel, methane and biohydrogen. This study aimed to identify and characterize a thermostable endoglucanase, a carbohydrate active enzyme which would be part of a cocktail of enzymes used to breakdown cellulose to liberate glucose. The glucose in turn can be fermented to produce bioethanol.

Twenty seven fungal isolates were screened for endoglucanase activity on agar plates supplemented with 1% carboxymethyl cellulose. Zones of hydrolysis were visualized by Congo red staining. Ten isolates were selected based on the diameters of hydrolytic zones (presenting high endoglucanase activity) for time course analysis for optimal production of endoglucanase. Isolate D9 (*Aspergillus fumigatus*) was found to be the highest producer and time course analysis revealed that the highest production is achieved on Day 9 after inoculation. The crude enzyme extract was produced in bulk and precipitated in 40 - 60% ammonium sulphate in order to remove unwanted proteins and to concentrate the endoglucanases. However, following dialysis, more than 80% proteins and endoglucanase activity was lost. Other avenues such as IEF fractionation and freeze drying were explored to achieve partial purification and concentration, respectively, but these proved fruitless.

Thus the focus of the study shifted to cloning, heterologously expressing and characterizing an endoglucanase from *Phialophora alba*, a fungus known to produce lignocellulosic enzymes. Whole genome sequencing was performed and open reading frames (ORFs) were identified using the CLC Main Workbench. These ORFs were analyzed using the CAZymes Analysis Toolkit to identify carbohydrate active enzyme genes. A single endoglucanase gene (GH5) was identified and selected for the duration of the study. Primer sets GH5_ORF_F; GH5_ORF_R and Exp_GH5_ORF_F; Exp_GH5_ORF_R, were specifically designed for

cloning into plasmids pTZ57R/T and pPIC9, respectively. The recombinant plasmid pPIC9_13_Eg_GH5 was successfully integrated into *Pichia pastoris* GS115 genome. The heterologous enzyme was produced; however, no endoglucanase activity was detected. The gene was expressed in *Escherichia coli* BL21 (DE3), however, no endoglucanase activity was detected in this instance as well. Bioinformatic analysis of the CAZome of *P. alba* shows a preference for xylan and chitin over cellulose. *P. alba* appears to be a poor cellulose degrader possessing only a single endoglucanase gene, however, it does possess auxillary activity enzymes enzymes that may also be involved in the breakdown on cellulose. This therefore provides insight into why this enzyme could be inactive as well as sheds light on the plant polysaccharide degradative abilities of this fungus.

LIST OF FIGURES

PAGE

Figure 1.1: Cellulose structure and its' arrangement into microfibrils and fibrils within the plant cell wall (http://www.bio.miami.edu/dana/pix/cellulose_microfibrils.jpg) **5**

Figure 1.2: Double-displacement mechanism (http://www.cazy.org/fam/ghf INV RET.html#inv) **10**

Figure 1.3: Single-displacement mechanism (http://www.cazy.org/fam/ghf INV RET.html#inv) **10**

Figure 2.1: Time course analysis for optimal endoglucanase production of fungal isolates displaying desirable endoglucanase activity **40**

Figure 2.2: Effect of pH on endoglucanase activity from crude culture filtrate of *A. fumigatus* **42**

Figure 2.3: pH Stability profile of endoglucanase from crude culture filtrate of *A. fumigatus* at A. pH 4, B. pH5, C. pH 6, D. pH 7, E. pH 8, F. pH 9 **44**

Figure 2.4: Effect of temperature on endoglucanase activity from crude culture filtrate of *A. fumigatus* **45**

Figure 2.5: Thermostability profile of endoglucanase from crude culture filtrate of *A. fumigatus* at A. 50 °C, B. 60 °C, C. 70 °C, D. 80 °C, E. 90 °C **47**

Figure 2.6: Silver stained 12% SDS-PAGE of day 9 unconcentrated crude culture filtrate of *A. fumigatus* (Lane 1: Marker: Spectra Multicolour Broad Range Protein Ladder; 2: Day 9 *A. fumigatus* culture filtrate) **48**

Figure 2.7: Silver stained 12% SDS-PAGE of 10× concentrated crude enzymes (1: Marker: PageRuler Plus Prestained Protein Ladder; 2 and 4: 10X concentrated sample of *A. fumigatus* culture filtrate) **49**

Figure 2.8: Zymogram analysis of Native PAGE (1: Marker: PageRuler Plus Prestained Protein Ladder; 2, 5 and 7: 10× concentrated sample of *A. fumigatus* culture filtrate) **50**

Figure 3.1 Functional distinction of putative CAZymes from *P. alba* genome (PCW: plant cell wall; FCW: fungal cell wall; BPG: bacterial peptidoglycan, ESR: energy storage and recovery; PG: protein glycosylation (1)) Total count excludes PLs, CBMs and those CAZymes whose exact function has not yet been determined (GHs: 92, 109, 125, 127, 128, 132 and CE10) **65**

Figure 3.2 Chord diagram illustrating the substrate specificities of carbohydrate active enzyme families from *P. alba* (Gal: Galactomannan; Pec: Pectin; Cel: Cellulose; Sta: Starch; Chi: Chitin; Xyg: Xyloglucan; Xyl: Xylan) (Particular attention [black arrows] is drawn to those families of GHs that potentially hydrolyse cellulose) **66**

Figure 4.1: PCR amplification using primer set F10EgF and F10EgR and double-stranded cDNA as template (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: cDNA using dT_{18}; 3: cDNA using F10EgR) **85**

Figure 4.2: PCR amplification using primers F10EgF and EgF10R at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: 1.25 mM; 3: 2 mM; 4: 2.5 mM; 5: 3 mM) **86**

Figure 4.3: PCR amplification using degenerate primers Eg-GH5-f and Eg-GH5-r at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: No template control; 3: 0 mM; 4: 1 mM; 5: 1.5 mM; 6: 2 mM; 7: 2.5 mM; 8: 3 mM) **87**

Figure 4.4: PCR amplification using cellulase core primers Eg5-F and Eg5-R at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: No template control; 3: 1.5 mM; 4: 2 mM; 5: 2.5 mM; 6: 3 mM) **87**

Figure 4.5: PCR amplification using degenerate GH45 primers and degenerate CBHII primers (Lane 1: No template control GH45; 2: No template control CBHII; 3: GH45(1.25 mM); 4: CBHII (1.5 mM); 5: Marker (bp): FastRuler Middle Range DNA Ladder) **88**

Figure 4.6: PCR re-amplification of excised bands from initial amplification with degenerate GH45 primers (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder; 2: ~2000 bp excised band; 3: ~1000 bp excised band; 4: ~850 bp excised band; 5: No template control) **88**

Figure 4.7: PCR amplification of endoglucanase gene from *P. alba* using GH5_ORF primers (G) and GH5_ORF_Flank primers (F) at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: G 1 mM, 3: G 1.5 mM, 4: G 2mM, 5: F 1 mM, 6: 1.5 mM, 7: F 2 mM) **89**

Figure 4.8: Restriction analysis of selected positive clones (Lane 1: Marker: FastRuler Middle Range DNA Ladder, 2: Plasmid G1, 3: Plasmid cDNA1, 4 Plasmid cDNA2, 5: Plasmid F1, 6: Plasmid F2, 7: Plasmid F3, 8: Plasmid F4) (G and F refer to the primers sets discussed previously, cDNA refers to cDNA being used as template DNA) **90**

Figure 4.9: The 416 amino acid long sequence of the *P. alba* GH5 endoglucanase translated from the nucleotide sequence **92**

Figure 4.10: The tertiary structures of EgGH5 from *P. alba* (A) and the GH5 protein from *M. thermophila* (B) (models predicted and rendered using the Phyre^2 web portal) **92**

Figure 4.11: Section of sequence alignment (1-160) of translated native GH5 endoglucanase gene (*P. alba*_GH5) and endoglucanase gene integrated into *Pichia* genome (amino acid changes indicated by black arrows) **94**

Figure 4.12: Congo red plate screening assay (CMC) of *Pichia* clones from transformations performed with recombinant vector linearlized with *Sal*I, *Bgl*II and *Not*I **95**

Figure 4.13: PCR amplification of endoglucanase gene from *Pichia* clones using GH5_ORF primers (Lane 1: Marker: GeneRuler 100 bp Plus DNA Ladder, 2: No template control, 3: *Bgl*II no insert linear pPIC9, 4: *Sal*I no insert linear pPIC9, 5: B2, 6:B6, 7: S1, 8: S2, 9:S3) (B and S: *Bgl*II and *Sal*I linearized vector clones) **96**

Figure 4.14: Induction of recombinant protein expression by *Pichia* clones with 0.5% methanol over a period of 4 days **97**

Figure 4.15: Silver stained SDS-PAGE gel of recombinant proteins expressed by *Pichia* clones after induction with methanol (Lane 1: Marker: PageRuler Plus Prestained Protein Ladder, Lanes: 2-4: *Pichia* pPIC9-no insert Day 0, 2, 4; Lanes: 5-7: *Pichia* pPIC9-S1 Day 0, 2, 4; Lanes: 8-10: *Pichia* pPIC9-B1 Day 0, 2, 4) **98**

Figure 4.16: Colony PCR of *E. coli* BL21 (DE3) clones (Lane 1: Marker: GeneRuler 100 bp Plus DNA Ladder, 2: No template control, 3: pET-22b-no insert, 4: pPIC9_13_Eg_GH5, 5-9: Clones eB1 – eB5)**101**

Figure 4.17: Congo red plate screening assay of intracellular proteins of *E. coli* BL21 (DE3) clones (Pos: Fungal extract known to possess endoglucanase activity, Spots 2 – 6: eB1 – eB5) **101**

Figure 4.18: Coomassie stain of SDS-PAGE of intracellular proteins from recombinant *E. coli* BL21(DE3) clones following 6 hours induction with 1 mM IPTG (Lane 1: Marker: Spectra Multicolor Broad Range Protein Ladder, 2: *E. coli* BL21(DE3) with pET-22b-no insert, 3: eB1, 4: eB3, 5: eB4, 6: eB5) **102**

Figure A1: Glucose standard curve for the determination of endoglucanase activity during screening phase **137**

Figure A2: BSA standard curve used for the determination protein concentration *via* the Bradford assay **137**

Figure A3: Glucose standard curve for the determination of endoglucanase activity at: A: pH 4; B: pH 5; C: pH 6; D: pH 7; E: pH 8; F: pH 9 **138**

Figure A4: Glucose standard curve for the determination of endoglucanase activity at: A: 50°C; B: 60°C; C: 70°C; D: 80°C; E: 90 **139**

Figure A5: Zymogram analysis of *Pichia* clones extracellular proteins (Lane 1: Marker: PageRuler Plus Prestained Protein Ladder, Lanes: 2-4: *Pichia* pPIC9-no insert Day 0, 2, 4; Lanes: 5-7: *Pichia* pPIC9-S1 Day 0, 2, 4; Lanes: 8-10: *Pichia* pPIC9-B1 Day 0, 2, 4) **140**

Figure C1: Nucleotide sequence of GH5 endoglucanase gene from *P. alba* **142**

LIST OF TABLES

PAGE

Table 1.1: Cellulose content of various plant biomass (Demirbas, 2005; Menon and Rao, 2012) **5**

Table 1.2: Hemicellulases and their respective substrates **14**

Table 1.3: Properties of thermostable endoglucanases **17**

Table 1.4: Examples of cellulase genes expressed in alternate hosts **27**

Table 2.1: Preliminary assessment of endoglucanase activity of fungal isolates **38**

Table 2.2: Ammonium sulphate precipitation of crude culture filtrate of *A. fumigatus* **51**

Table 3.1: Distribution and comparison of the number of putative CAZyme genes between important fungal CAZyme producers **56**

Table 3.2: Putative carbohydrate-active enzymes present in the genome of *Phialophora alba* **61**

Table 4.1: List of primers used to amplify endoglucanase genes or fragments **78**

Table 4.2: Recombinant vectors and strains generated during the course of the study **82**

Table 4.3 DNA, RNA and poly-A-tailed mRNA quantities and qualities **83**

CHAPTER 1

Literature Review

1.1 Introduction

There is a progressive decrease of the world's energy resources which are mainly based on non-renewable fuels and it is predicted that global energy usage will almost double within the next two decades (Fortman *et al*., 2008; Sánchez and Cardona, 2008). Apart from powering 96% of the world's transportation, petroleum is also a key ingredient in plastics and fertilizers (Muanya, 2009). Its integral role in human civilization cannot be overestimated and if production should decline and fail, modern life would be practically impossible. For many years scientists, government officials, and business leaders have warned of peak oil. Peak oil, otherwise known as the Hubbert peak theory is the point in time when the maximum rate of global petroleum extraction and production is reached, after which the rate of production declines (Sorrell *et al*., 2010).

The diminishing petroleum reserve is only one half of the problem that the world faces. The second major problem is that combustion of fossil fuels results in the emission of carbon dioxide (CO_2). Carbon dioxide is a greenhouse gas which, with rising emission levels, contributes to global warming (Starr and Taggart, 2004). Thus, focus has shifted to trying to produce alternate fuels from renewable resources, which are economically sustainable as well as environmentally friendly.

Microbes are fast becoming the solution to the energy crisis. Alternative fuels that can be produced by microbes include, bioethanol, biodiesel, methane and biohydrogen (Fortman *et al*., 2008). Ethanol is the current leader in the biofuels market and is the most employed liquid biofuel either as a fuel or as a gasoline enhancer (Sánchez and Cardona, 2008). A host of countries, including South Africa have implemented or are going to implement the

addition of ethanol into gasoline. This fuel ethanol can be obtained from a variety of feedstocks, i.e. energy crops and lignocellulosic biomass. The main feedstock for ethanol production is sugar cane (cane juice or molasses) (Antoni *et al*., 2007). *Saccharomyces cerevisiae* is the microorganism of choice for bioethanol production as it is able to hydrolyse cane sucrose into its monomers which are the substrates that are fermented to ethanol in a batch fermentation process (Antoni *et al*., 2007; Sánchez and Cardona, 2008).

Other feedstocks for the production of ethanol include starch, starchy crops and lignocellulosic biomass (Hahn-Hägerdal *et al*., 2006). In order for starch to be utilized, it must first be hydrolysed into its glucose monomers before it is converted to ethanol by fermentation. The starch is hydrolysed as a result of the catalytic action of amylases. Starch may also be extracted from crops such as corn, wheat and cassava and used for the production of ethanol (Sánchez and Cardona, 2008). However, there is the debate over the ethics of using food crops for the production of ethanol. The use of edible crops as feedstocks for bioethanol production will inevitably lead to an increase in food prices (Cockerill and Martin, 2008).

Thus an appropriate alternative feedstock is lignocellulose which is a complex substrate and the most abundant polymer on earth. Lignocellulosic material for fuel production may be divided into six categories: crop residues, hardwoods, softwoods, cellulose wastes, herbaceous biomass and municipal solid wastes (Cardona *et al*., 2010). The processing challenge for ethanol production from this feedstock is the complexity of the composition of the polymer which consists of a matrix made up of cellulose, lignin and hemicellulose. This problem can be overcome by subjecting the material to a pre-treatment (Brodeur *et al*., 2011). The pre-treatment step plays an important role in the hydrolysis of cellulose in preparation for ethanol production. In the absence of a pre-treatment, hydrolysis of cellulose is less than

20%, while after a pre-treatment step; the hydrolysis of cellulose exceeds 90% (Sánchez and Cardona, 2008). Thus the main aims for the pre-treatment are to remove lignin and hemicellulose, as well as reduce crystalline cellulose and increase the porosity of the material. Pre-treatment methods include: physical methods (mechanical comminution and pyrolysis), physical-chemical methods (steam or CO_2 explosion, liquid hot water and ammonia fiber explosion), chemical methods (ozonolysis, dilute/concentrated-acid hydrolysis and oxidation) or biological methods (fungal enzyme treatment and bioorganosolv treatment) (Sánchez and Cardona, 2008; Brodeur *et al*., 2011).

The recalcitrance of lignocellulosic biomass can be overcome with the above pretreatments, however, a major impediment to their use is that they are not low-cost technologies (Brodeur *et al*., 2011). With a focus on biological pretreatment, lignocellulose can be degraded using a cocktail of carbohydrate-active enzymes. The cost implications involved in biological pretreatments may be reduced by identifying microorganisms that produce hydrolytic enzymes with desirable characteristics and produce them high titres by heterologous expression. Endoglucanases in particular are of significant importance as they are responsible for cleaving the internal glycosidic bonds of cellulose thus making the substrate more accessible to other cellulases such as CBHs and β-glucosidases in order to bring about efficient degradation of cellulose (Yennamalli *et al*., 2013).

1.2 Structure of lignocellulose

Lignocellulosic biomass consists mainly of three types of polymers, cellulose, hemicellulose and lignin making up 90% of its dry matter with the rest consisting of small amounts of minerals known as ash and extractives (Wyman, 1999; Sánchez, 2009; Balat, 2011). These constituents offer structural support and rigidity to plants. The compositions and proportions

of these constituents vary between plant species due to environmental and genetic influences (Balat, 2011).

1.2.1 Cellulose

Cellulose is the most abundant polysaccharide on Earth (Bayer *et al*., 1998; Zhang and Zhang 2013). Apart from constituting the bulk of plant cell walls (40-55%), cellulose is present in bacteria, fungi, algae as well as in animals (O'Sullivan, 1997; van Zyl *et al*., 2011). Cellulose exists in crystalline and non-crystalline (amorphous) forms with crystalline cellulose, also known as native cellulose, possessing two allomorphs: cellulose I and cellulose II (O'Sullivan, 1997; Bayer *et al*., 1998; Brown Jr, 1999).

Cellulose I is the most prevalent form found in nature and consists of a distinct number of glucan chains arranged in parallel to form a microfibril and is stabilized by hydrogen bonds (Figure 1.1) (Brown, 2003; Menon and Rao, 2012). Those parts of the microfibrils that are less ordered form the non-crystalline amorphous regions. The crystalline structure is relatively impermeable to enzymes and water making it difficult to break due to the orientation of the linkages and additional hydrogen bonding (Sukumaran *et al*., 2005; Balat, 2011). At the molecular level, cellulose is a homopolymer comprised of repeating glucose units (the basic unit of cellulose is cellobiose which are two β-1-4 linked glucose units) linked by β-1-4 glycosidic bonds to form the crystalline structure (Brown Jr, 2004).

Cellulose II is rare in nature and may only be found in some algae and bacteria and differ from cellulose I in that the glucan chains are oriented in an antiparallel fashion and tends to be more thermodynamically stable (Brown, 2003). Cellulose II may be prepared and studied by subjecting cellulose I to solubilisation in a solvent followed by precipitation in water, or by mercerization (sodium hydroxide treatment) (O'Sullivan, 1997).

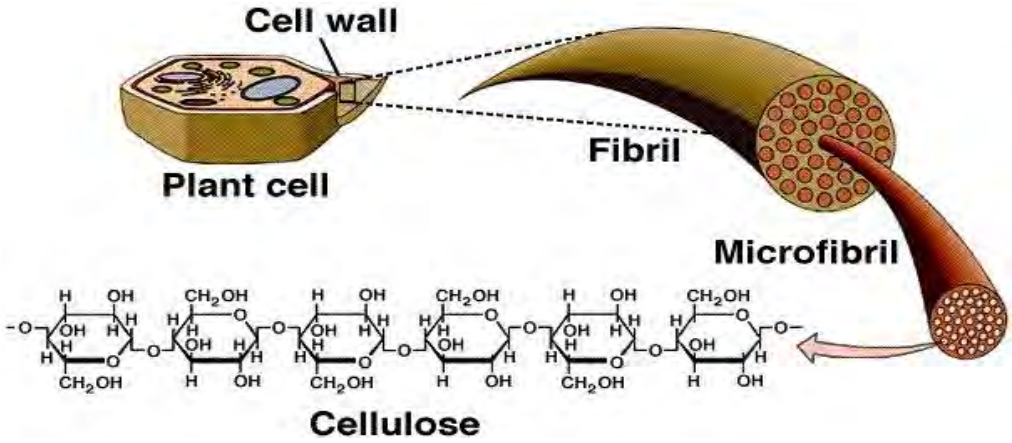


Figure 1.1: Cellulose structure and its' arrangement into microfibrils and fibrils within the plant cell wall (http://www.bio.miami.edu/dana/pix/cellulose_microfibrils.jpg)

Plant cell walls may be subdivided into primary and secondary walls with the secondary wall containing the major portion of cellulose (Menon and Rao, 2012). The cellulose content varies among plants types as is shown in Table 1.1 with waste residues such as cornstover and sugarcane bagasse having similar cellulose contents to that of corn and much more than wheat which therefore supports their use as a feedstock for biofuel production (Demirbras, 2005).

Table 1.1: Cellulose content of various plant biomass(Demirbas, 2005; Menon and Rao, 2012)

Material	Cellulose content (% dry weight)
Cornstover	35.1-39.5
Corn cob	32.3-45.6
Cotton	85-95
Grasses	25-40
Hardwood barks	22-40
Hardwoods	45±2
Softwood barks	18-38
Softwoods	42±2
Sugarcane bagasse	25-45
Wheat straw	37-41
Wheat bran	10.5-14.8

1.2.2 Hemicellulose

Hemicellulose is the second most abundant polysaccharide in nature. Hemicellulose is composed mainly of pentoses (xylose and arabinose) and hexoses (mannose, glucose and galactose) and has a lower molecular weight than that of cellulose (Brodeur *et al*., 2011). The basic components of hemicellulose include: xylan, mannan, galactoglucomannan, arabinan and arabinogalactan which are substituted with a variety of side chains (Shallom and Shoham, 2003; Sánchez, 2009). Hemicelluloses are attached to cellulose microfibrils and are generally found in the secondary cell wall of plants, however, they may also be present in the primary cell wall (Jeffries, 1994). The four main types of backbones in hemicelluloses are xylans (D-xylose), xyloglucan (β-1,4-D-glucopyranan), mannan (β-1,4-D-mannopyranose) and β-glucans (β-(1,3-1,4)-D-glucan) (Segato *et al*., 2014) with xylan being the major component of hemicellulose in plant cell walls.

Xylan is covalently and non-covalently linked to cellulose, lignin and other polysaccharides (Watanabe *et al*., 2008). Xylan is a complex and heterogeneous structure that consists of a β-1,4-linked D-xylose backbone commonly substituted with acetyl, uronyl and in some cases, arabinosyl side chains (Gupta *et al*., 2000; Lee *et al*., 2007; Watanabe *et al*., 2008). Softwoods and hardwoods have evolved separately and thus possess differing hemicellulose constituents, with softwood xylans being substituted with arabinofuranose units while xylans of hardwoods are α-1,2-D-glucuronic acids (Jeffries, 1994). Because of the complexity and diversity with regard to sidechains, the complete hydrolysis of this polysaccharide can only be accomplished by the concerted efforts of enzymes cleaving the xylan backbone as well as those capable of debranching and removing sidechains (refer to section 1.3.2).

1.2.3 Lignin

Lignin is a heterogeneous non-sugar based polymer composed of phenolpropane units linked via C-C and aryl-ether linkages to form an impenetrable matrix. The phenolpropane backbone is substituted with a number of hydroxyl groups, namely, hydroxyl, methoxyl and carbonyl groups (Balat, 2011). Lignin is present in plant cell walls and offers rigidity and structural support to aerial parts of the plant. It also confers impermeability and resistance against microbial attack and oxidative stress (Sánchez, 2009). This apparent recalcitrance that lignin confers to lignocellulosic material is one of the major obstacles to the production of bioethanol (Balat, 2011).

1.3 Enzymatic degradation of lignocellulose

1.3.1 Cellulases

The biological degradation of cellulose resulting in the release of glucose monomers (saccharification) is brought about by the action of biological catalysts belonging to the glycoside hydrolase family of enzymes (enzymes that degrade polysaccharides). These enzymes are collectively termed cellulases. Due to its crystalline structure, cellulose is particularly resistant to microbial degradation; hence, multiple cellulases are required for efficient cellulose hydrolysis. This cellulase system is comprised of three types of cellulases, namely, endoglucanase (endo-1,4-β-D-glucanase (EG), EC 3.2.1.4), cellobiohydrolase (CBH)/exoglucanase (exo-1,4-β-D-glucanase, EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21), which either function independently or form complexes known as cellulosomes (Bhat and Bhat, 1997; Bayer *et al*., 1998; Ye *et al*., 2001; Zhang and Zhang, 2013). EGs act by randomly cleaving internal β 1-4 glycosidic linkages of the cellulose chain producing cello-oligosaccharides, while CBH cleaves cellobiose units from the non-reducing ends of the cellulose chains. The cellobiose that is generated by the concerted efforts of EGs and CBHs

is then cleaved by β-glucosidase to yield glucose and eliminate cellobiose inhibition (Bhat and Bhat, 1997; Ye *et al*., 2001; Li *et al*., 2003; Zhang *et al*., 2006).

These three enzymes work in synergy to bring about efficient hydrolysis of cellulose. Most of these enzymes are relatively ineffective in breaking down crystalline cellulose individually (Singh and Hayashi, 1995). Eriksson and Wood (1985) demonstrated that the individual EGs, CBHs and β-glucosidases are unable to degrade crystalline cellulose while a cocktail of these three enzymes brings about extensive hydrolysis. There are two forms of EGs and exoglucanases that are required for the hydrolysis of crystalline cellulose; those that have a relatively low affinity and do not adsorb tightly onto cellulose and those that have a high affinity and tightly adsorb onto cellulose (Goyal *et al*., 1991; Rabinovich *et al*., 2002). The degradation of cellulose begins with the attack on the outer amorphous regions of a cellulosic substrate by low affinity EGs which wedge into defects in the cellulose fibres and open-up sites which are then attacked by high affinity EGs and CBHs (which is the major component of the fungal cellulase system). The resulting glucose dimers (cellobiose) that are released are cleaved by β-glucosidases to yield glucose monomers (Rabinovich *et al*., 2002; Sánchez, 2009).

1.3.1.1 Modular nature of cellulases

Cellulases are modular multidomain enzymes that comprise at least three independently folding structural units presenting different functionalities, which are required to cope with the structural heterogeneity of the substrate (Bhat and Bhat, 1997; Bayer *et al*., 1998; Rabinovich *et al*., 2002). These three elements, referred to as modules or domains include a catalytic domain (CD), carbohydrate binding module (CBM), and an interdomain linker (Rabinovich *et al*., 2002). The CD is that domain that contains the catalytic site or active site, the CBM plays a role is recognition of the substrate and affinity of the enzyme to the

substrate while the linker region separates the CD and CBM and allows them to function autonomously (Bayer *et al*., 1998; Rabinovich *et al*., 2002).

1.3.1.2 Catalytic mechanism of action

The enzymatic hydrolysis of glycosidic bonds takes place *via* an acid-base mechanism involving two residues, i.e. a proton donor and a nucleophile (Davies and Henrissat, 1995). Glycosyl hydrolases, and by extension, cellulases follow one of two catalytic mechanisms, namely, either a retention (double-displacement mechanism) or inversion mechanism (single-displacement mechanism) (Davies and Henrissat, 1995; Bhat and Bhat, 1997; Zhang and Zhang, 2013). Hence, cellulose hydrolysis may result in either the overall retention (displaying the same stereochemistry as the substrate) or inversion (displaying the opposite stereochemistry to the substrate) of the anomeric carbons' configuration of the resulting glucose monomers (Bayer *et al*., 1998). The double-displacement mechanism, as illustrated in Figure 1.2, proceeds with the protonation of the substrate (at the glycosidic oxygen) by the acid catalyst which is subsequently attacked by the nucleophile resulting in the formation of an α-glycosyl enzyme intermediate which is then hydrolysed by water to yield a product with the same stereochemistry as the substrate (Davies and Henrissat, 1995; Bhat and Bhat, 1997; Bayer *et al*., 1998).

The single-displacement mechanism (Figure 1.3) involves a single step in which the glycosidic oxygen is protonated and subsequently cleaved by a nucleophilic water molecule (formed by the action of the residue acting as the base) resulting in the formation of a product with opposite stereochemistry to the substrate (Davies and Henrissat, 1995; Bhat and Bhat, 1997; Bayer *et al*., 1998).

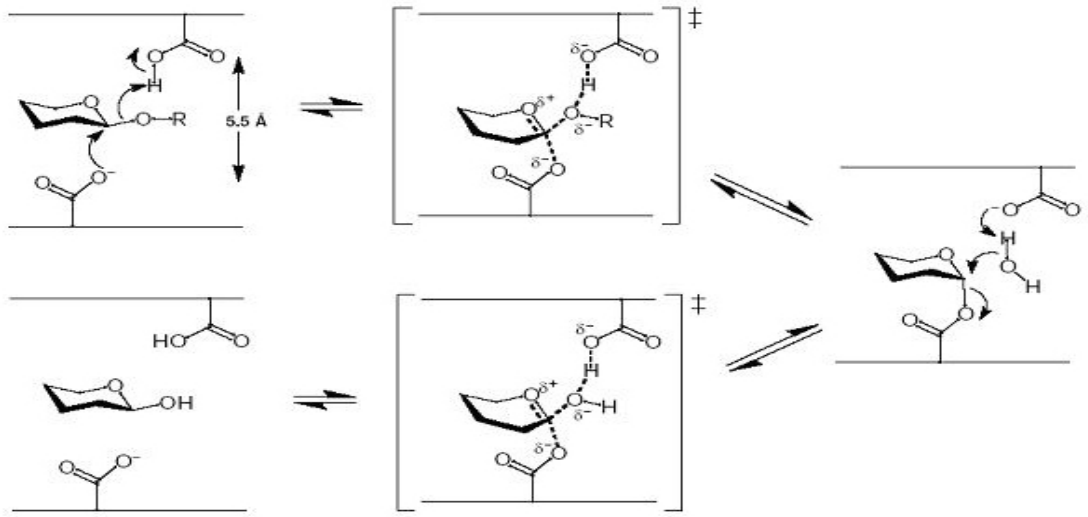


Figure 1.2: Double-displacement mechanism (http://www.cazy.org/fam/ghf INV RET.html#inv)

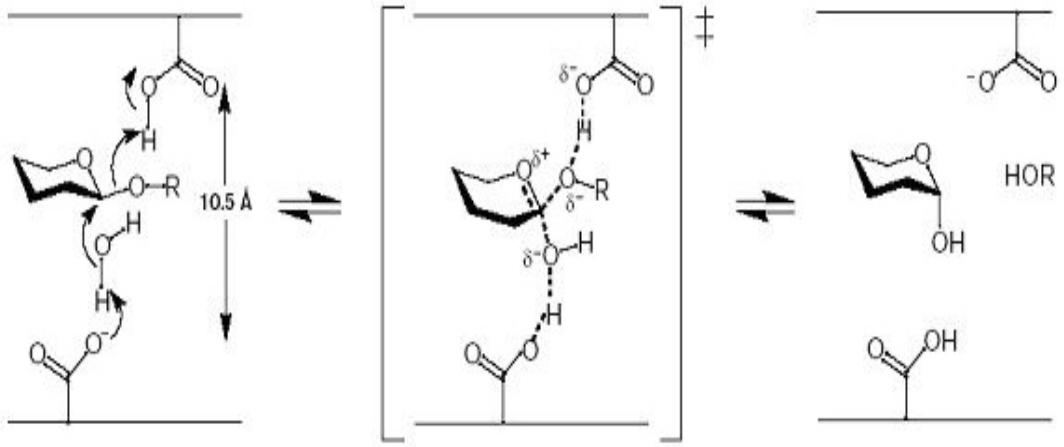


Figure 1.3: Single-displacement mechanism (http://www.cazy.org/fam/ghf INV RET.html#inv)

The structural differences between CBH and EG active sites provide clues that highlight differences in functionality, particularly with regard to processivity of cellulose (Payne *et al*., 2015). CBHs tend to remain attached to cellulose chain ends and continuously removes cellobiose units while EGs move randomly along the cellulose backbone, cleaving internal glycosidic bonds then detaches from the substrate and moves to a new location. The CBH active site is characteristic of a tunnel-shaped conformation with the roof of the tunnel formed by long flexible loops (Divn *et al*., 1994; Munoz *et al*., 2001; Grassick *et al*., 2004;

Parkkinen *et al*., 2008). EGs on the other hand lack these loops thus their active site topology is that of an open cleft (Kleywegt *et al*., 1997; Poidevin *et al*., 2013). CBHs exhibit specificity to either the reducing or non-reducing ends of cellulose. The open cleft of EGs are thought to allow random binding and cleavage of cellulose chains and in so doing increase the concentration of chain ends (Zhang and Lynd, 2004; Poidevin *et al*., 2013). These cellulose chain ends are thread into the tunnel active site of the CBHs to form at catalytically active complex and subsequenctly glycosidic bonds are cleaved and cellobiose units are released (Ghattyvenkatakrishna *et al*., 2013; Poidevin *et al*., 2013).

It has also been suggested that in addition to the endo- and exo-acting enzymes, physical disruption of insoluable cellulose is also required for the hydrolysis of cellulose as it renders crystalline cellulose more accessible to cellulases (Zhang and Lynd, 2004). This increase in cellulose activity may be brought about by a number of factors such as the CBMs (Din *et al*., 1991; 1994), the CDs of some CBHs (Woodward *et al*., 1992; Lee *et al*., 1996) and proteins called swollenin (Saloheimo *et al*., 2002). The cellular component involved in the disruption as well as their importance is still not entirely clear (Zhang and Lynd, 2004).

1.3.1.3 Classification of glycoside hydrolases

Glycoside hydrolases (GH) are classified according to their structural properties by comparison of the amino acid sequence similarities of their CDs rather than the enzyme's substrate specificity, as was previously the traditional nomenclature of GHs set by the International Union of Biochemistry (IUB) (Henrissat, 1991; Henrissat and Bairoch, 1993; Davies and Henrissat, 1995; Mathew *et al*., 2008). The CAZy Database is an excellent source of information on enzymes involved in the building, breakdown and modification of complex carbohydrates and glycoconjugates (Cantarel *et al*., 2009). The database entries are classified according to the new criteria, based on both their amino acid sequence and three-dimensional

structure which results in correlations with enzyme mechanisms and protein folds rather than enzyme specificity (Davies and Henrissat, 1995; Cantarel *et al*., 2009). This allows one to distinguish between enzymes that possess different functions yet belong to the same family, or between enzymes possessing the same function yet belonging to different families (Busk and Lange, 2013; Lombard *et al*., 2014). With respect to the current method of classification, cellulases are classed into the following GH families: EGs are found in families 5-8, 12, 16, 44, 45, 48, 51, 64, 71, 74, 81, 87, 124 and 128; CBHs are found in families 5-7 and 48 and β-glucosidases are found in familes 1, 3, 4, 17, 30 and 116 (Mathew *et al*., 2008; Wang *et al*., 2010; Li *et al*., 2011; Juturu and Wu, 2014; www.cazy.org).

1.3.1.4 Occurrence in nature

Cellulolytic enzymes are produced by a variety of aerobic and anaerobic, mesophilic and thermophilic bacteria and fungi. However, more emphasis has been placed on fungal cellulases as fungi are able to produce higher amounts of cellulases and are less complex than bacterial cellulases in that they are not (however several anaerobic fungi may produce cellulosomes) arranged in cellulosomes and thus allowing for easier cloning and recombinant production (Bayer *et al*., 1998, Maki *et al*., 2009). Examples of bacterial cellulase producers include: *Cellulomonas*, *Cytophaga*, *Cellovibrio*, *Bacillus* spp. and *Thermomonospora* sp. (Bhat and Bhat, 1997; Sukumaran *et al*., 2005; Maki *et al*., 2009). Fungal cellulase producers include: *Phanerochaete chrysosporium*, *Trichoderma* spp., *Penicillium* spp., *Humicola* spp. and *Aspergillus* spp. (Sukumaran *et al*., 2005).

The characteristics and set of cellulases differ depending on the various microbial producers. The cellulase system of the soft rot fungus *Trichoderma reesei* has received the most attention and is the best characterized (Reczey *et al*., 1996, Payne *et al*., 2015). *T. reesei* has two CBHs, CBH I and CBH II. CBH I is the major cellulase enzyme making up 60% of

secreted proteins while CBH II makes up 20% of the secreted proteins. CBH I range in molecular weight from 42–72 kDa with iso-electric points (pI) from 3.5–4.2, while CBH II ranges in molecular weight from 50–58 kDa and pI from 5.0–6.3 (Goyal *et al*., 1991). The cellulase system of *T. reesei* consists of three EGs, namely, EG I, EG II and EG III, which also occur in multiple forms. The characteristics of these enzymes are as follows: EG I: 43–55 kDa and pI of 4.7; EG II: 49.8 kDa and pI of 5.5; EG III: 20–23.5 kDa and pI of 7.5 or 7.7 (Goyal *et al*., 1991; Singh and Hayashi, 1995). The least studied enzyme in cellulose degradation is β-glucosidase, the *T. reesei* cellulase system expresses two β-glucosidase with molecular weights of 50 and 35 kDa and pIs of 6.1 and 5.7, respectively (Singh and Hayashi, 1995).

The majority of industrial cellulases are of bacterial and fungal origin (Huy *et al*., 2015). Fungal cellulases, such as those produced by *Aspergillus* spp., *Trichoderma* spp. and *Penicillium* spp. are mesophilic and not highly thermostable (Lynd *et al*., 2002; Zhao *et al*., 2011). An EG produced by *A. niger* displayed optimum activity at pH 4 and 50°C and only retained 60% activity at 55°C after 30 minutes (Quay *et al*., 2011). The EG from *T. viride* AS 3.3711 displayed optimal activity at pH 5 and 55°C with activity dropping drastically after incubation at 60°C (Fan *et al*., 2012). Staying true to the trend observed, the EG from *P. pinophilum* MS 20, also displayed a mildly acidic pH optimum at pH 5 and an optimum temperature at 50°C and did not show desirable stability at higher temperatures (Pol *et al*., 2012). The industrial applications that make use of such enzymes require that the enzyme be more hardy and stable under specific reaction conditions (Pol *et al*., 2012; Zhao *et al*., 2011). Thus, there is a need for the discovery of novel cellulases that possess extremophilic physicochemical properties for exploitation in industry.

1.3.2 Hemicellulase

Due to its structural heterogeneity, the complete degradation of hemicellulose requires the concerted efforts of a variety of enzymes, many more than what is required for the degradation of cellulose. Xylanases, together with auxiliary enzymes (those specific for the cleavage of side chains) catalyse the hydrolysis of xylan and hemicellulose. The enzymes involved in the degradation of the polysaccharide backbones and the release of their respective side chains are listed in Table 1.2.

Table 1.2: Hemicellulases and their respective substrates

Enzyme	Substrate	EC Number	Reference
Endo-β-1,4-xylanase	β-1,4-xylan	3.2.1.8	Smith *et al*. (1991)
Exo-β-1,4-xylosidase	β-1,4-xylooligomers	3.2.1.37	Smith *et al*. (1991)
Endo-β-1,4-mannanase	β-1,4-mannan	3.2.1.78	Shallom and Shoham (2003)
Exo-β-1,4-manosidase	β-1,4-mannooligomers	3.2.1.25	Shallom and Shoham (2003)
α-L-arabinofuranosidase	α-L-arabinofuranoside residues	3.2.1.55	Koseki *et al*. (2009)
Endo-α-1,5-arabinanase	α-1,5-arabinan	3.2.1.99	Sánchez, (2009)
α-Glucuronidase	4-*O*-methyl-D-glucuronic acid sidechains	3.2.1.131	Chong *et al*. (2009)
α-Galactosidase	Melibiose	3.2.1.22	Shallom and Shoham (2003)
β-Glucosidase	β-Glucopyranose Mannopyranose	3.2.1.21	Juturu and Wu (2014)
Endo-galactanase	β-1,4-galactan	3.2.1.89	Sánchez, (2009)
Acetyl xylan esterase	2- or 3-*O*-acetyl xylan	3.1.1.72	Maheshwari *et al*. (2000)
Acetyl mannan esterase	2- or 3-*O*-acetyl mannan	3.1.1.6	van den Brink and de Vries (2011)
Ferulic acid esterase	Feruloyl-polysaccharide	3.1.1.73	Koseki *et al*. (2009)

1.3.3 Ligninases

In nature, wood decay and the biodegradation of lignin in particular is largely facilitated by basidiomycetes white rot fungi as they secrete extracellular enzymes collectively termed ligninases (Hammel *et al*., 1993; Dashtban *et al*., 2010). Lignin degradation is an oxidative process and the enzymes involved in this process may be classified as either phenol oxidases or heme peroxidases (Sánchez, 2009; Dashtban *et al*., 2010). The phenol oxidase that is involved in lignin degradation is laccase (EC 1.10.3.2) while the heme peroxidases include lignin peroxidase (LiP) (EC 1.11.1.14), manganese peroxidase (MnP) (EC 1.11.1.13) and versatile peroxidase (VP) (EC 1.11.1.16) (Dashtban *et al*., 2010; Iqbal *et al*., 2011).

Lignin peroxidases degrade up to 90% of lignin (which comprises non-phenolic lignin units) *via* hydrogen peroxide-dependant (H_2O_2) oxidation reactions with the formation of intermediate radicals. These free radicals are involved in non-enzymatic reactions to bring about side chain cleavage, demethylation, aromatic ring cleavage and C-4-ether breakdown (Sánchez, 2009; Dashtban *et al*., 2010). Manganese peroxidases oxidise Mn^{2+} to Mn^{3+}, which acts as a low molecular weight diffusible redox-mediator when chelated to organic acids (such as oxalate and malonate) and brings about the oxidation of phenolic and non-phenolic lignin units. Versatile peroxidases are termed as such, due to their ability to oxidise substrates typical of LiP and MnP, those being phenolic and non-phenolic aromatic compounds and Mn^{3+}, respectively (Sánchez, 2009; Dashtban *et al*., 2010).

1.4 Industrial applications of cellulolytic enzymes

Microbial cellulases find applications or potential applications in various industrial sectors which include: paper and pulp, textile, detergent, food and animal feed, brewing and biofuel industries as well as for research purposes (Bhat, 2000; Sukumaran *et al*., 2005; Kuhad *et al*., 2011). One of the most important applications being investigated is the degradation of

lignocellulosic material for the production of biofuels, typically ethanol, where cellulases are required to convert cellulosic materials to glucose and other fermentable sugars (Sukumaran *et al*., 2005; Wackett, 2008; Zhang and Zhang, 2013).

Thermostable cellulases are gaining widespread interest in industry. With regard to the hydrolysis of lignocellulosic biomass, thermostable cellulases have higher specific activity, higher stability and allow for increased flexibility of process configurations (Viikari *et al*., 2007). The hydrolysis of lignocellulosic biomass at elevated temperatures provides additional benefits such as deceased risks of contamination and reduced energy costs for cooling (Yang *et al*., 2010). It is also a known fact that cellulose swells at higher temperatures, thus becoming more amenable to enzymatic breakdown (Li *et al*., 2011). Examples of thermostable EGs and their properties produced by various microorganisms are tabulated below (Table 1.3).

1.4.1 Fuel ethanol

Ethanol is a renewable energy source that is produced through the fermentation of hexose and pentose sugars that can be used to supplement petroleum fuels. Bioethanol is currently being produced from starch and sucrose producing crop, however, there is much debate over the use of animal and human food commodities as a feedstock for the production of bioethanol (Zheng *et al*., 2009). This sparked the interest in the use of lignocellulosic biomass such as agricultural and forestry waste residues as a feedstock (Hahn-Hägerdal *et al*., 2006; Umamaheswari *et al*., 2010). The sources that show most promise for bioethanol include rice straw, wheat straw, wood and sugarcane bagasse. Fuel ethanol can be produced from lignocellulosic biomass *via* a four step process which includes a pretreatment process which breaks down the lignocellulosic matrix, enzymatic hydrolysis of cellulose into glucose

Table 1.3: Properties of thermostable endoglucanases

Organism	Molecular Weight (kDa)	pI	pH Optimum	Temperature Optimum (°C)	Reference
Clostridium stercorarium	100	NR	6.0 – 6.5	90	Viikari *et al*. (2007)
Bacillus sp. C1AC5507	55	NR	7.0	70	Padilha *et al*. (2015)
Bacillus subtilis	NR	NR	5.0 – 6.5	65 – 70	Haki and Rakshit (2003)
Thermoascus aurantiacus	34	3.7	4.0 – 4.4	70 – 80	Parry *et al*. (2002)
Bacillus sp. KSM-S237	86	NR	8.6 – 9.0	45	Viikari *et al*. (2007)
Thermotoga neapoltana (Endocellulase A)	NR	NR	6.0	95	Haki and Rakshit (2003)
Paenibacillus barcinonensis MG7	58.6	NR	6.0	65	Asha *et al*. (2012)
Daldinia eschscholzii (Ehrenb.:Fr.) Rehm	46.4	4.9	6.0	70	Karnchanatat *et al*. (2008)
Bacillus strain M-9	54	NR	5	60	Bajaj *et al*. (2009)
Talaromyces emersonii (EG I)	35	3.19	5.5 – 5.8	75 - 80	Maheshwari *et al*. (2000)

NR: not reported

monomers by cellulolytic enzymes, the fermentation of sugars into ethanol and finally distillation to purify the ethanol (Margeot *et al*., 2009). The major hindrance to the use of lignocellulosic biomass for fuel ethanol production is its recalcitrance and the requirement of pretreatment processes to render cellulose accessible to hydrolytic enzymes (Zheng *et al*., 2009).

1.4.1.1 Pretreatment strategies

Pretreatment of lignocellulosic biomass is required to render the biomass more accessible to hydrolytic enzymes by increasing the surface area, cellulose decrystallisation and partial depolymerisation, hemicellulose and lignin solubilisation and modification of the lignin structure (Margeot *et al*., 2009). It has been reported that sugar yields can be improved by more than 90% when biomass such as grasses, corn and wood are subjected to pretreatment (Brodeur *et al*., 2011). An effective pretreatment should meet the following criteria: it should avoid size reduction and avoid the degradation or loss of carbohydrates, limit the formation byproducts and inhibitory compounds, have a reduced energy output and be cost effective (Sun and Cheng, 2002; Zheng *et al*., 2009). Pretreatment methods encompass physical (e.g. uncatalyzed steam explosion, liquid hot water pretreatment and mechanical comminution), chemical (catalyzed steam-explosion, acid and alkaline pretreatments and organosolv), biological (wood degrading microorganisms) or a combination of these methods (Martin *et al*., 2006; Balat, 2011).

1.4.1.1.1 Physical pretreatment

Physical pretreatments do not make use of any chemical agents but employ methods that bring about size reduction of biomass by mechanical means such as chipping, grinding and milling or a combination thereof to yield sizes of a few centimetres to 1-3 mm (Sun and Cheng, 2002). The reduced particle size provides a larger surface area for the enzymes active

site during the hydrolysis step to penetrate the fibres (Hamelinck *et al*., 2005). The final particle size is dependant the amount of power one is willing to utilise as fine particles of micrometer size require extremely high amounts which is not economically feasible or acceptable from an engineering standpoint (Zhao *et al*., 2011). Thus more common physical pretreatment methods are uncatalysed steam explosion and liquid hot water pretreatment (LHW) (Zheng *et al*., 2009).

Uncatalyzed steam-explosion (also known as steam explosion or autocatalysis) is one of the most common methods employed and is considered the most cost effective method when utilizing agricultural residues (Balat, 2011). Steam explosion is characterized by rapidly heating the lignocellulosic biomass with steam at temperatures of 160-270°C for a few seconds or minutes and then terminating by decompression and cooling. During pretreatment, the hemicellulose is hydrolysed by organic acids that are formed from acetyl and other functional groups that are released from the biomass. At these high temperatures, water displays acid properties that aid in hemicellulose hydrolysis. Lignin is also transformed and disrupted which decreases recalcitrance (Martin *et al*., 2006; Zheng *et al*., 2009; Balat, 2011).

Liquid hot water pretreatment involves the cooking of the biomass in hot water at temperatures between 200-230°C under high pressure for up to 15 minutes (Mosier *et al*., 2005). Much of the biomass is dissolved in the process and as in the case of steam explosion, the water acts as an acid and hydrolyses hemicellulose to release acids and sugars (Zheng *et al*., 2009).

1.4.1.1.2 Chemical pretreatment

Chemical pretreatment brings about delignification of lignocellulosic biomass after exposure to chemical agents such as acids, alkaline solutions, organic acids, pH-controlled liquid hot water and ionic liquids (Balat, 2011; Menon and Rao, 2012). Acid pretreatments make use of

nitric, sulphuric or hydrochloric acids to solubilise hemicellulose and precipitate solubilised lignin from lignocellulosic biomass thus improving the biodegradability of cellulose (Hendriks and Zeeman, 2009; Balat, 2011). Acid pretreatment strategies may utilize concentrated or dilute acids and may operate under low or high temperature, respectively, with the use of dilute acid and high temperature being the more favourable strategy as the use of concentrated acids is toxic, hazardous and corrosive (Zheng *et al*., 2009; Alvira *et al*., 2010). Dilute acid pretreatment may function as a means of pretreatment of lignocellulosic biomass for enzymatic hydrolysis or as a means of actual hydrolysis of carbohydrates into fermentable sugars (Taherzadeh and Karimi, 2008). Dilute acid pretreatment has received much attention and has been studied for pretreating a variety of lignocellulosic biomass (Alvira *et al*., 2010). Dilute sulphuric acid is the most commonly used acid and has been used commercially for the pretreatment of biomass such as corn stover, poplar and spruce (softwood) (Brodeur *et al*., 2011; Menon and Rao, 2012). Pretreatment with dilute sulphuric acid brings about hydrolysis of hemicellulose into its monomeric sugars, exposure of cellulose for enzymatic hydrolysis and the solubilisation of heavy metals which may contaminate the feedstock (Balat, 2011). However, there are still some drawbacks to the use of dilute acids which include: the requirement for corrosive resistant materials which may be expensive; neutralisation of prehydrolysates is required prior to fermentation; the disposal of neutralisation salts; and the formation of fermentation inhibitors such as carboxylic acids, furans and phenolic compounds (Taherzadeh and Karimi, 2008; Balat, 2011).

Alkaline pretreatment refers to the use of alkaline solutions such as sodium hydroxide, potassium hydroxide, calcium hydroxide (lime), ammonium hydroxide, aqueous ammonia and sodium hydroxide in combination with hydrogen peroxide and others (Taherzadeh and Karimi, 2008; Zheng *et al*., 2009; Menon and Rao, 2012). The mechanism of action of alkali solutions involves: the degradation of ester and glycosidic side chains which brings about the

structural modification of lignin, swelling of cellulose which makes it more accessible to enzymes; decreasing the degree of polymerisation and crystallinity of cellulose; and the partial solubilisation of hemicellulose (Alvira *et al*., 2010). Alkali pretreatment can be performed at lower temperatures to that of acid pretreatments. Pretreatment can be performed at room temperature, however, the time of exposure can range from seconds to days (Taherzadeh and Karimi, 2008; Alvira *et al*., 2010). Sodium hydroxide is able to disrupt the lignin structure and increase accessibility of enzymes to hemicellulose and cellulose (Menon and Rao, 2012). The most cost effective alkaline pretreatment is lime, when compared to other popular alkaline pretreatments such as sodium and potassium hydroxide (Balat, 2011). Lime also boasts fewer safetly requirements and is also recoverable from the hydrolysate by reaction with carbon dioxide to form calcium carbonate which can be reconverted into lime (Brodeur *et al*., 2011). Apart from the removal of lignin, lime also cleaves acetyl groups from hemicellulose and enhances cellulose degradability by reducing steric hinderances to the enzymes (Alvira *et al*., 2010). The removal of hemicellulose has a postive effect on cellulose degradation, however, there is often loss of fermentable sugars which is not ideal for ethanol production and the solubilised lignin often has an inhibitory effect which also makes alkaline pretreatment not very attractive for ethanol production (Hendriks and Zeeman, 2009).

1.4.1.1.3 Biological pretreatment

Biological pretreatments employ wood degrading microorganisms such as bacteria, brown-, white, and soft-rot fungi to solubilise lignin (Hamelinck *et al*., 2005). Fungi have distinct characteristics with regard to lignocellulosic biomass degradation. White-rot fungi are the most effective biological pretreatment as they are able to degrade lignin more aggressively that brown- and soft-rot fungi which mainly attack cellulose (Menon and Rao, 2012). *Phlebia* sp. MG-60, a marine fungus, proved to be an excellent lignin degrader by being able to

efficiently delignify sugarcane bagasse while leaving the cellulose intact when the culture was supplemented with Kirk medium (Li *et al*., 2002). Pretreatment of rice straw with *Pleurotus ostreatus* resulted in selective degradation of lignin and left the substrate more susceptible to enzymatic hydrolysis of cellulose (Taniguchi *et al*., 2005). The lignin degrading ability of *P. chrysosporium* has been well documented and it was reported by Jian and colleagues (2008) that the fungus was capable of the removal of lignin and hemicellulose from cotton stalks to facilitate fuel ethanol production.

Biological pretreatment appears to be a promising technology due to: its low energy requirements, mild environmental conditions and environmentally friendly nature (Balat, 2011). However, its disadvantages pose a stark contrast to its advantages; the treatment rate is very slow and requires careful control of growth conditions, a large amount of space is required for the treatment and most lignolytic microorganisms metabolise not only lignin but hemicellulose and cellulose as well (Taherzadeh and Karimi, 2008; Zheng *et al*., 2009; Balat, 2011; Menon and Rao, 2012). These limitations of using microorganisms themselves may be overcome by the use of only their enzymes in combination with other pretreatments discussed above.

1.4.1.2 Cellulose hydrolysis and fermentation

The cellulose obtain from pretreated lignocellulosic biomass thereafter needs to be hydrolysed into glucose using acids or enzymes (Sun and Cheng, 2002; Cardona *et al*., 2010). Cellulose hydrolysis using dilute or concentrated acids are performed under extreme temperatures resulting in high glucose yields, however, these conditions result in the formation of undesirable products such a furfural which inhibits fermentation (Hamelinck *et al*., 2005). In light of this, cellulose hydrolysis is currently being carried out with the aid of

microbial cellulases as the subsequent fermentation yields better results as they do not produce inhibitory compounds (Sánchez and Cardona, 2008).

The hexose and pentose sugars liberated as a result of pretreatment and hydrolysis are thereafter fermented by bacteria, yeast or fungi to produce bioethanol under anaerobic conditions. *Saccharomyces cerevisiae* is the most ultilized organism for the fermentation of carbohydrates liberated from lignocellulosic substrates. However, this yeast is only capable of fermenting hexose sugars (Cardona *et al*., 2010). Therefore efforts have been made to genetically improve microorganisms to be able to ferment both hexose and pentose sugars (Fernandes and Murray, 2010).

Cellulose hydrolysis and fermentation may be carried out in different units in a configuration termed separate hydrolysis and fermentation (SHF) (Hamelinck *et al*., 2005). A defining feature of this method of fermentation is that cellulose hydrolysis and the subsequent fermentation of the carbohydrates can be performed at their respective optimal conditions (Sánchez and Cardona, 2008). The alternate method of fermentation, simultaneous saccharification and fermentation (SSF) where cellulolytic enzymes and microorganisms are added into the same processing unit such that once glucose is liberated from cellulose, they are immediately metabolized by the microorganism and converted into ethanol (Cardona *et al*., 2010). SSF tends to be a more attractive method of SHF as SSF results in a higher ethanol yield as well as less energy demanding (Sánchez and Cardona, 2008).

1.5 Heterologous expression of cellulases

There is an increasing demand for the discovery of novel cellulases with varying conditions of optimal activity to meet the requirements of a variety of industrial applications (Liu *et al*., 2011). Cellulases are being employed in the textile, paper and pulp, food and animal feed

industries with an increased focus on the use of EGs for the enzymatic hydrolysis of lignocellulosic biomass for the production of bioethanol (Sukumaran *et al*., 2005).

Although there are several microorganisms capable of cellulase production, these native cellulase producers are unable to produce copious amounts to satisfy the industrial demand as this would impose a metabolic burden on the organism (Mathew *et al*., 2008; Vinuselvi and Lee, 2012). Another challenge is that the cost of commercial production of cellulolytic enzymes from native producers for bioconversion processes remains too high, thus genetic engineering approaches are being explored to reduce the cost of cellulase production as well as to obtain greater yields and higher titres of the enzymes of interest (Yang *et al*., 2010). A host of cellulolytic enzymes have been cloned and heterologously expressed for the purposes of enzymological studies as well as for the development of consolidated bioprocessing (CBP) organisms. CBP organisms are those organisms that are capable of hydrolysing biomass into fermentable sugars as well as being able to ferment the resulting sugars into ethanol (Lynd *et al*., 2005; Xu *et al*., 2009). This is accomplished by introducing cellulolytic genes into organisms with high product (ethanol) yields, thus allowing them to hydrolyse cellulose or by engineering cellulolytic organisms to improve product-related properties and yields (la Grange *et al*., 2010). The following organisms have been genetically engineered to heterologously express different cellulase components:

1.5.1 *Escherichia coli*

E. coli serves as an ideal host for the heterologous expression of protein and genetic engineering as it is a well studied microorganism with a host genetic tools available (Vinuselvi and Lee, 2012). *E. coli* is able to assimilate all sugar monomers that are present in lignocellulosic biomass, however, it is unable to hydrolyse cellulose nor is it able to produce ethanol in considerable amounts (la Grange *et al*., 2010). This, along with *E. coli*'s limited

ability to secrete proteins into the extracellular medium precludes its use as a CBP microorganism (la Grange *et al*., 2010; Vinuselvi and Lee, 2012). This, however, does not prevent the use of *E. coli* as a heterologous expression host for the puposes of studying the enzymatic properties of cellulases or for its overexpression. A number of bacterial and fungal cellulases have been cloned, expressed and characterized through the use of *E. coli*. For example, the production of a thermostable EG (CelDR) from a *Bacillus subtilus* strain DR, was enhanced three fold in *E. coli* BL21 (DE3). The enzyme displayed optimum activity at 50°C and retained 70% of its activity after incubation at 75°C for 30 minutes (Li *et al*., 2008). This study also provided evidence through sequencing data, that supports the notion that amino acids with strong hydrophobic substituents increase the tolerance of cellulases and related enzymes to high temperatures. Fifteen novel fungal EGs were cloned and analysed with the use of *E. coli* XL1Blue and *A. niger* var. *awamori* (Goedegebuur *et al*., 2002). Sequence analysis showed that family 12 EGs can be subdivided into four subfamilies: fungal group I (12-1), fungal group II (12-2), *Streptomyces* group (12-3) and *Thermophiles* group (12-4).

1.5.2 *Saccharomyces cerevisiae*

S. cerevisiae is an efficient industrial ethanol producer, capable of growing at a low pH, osmo- and ethanol-tolerant, naturally robust, is generally regarded as safe (GRAS) and has well-developed genetic manipulation tools making it an attractive host for heterologous expression studies as well as for development as a CBP organism (Den Haan *et al*., 2007a; la Grange *et al*., 2010). A number of cellulase encoding genes from a range of bacteria, filamentous fungi and plants have been cloned and expressed in *S. cerevisiae* (Lynd *et al*., 2002). Two genes encoding an EG (*egI*) and a β-glucosidase (*bgl1*) from *T. reesei* and *Saccharomycopsis fibuligera*, respectively, were cloned and expressed in *S. cerevisiae* making it the first recombinant strain capable of growth on pure amorphous cellulose and

conversion of cellulose to ethanol in a single step (Den Haan *et al*., 2007a). Four fungal CBH genes (*cbh1* and *cbh2* from *T. reesei*, *cbhB* from *A. niger* and *cbh1-4* from *P. chrysosporium*) were also successfully expressed in *S. cerevisiae* with similar specific activity to that of the native enzymes. However, secretion of the heterologously expressed CBHs was relatively low (Den Haan *et al*., 2007b). A thermostable EG gene, *egI*, from *Thermoascus aurantiacus* was expressed in *S. cerevisiae* for the purposes of investigating its properties. It was found to be optimally active at 70°C and pH 6, while remaining stable for 2 hours at 70°C and between pH 3 and 10 (Hong *et al*., 2003).

1.5.3 *Pichia pastoris* and other potential hosts

Both the *E. coli* and *S. cerevisiae* expression systems have been used extensively in the reaserch arena, however, they both possess disadvantages that either limit or preclude their use as an efficient expression host or CBP organism. With regard to *E. coli*, high cell densities result in the accumulation of acetate which is toxic to the cells (however, this can be alleviated by controlling oxygen levels), proteins may inactive, insoluable and require refolding as *E. coli* is unable to produce proteins with disulfide bonds as well as those that require glycosylation (Demain and Vaishnav, 2009). *S. cerevisiae* may be able to succeed where *E. coli* fails, however, expression levels and the activity of the recombinant enzymes are generally lower than that of the native enzyme (Huang *et al*., 2010; Rosano and Ceccarelli, 2014). With this in mind, researchers have taken to using alternate expression hosts, such as filamentous fungi, yeasts other than *S. cerevisiae* as well as prokaryotes other than *E. coli*, making use of their unique metabolic pathways for the proper folding and post-translational modifications as well as achieving high levels of secretion (Table 1.4).

Of particular interest is *P. pastoris*, recently reclassified as *Komagataella pastoris*; a methylotrophic yeast that has been used extensively as a host for heterologous protein production (Ahmad *et al*., 2014). *P. pastoris*, like *S. cerevisiae*, is relatively easy to

genetically manipulate, and compared to mammalian cells, does not require complex growth media or culture conditions and can be grown to high cell densities (Macauley-Patrick *et al*., 2005).

Table 1.4: Examples of cellulase genes expressed in alternate hosts

Donor	Gene/Enzyme	Host	Reference
Acremonium thermophilum	*cel7a*	*Trichoderma reesei*	Li *et al*. (2011)
Chaetomium thermophilum	*cbh3*	*Pichia pastoris*	
Humicola grisea	*egl2, egl3, egl4*	*Aspergillus oryzae*	
Humicola insolens	*avi2*	*Humicola insolens*	
Trichoderma reesei	Endoglucanase II	*Kluyveromyces maxianus*	Van Zyl *et al*. (2011)
Aspergillus aculeatus	β-glucosidase		
Erwinia chrysanthemi	*cel5Z* and *cel8Y* (endoglucanases)	*Klebsiella oxytoca*	
Cellulomonas uda	Endoglucanase	*Zymomonas mobilis*	Singh and Hayashi (1995)

Expression in *P. pastoris* has distinct advantages over the use of *S. cerevisiae*. It is able to express 10- to 100-fold higher heterologous proteins and does not hyperglycosylate proteins as *S. cerevisiae* does (*Pichia* Expression Kit Manual; Lindenmuth and McDonald, 2011). A major advantage is that *P. pastoris* secretes very low levels of native proteins making it fairly easy to purify heterologously expressed proteins if they are secreted into the culture medium (Macauley-Patrick *et al*., 2005; Ahmad *et al*., 2014; Fickers, 2014).

Cellulases have been expressed at high titres in *P. pastoris* and have displayed comparable activity and stability to their native forms (Lambertz *et al*., 2014). Two thermostable GH family 5 EGs, Egl2 and Egl3 from *A. fumigatus* Z5 was successfully cloned and expressed in *P. pastoris* X33 and displayed optimal actvity at pH 5 and 50°C and pH 4 and 60°C, respectively (Liu *et al*., 2011). Multiplicity is a common phenomenon with regard to

carbohydrate-active enzymes and the question arises as to whether it a matter of each enzyme having a distinct fucntion or genetic redundancy (Poidevin *et al*., 2013). Poidevin and co-workers (2013) underook to clone four putative GH family 6 cellulases from *Podospora anserina* to answer this question and found that although three of the enzymes had similar substrate specificities, they differed in terms of their catalytic activity, stability and modes of action. A thermophilic bacterial EG from *Acidothermus cellulolyticus* has also been expressed in *P. pastoris* as it has the advantage of being grown at lower temeratures and offered tight control of expression *via* methanol induction (Lindenmuth and McDonald, 2011).

1.6 Scope of the study

Thermophilic and thermostable EGs offer several potential advantages in the saccharification of lignocellulose which include but are not limited to: decreased risks of contamination, reduced hydrolysis times, increased solubility of reactants and a decrease in the cost of energy required for cooling following thermal pretreatment (Huy *et al*., 2015). The present study aims to identify a fungal isolate that is capable of producing thermophilic or thermostable EGs. Once a fungal isolate has been identified *via* EG screening assays, the EG is to be purified or cloned such that its biochemical properties and hydrolytic potential may be assessed for industrial application. The long term objective would be to utilize the purified or recombinant EG and test its activity on locally sourced lignocellulosic biomass such as sugarcane bagasse for the production of bioethanol.

1.7 Hypothesis

It is hypothesised that endoglucanases produced by fungal strains isolated from *Eucalyptus* spp. woodchips will possess desirable properties for application in industry.

1.8 Objectives

1.8.1 To screen the Lab 2 fungal culture collection at UKZN Westville Campus (Discipline of Microbiology) and identify isolates displaying exceptional endoglucanase activity

1.8.2 And to partially purify the endoglucanase

1.8.3 To data-mine the genome of *Phialophora alba* and identity putative carbohydrate-active enzymes

1.8.4 To clone and heterologously express the endoglucanase

CHAPTER 2

Screening of fungal isolates from *Eucalyptus* spp. woodchips and the characterization and partial purification of a thermostable endoglucanase

2.1 Introduction

There are numerous reports and reviews that detail the potential of filamentous fungi as sources of lignocellulosic enzymes. From an industrial standpoint, fungal enzymes offer distinct advantages over their bacterial counterparts. These include, higher yields, extracellular secretion and the production of auxiliary enzymes that are responsible for the efficient removal of polysaccharide side chains (Ja'afaru, 2013).

A key cellulase enzyme involved in the depolymerisation of lignocellulosic wastes into simple sugars for biofuel production are endoglucanases (β-1,4-endoglucanases; EG) (Narra *et al*., 2014). EGs are grouped into the glycoside hydrolase (GH) superfamily and have representatives in families 5-8, 12, 16, 44, 45, 48, 51, 64, 71, 74, 81, 87, 124 and 128 (Juturu and Wu, 2014). They generate oligosaccharides and new chain ends by randomly cleaving internal amorphous sites in the cellulose polysaccharide chain (Narra *et al*., 2014; Pol *et al*., 2012).

The majority of industrial cellulases and EGs in particular are of bacterial and fungal origin (Huy *et al*., 2015). Many industrial applications that make use of EGs require that the enzyme be robust and highly stable under specific reaction conditions as is the case with the bioconversion of lignocellulosic biomass which favours the use of thermoacidophilic and thermostable cellulases (Pol *et al.,* 2012; Zhao *et al*., 2011). However, fungal cellulases, such as those produced by *Penicillium* spp., *Trichoderma* spp. and *Aspergillus* spp. are mesophilic and not highly thermostable (Lynd *et al*., 2002; Zhao *et al*., 2011). Thus there is a constant

need for the discovery of novel EGs that possess extremophilic physicochemical properties for exploitation in industry.

Novel cellulases possessing extremophilic physicochemical properties can be obtained in one of two ways; by protein engineering or by isolating microorganisms from extreme environments and screening them for pre-existing extremophilic cellulases (Dutta *et al*., 2008). Cellulase properties can be improved through rational design or directed evolution (Zhang *et al*., 2006). The problems with these two strategies are that rational design requires detailed information on the structure of the protein which may not always available while the greatest challenge to directed evolution is the development of tools to efficiently evaluate the performance of recombinant mutants (Zhang *et al*., 2006). Screening environmental samples overcomes these limitations as it is very likely that microorganisms living in unique and extreme environments will possess enzymes (and possibly cellulases) that are active under extreme conditions. Extremophilic microorganisms maybe be found is environments such as salt pan, thermal springs, hydrothermal vents, soil and compost heaps (Maheshwari *et al*., 2000). These environments serve as potential reservoirs for cellulases with desirable properties.

The use of enzymes for industrial biotechnological applications is dependent on the economic feasibility of the enzyme production and recovery process. The choice of downstream processing is a major cost factor as the loss of the enzyme or bioproduct as a result of multiple separations, concentrations and purifications may compromise the efficiency of the entire process (Farinas *et al*., 2011). Thus, there is a need for the development of techniques and methods for downstream processing that preserves the native structure and function of the enzymes of interest (Herculano *et al*., 2012). With cost reduction in mind, precipitation can be an effective means of concentrating and purifying an enzyme (Farinas *et al*., 2011).

Precipitation takes advantage of the changes in the solubility of proteins relative to others by changing solvent conditions; with one of the most commonly used methods being ammonium sulphate precipitation (Burgess, 2009). Several studies focussing on the isolation and purification of cellulases and EGs of bacterial and fungal origin have used ammonium sulphate precipitation as a first step to achieve a pure or partially pure enzyme for characterization studies (Sul *et al*., 2004; Onyike *et al*., 2008; Pol *et al*., 2012; Shanmugapriya *et al*., 2012; Dave *et al*., 2015). However, the primary drawback to using ammonium sulphate precipitation to purify proteins is that contaminants often precipitate together with the target protein (Duong-Ly and Gabelli, 2014).

Other groups have chosen to pursue the purification of cellulases and EGs using different techniques or methods of precipitation. Elakkiya and Muralikrishnan (2014) compared acetone precipitation and ammonium sulphate precipitation of cellulase from *Trichoderma viride* and found that the recovery of the enzyme with ammonium sulphate precipitation was superior to the use of acetone. Initial purification of the EG from *Aspergillus terreus* DSM 826 with acetone resulted in a 60% recovery, however, following chromatography, even though a 27-fold purification was obtained, only 10.5% was recovered (Elshafei *et al*., 2009). Farinas and co-workers (2011) evaluated ammonium sulphate and ethanol as precipitation agents for the recovery EGs and xylanases and found that both agents resulted in similar recoveries, however, the recovery of xylanases in both instances were much lower (27 and 25% for ammonium sulphate and ethanol, respectively) than for EGs (65 and 61% for ammonium sulphate and ethanol, respectively). Protein may also be precipitated by the use of non-ionic polymers such as polyethylene glycol (PEG) (Burgess, 2009). This is accomplished by setting up an aqueous two-phase system (ATPS), formed by mixing a polymer with an inorganic salt or by mixing to incompatible polymers (Herculano *et al*., 2012). The cellulase complex from *A. japonicus* URM5620 was fractionated using a PEG/citrate ATPS

(Herculano *et al*., 2012). It was found that β-glucosidases partitioned to the salt phase while the endoglucanases and other cellulases associated with the top PEG phase, however, the highest EG activity yield was only 1.64%.

This initial phase of work aimed to identify fungal strains that produce thermophilic EGs (with high activities and stabilities above 50°C) that could be purified *via* ammonium sulphate precipitation and would have potential application in the liberation of fermentable sugars from lignocellulosic biomass for bioethanol production.

2.2 Materials and methods

2.2.1 Screening for endoglucanase production

Based on preliminary screening of fungal isolates from *Eucalyptus* spp. woodchips for cellulase activity by Govender (2012), twenty seven fungal isolates were screened for cellulase activity on potato dextrose agar (PDA) supplemented with 1% carboxymethyl cellulose (CMC). The cultures were incubated for five days at 50°C, thereafter; the plates were stained with a 0.1% Congo red solution for 30 minutes and destained with 1 M NaCl for 15 – 20 minutes (Teather and Wood, 1982). Cellulose hydrolysis presented as clear zones around the fungal colonies. The extent of EG activity was determined based on the diameter and intensity of hydrolysis of the CMC. Ten isolates presenting highest activity were selected for time course analysis to determine the optimal production of EG. Working stocks were maintained on PDA at 4°C.

2.2.2 Time course analysis for optimal production of endoglucanases

The EG producers identified in 2.2.1 were grown in CMC broth: basal culture medium (2.0 g KH_2PO_4, 0.3 g $CaCl_2$, 1 g $MgSO_4.7H_2O$, 2 g NH_4NO_3, 5 mg $FeSO_4.7H_2O$, 1.6 mg $MnSO_4.4H_2O$, 3.45 mg $ZnSO_4.7H_2O$ and 2 mg $CoCl_2.6H_2O$ per liter [Gilna and Khaleel, 2011]) supplemented with 1% CMC. One hundred millilitres of CMC broth in 250 ml

Erlenmeyer flasks were inoculated with a 1 cm^3 piece of mycelium from a 4 day old actively growing culture for each isolate (in triplicate). The flasks were incubated at 50°C on an orbital shaker (New Brunswick, Innova 44, Eppendorf, Hamburg, Germany) at 200 rpm for 7 days. One millilitre aliquots were collected from each flask every 24 hours. The aliquots were centrifuged at 13 000 rpm for 5 minutes and the supernatant was collected and assayed for EG activity. Protein concentrations were determined with the method of Bradford (Bradford, 1976), using bovine serum albumin as the standard. The isolate that displayed the highest EG activity was selected be used for the continuation of the study.

2.2.3 Determination of endoglucanase activity

Endoglucanase activity was determined using the dinitrosalicylic acid (DNS) method (Bailey *et al*., 1992). The substrate for the cellulase assay was prepared by homogenising 1 g of CMC in 80 ml 50 mM citrate buffer, pH 5 at 60°C and heating to boiling. The solution was cooled with continuous stirring overnight and then made up to 100 ml with buffer and stored at 4°C until required (Bailey *et al*., 1992).

The assay was conducted by incubating 100 µl of crude enzyme extract with 900 µl of substrate solution for 10 minutes at 50°C. One and a half millilitres DNS reagent was added to the enzyme-substrate mixture and then boiled for 10 minutes and thereafter cooled in cold water. The absorbance of the solution was measured at using a spectrophotometer (Biochrom Libra L2 (S12), Cambridge, UK) at a wavelength of 540 nm. A glucose standard curve was constructed to determine the activity of the endoglucanase, where one unit of cellulase activity (IU) was defined as the amount of enzyme required to release 1 µmol of reducing sugar (glucose) per minute under the assay conditions. One IU is equal to 16.67 $nkat.ml^{-1}$.

The buffer blanks were treated in the same manner as sample aliquots, except 100 µl of the 50 mM citrate buffer was used instead of enzyme. Enzymes blanks were prepared by the

addition of the DNS reagent first, followed by the addition of 100 µl extract, thus eliminating any background of reducing sugars in the culture supernatant. All assays were performed in triplicate and used to caluculate the average value and standard deviations which were used to plot the graphs.

2.2.4 Determination of pH and temperature optima of crude extracts

The optimum pH of the EG was determined by incubating the crude enzyme for 10 minutes with substrates prepared in various buffers, namely, citrate buffer (pH 4-5), phosphate buffer (pH 6-8), and Tris-HCl (pH 9) after which the reaction was stopped using the DNS reagent and the amount of reducing sugar liberated was determined as described in section 2.2.3. The temperature optimum was determined by assaying EG activity at different temperatures ranging from 50°C-90°C at the optimum pH.

2.2.5 Determination of pH and temperature stability of crude extracts

pH stability of the EG was determined at each pH as described above at the optimal temperature. The crude enzyme was subjected to buffer exchange using centrifugal concentrators and incubated at each pH for 2 hours. Centrifugal concentration was performed by centrifuging 500 µl of the crude enzyme extract through centrifugal concentrators down to 50 µl. The supernatant was collected and topped up to 500 µl with the appropriate buffer. Five hundred microlitre aliquots were withdrawn every 30 minutes and assayed for residual EG activity using the DNS assay (Bailey *et al*., 1992).

Temperature stability was determined in a similar fashion. The enzyme was incubated at the optimum pH at varying temperatures, ranging from 50-90 °C for 2 hours. Five hundred microlitre aliquots were withdrawn every 30 minutes and assay for residual EG activity using the DNS assay. Reagent and enzyme controls were used. A tube containing only substrate was also incubated with the test samples to monitor degradation of the substrate due to the

high temperatures and in so doing, back ground hydrolysis was eliminated when calculating the amount of reducing sugar being released (Bailey *et al*., 1992).

2.2.6 Determination of molecular weight of the endoglucanase

The molecular weight of the EG was determined by subjecting the crude enzyme extract to 12% SDS-PAGE followed by renaturation and zymogram analysis and native PAGE and zymogram analysis. SDS-PAGE was run according to Laemmli (1970) at 90 V for 3 hours. Renaturation was performed by first rinsing the gel in distilled water and then washing overnight in a solution of water, ethanol and glacial acetic acid (4:1:5). The solution was discarded and the gel was subsequently washed in fresh water:ethanol:acetic acid solution for 20 minutes. Thereafter, the gel was rinsed twice with distilled water and then twice with phosphate buffer (pH 6). The proteins were renatured by placing the gel in a 1% Triton-X 100–phosphate buffer (pH 6) solution with gentle agitation for 30 minutes. The solution was discarded and the gel was rinsed twice for 20 seconds in phosphate buffer (pH 6) and then washed in phosphate buffer (pH 6) for one hour.

Zymogram analysis was performed by placing the renatured gel above on a 0.1% CMC substrate gel (0.1% CMC in phosphate buffer pH 6) and incubating at 50°C for 1 hour, followed by staining of the substrate gel with 0.1% Congo red for 15 minutes and destaining with 1 M NaCl until clear zones of hydrolysis were visible. The substrate gel was then fixed in 5% acetic acid (Teather and Wood, 1982).

Native-PAGE was performed by incorporating CMC to a final concentration of 0.1% in phosphate buffer (pH 6) into the resolving gel. Following electrophoresis, the gel was washed twice in 25% isopropanol (in phosphate buffer, pH 6) for 30 minutes and then twice in phosphate buffer (pH 6). The excess fluid was removed and the gel was placed in an

incubator at 50°C for one hour. The gel was thereafter stained with 0.1% Congo red for 15 minutes and destained with 1 M NaCl to observe EG activity.

2.2.7 Partial purification of the endoglucanase

All procedures were carried out at 4°C unless otherwise stated. The crude culture extract was subjected to fractional $(NH_4)_2SO_4$ (ammonium sulphate) precipitation (20–100%) (Burgess, 2009). The precipitate was collected by centrifugation (Eppendorf Centrifuge 5430R, USA) at 12 000 rpm for 20 minutes. The supernatant was collected for further precipitation while the protein pellet was resuspended in 50 mM phosphate buffer buffer (pH 6) and dialyzed against the same buffer with three buffer changes at four hour intervals (Li *et al*., 2003). Following each saturation step, both the supernatant and protein pellet were assayed to determine EG activity and protein concentration. Crude culture filtrate was also subjected to IEF fractionation using the ZOOM® IEF Fractionator (ThermoFisher Scientific, Massachusetts, USA) according to manufacturers instructions. Each fraction was thereafter assayed for EG activity. Attempts to concentrate the protein were also made by freeze drying the culture filtrate to a fine power and thereafter resuspending ~100 mg of the power in 200 µl of phosphate buffer pH 6 and assaying for EG activity.

2.3 Results and Discussion

2.3.1 Screening for endoglucanase production

Endoglucanases, being mainly extracellular enzymes may be produced in both liquid and solid media, however, solid substrates allow for fast screening of large collections of fungal isolates as well as allowing for the detection of specific enzymes (Doolotkeldieva, 2011). The cellulolytic activity of 27 fungal isolates known to possess cellulolytic activity was assessed on PDA plates supplemented with 1% CMC. EG activity produced zones of clearing, or

halos, around the fungal mycelia following the Congo red plate screening assay. The extent of EG activity for each fungal strain assessed is tabulated in Table 2.1.

EG activity was assessed based on the size and intensity of the clear zones following the Congo red assay. Apart from assessing EG activity, an environmental constraint was placed on the fungi by growing them under thermophilic conditions (50°C); this was in force in order to eliminate mesophilic fungi and enrich for those isolates which could potentially produce thermophilic endoglucanases. The most efficient degraders as determined by this assay are highlighted in yellow in Table 2.1 with the majority of fungi displaying good activity being isolated from *Eucalptus grandis* woodchips. Plate screening assays such as this are semi-quantitative as the relationship between clear zone diameters and enzyme activities are non-linear as well as there being difficulty in observing activity when low levels of enzyme are secreted (Zhang *et al*., 2006). Thus the selected isolates were grown in broth culture to quantitatively determine EG activity and the time course for optimal EG production.

Table 2.1: Preliminary assessment of endoglucanase activity of fungal isolates

Source of fungi	Isolate name	Activity
Mixed species woodchips	*Paecilomyces* sp.	+ (I)
	Phanerochaete chrysosporium	+/-
	Unknown isolate 8	No growth
	Unknown isolate 9	+ (M)
	Phialophora alba	++ (M)
	Curvularia sp.	+ (M)
	Unknown isolate 14	+++ (I)
Eucalyptus nitens woodchips	Isolate N1	+ (M)
	Isolate N2 - *Aspergillus fumigates*	++ (M)
	Isolate N3	++ (I)
	Isolate N4 – *Penicillium spinulosum*	++ (I)
	Isolate N5	Overgrown – could not stain[b]

	Isolate N6	Overgrown – could not stain[b]
Eucalyptus grandis woodchips	Isolate G4	++ (M)
	Isolate G6 – *Aspergillus fumigates*	+++ (I)
	Isolate G7	+++ (I)
	Isolate G8 – *Penicillium spinulosum*	+++ (M)
	Isolate G9	+++ (M)
Eucalyptus dunnii woodchips (Sappi)	Isolate D1	+/-
	Isolate D3	+++ (M)
Eucalyptus dunnii woodchips (CSIR)	Isolate D1 (FFP) - *Penicillium spinulosum*	+ (I)
	Isolate D3 (FFP)	No growth
	Isolate D8 (FFP)	+ (M)
	Isolate D9 (FFP) - *Aspergillus fumigatus*	++ (I)
	Isolate D10 (FFP)	++ (M)
	Isolate D13 (FFP)	Overgrown – could not stain[b]
	Isolate D14 (FFP)	Overgrown – could not stain[b]

Key: +/- = Uncertain (Zones of hydrolysis could not be visualized with certainty even after repeating); + = Activity (Zone of hydrolysis less than 10 mm); ++ = Good activity (Zones of hydrolysis greater than 10 mm), +++ = Very good activity (Majority of plate hydrolysed); I = Intense clearing; M = Mild to opaque clearing

[a]Those isolates with no genus and species names have not been identified

[b]Cultures were re-grown and incubated for a shorter period, however, the shortened incubation period was not sufficient to observe endoglucanase activity *via* plate screening

2.3.2 Time course analysis for optimal production of endoglucanases

All ten fungal isolates were cultivated in shake flask culture in a minimal medium supplemented with CMC to induce expression and secretion of extracellular EGs. The optimal time for EG production for each isolate was determined by subjecting the culture supernatant to the DNS assay every 24 hours. The time course for optimal production is shown in Figure 2.1. EG activity was present in each culture supernatant as early as day one with a steady increase each day. The highest producers, unknown isolate 14 and D9 (FFP) showed highest activity around day five. Preliminary characterization screening assays were performed on the EGs from isolates 14 and D9 to determine pH and temperature optima. The EG from isolate D9 (FFP), *Aspergillus fumigatus*, proved to be superior.

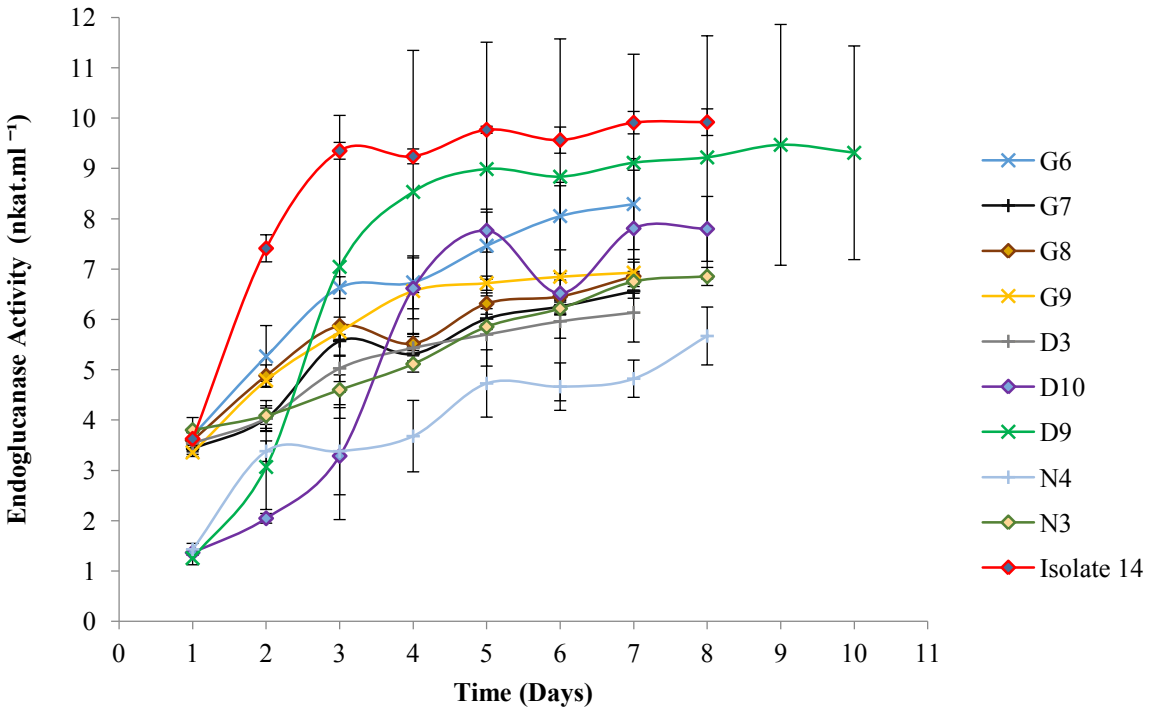


Figure 2.1: Time course analysis for optimal endoglucanase production of fungal isolates displaying desirable endoglucanase activity

A similar trend was noted for *A. fumigatus* Z5 with EG production having a steady increase from Day 1 with optimal production on Day 9 (Liu *et al*., 2011). The first discrepancy however, lies in the level of activity and amount of protein being produced by *A. fumigatus* D9 (FFP). The highest EG activity obtained in this study was 0.6 U/ml while the activity of *A. fumigatus* Z5 was several times higher at 341.3 $U.g^{-1}$ dry weight of substrate under solid state fermentation (Liu *et al*., 2011). Difficulty was encountered in comparing the activity of the EG of a *A. fumigatus* to those in literature as this depends on how individual authors choose to firstly, treat their samples prior to assaying (concentrating or purification) and secondly, how they choose to express their activity ($nkats.ml^{-1}$, Units [U], specific activity, relative activity or residual activity).

The second factor that needs to be considered when inducing the production of EGs is the mode of fermentation. The present study opted for shake flask culture, or submerged

fermentation; however, cellulases may also be produced using solid state fermentation. Under solid state fermentation, enzymes are produced at a higher yield and stability, however, it tends to be technically more difficult to perform than submerged fermentation (Hölker and Lenz, 2005; Singhania *et al*., 2009). And thus the third factor that will need to be considered to achieve high yields of fungal EGs are the effects of different carbon and nitrogen sources (Panagiotou *et al*., 2003; Narra *et al*., 2014). Durrant (1996) showed that by even varying the amount of oxygen (aerobic, microaerophilic and anaerobic) can result in varying degrees of EG multiplicity. Unfortunately, the optimization of fungal EG production lay outside the scope of this study and was thus not pursued.

2.3.3 Characterization of crude enzyme extract

The enzymes produced by *Aspergillus* spp. for the degradation of plant cell wall are of great importance in a number of industries (de Vries and Visser, 2001). The crude enzyme supernatant was concentrated 10 fold, using centrifugal concentrators with a molecular weight cut-off of 10 kDa. Characterization of the crude cellulase extract of *A. fumigatus* was carried out to determine optimum pH and temperature for activity as well as to investigate the pH- and thermo-stability of the enzyme.

The pH profile of the crude cellulase extract (Figure 2.2) was determined between pH 4-9 and displayed a broad spectrum of activity with mildly acidic and mildly alkali pH optima at pH 6 and pH 8. The crude extract displayed higher activities at acidic pHs peaking at 9.8 $nkat.ml^{-1}$ with a decrease in activity to 7.8 $nkat.ml^{-1}$ to 6.9 $nkat.ml^{-1}$ at pH 7 and pH 9, respectively. The pH stability of the crude enzyme was determined over a two hour period at 50°C for the various pHs and can be found in Figure 2.3 (A-F).

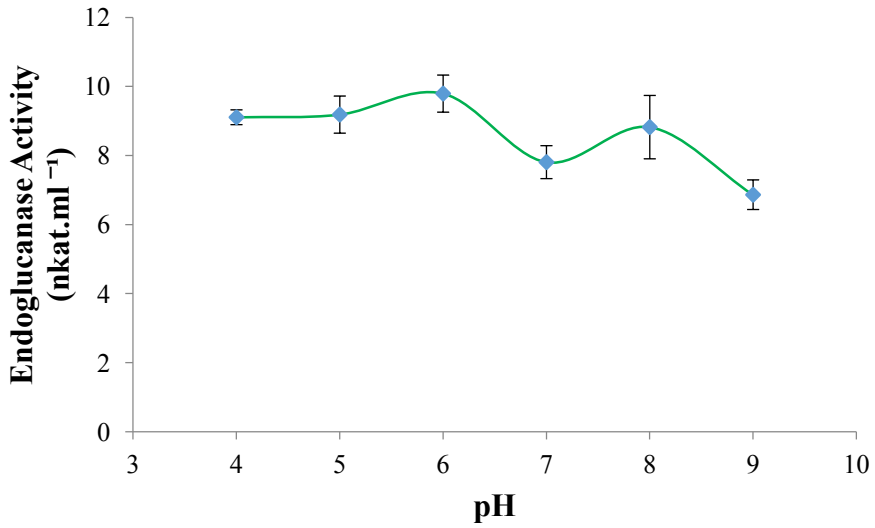


Figure 2.2: Effect of pH on endoglucanase activity from crude culture filtrate of *A. fumigatus*

The pH profile of the crude enzyme extract is similar to reports of EGs from other *Aspergillus* spp. with optimum activity under acidic conditions with activity proving to be stable over a wide range of pH values. The EG displayed an optimum pH of 6 and retained almost 100% activity for up to 90 minutes with a slight decrease to 76% at 120 minutes. The family 12 EG from *A. terreus* was active over a pH range of 3.6-5 with optimal activity at 4.8 but only retaining 50% activity after 120 minutes at pH 4 (Narra *et al*., 2014), while the EG purified from *A. aculeatus* displayed a pH optimum of 5 with pH stability not being reported (Naika *et al*., 2007). The EG produced by *A. glaucus* possessed a pH optimum of 4 and was stable at pH 3.5-7.5 after four hours, however, this was determined at 4°C and not at the optimum temperature (Tao *et al*., 2010). pH stability of fungal EGs under alkaline conditions has not been widely documented. The *A. fumigatus* crude EG was stable at alkaline pHs having retained 89% activity after 75 minutes with very erratic activity toward to the end of the two hour incubation period at pH 8 as well as retaining 95% activity after 90 minutes at pH 9.

The presence of a second, albeit lower peak in activity at pH 8 may be indicative of endoglucanase isozymes which may possess differing properties as fungal culture supernatants are known to contain multiple cellulases (Juturu and Wu, 2014). Onsori and colleagues (2005) demonstrated this via zymogram analysis of crude EG from an *Aspergillus* sp. that displayed three proteins with cellulolytic activity while Grigorevski-Lima and colleagues (2009) demonstrated that an *A. fumigatus* strain FBSPE-05 produced six protein bands with CMCase activity. This multiplicity could account for the erratic results seen on the stability profiles. The charge of acidic and basic amino acids are susceptible to change based on the pH of the environment. A change in pH may effect a change in the net charge and confirmation of a protein which may result in an increase or decrease in activity. If there are multiple endoglucanases present, the isozymes may be active or inactive (or less or more active) at different pHs and possibly even at different times if ionization of amino acid side chains are constantly in flux.

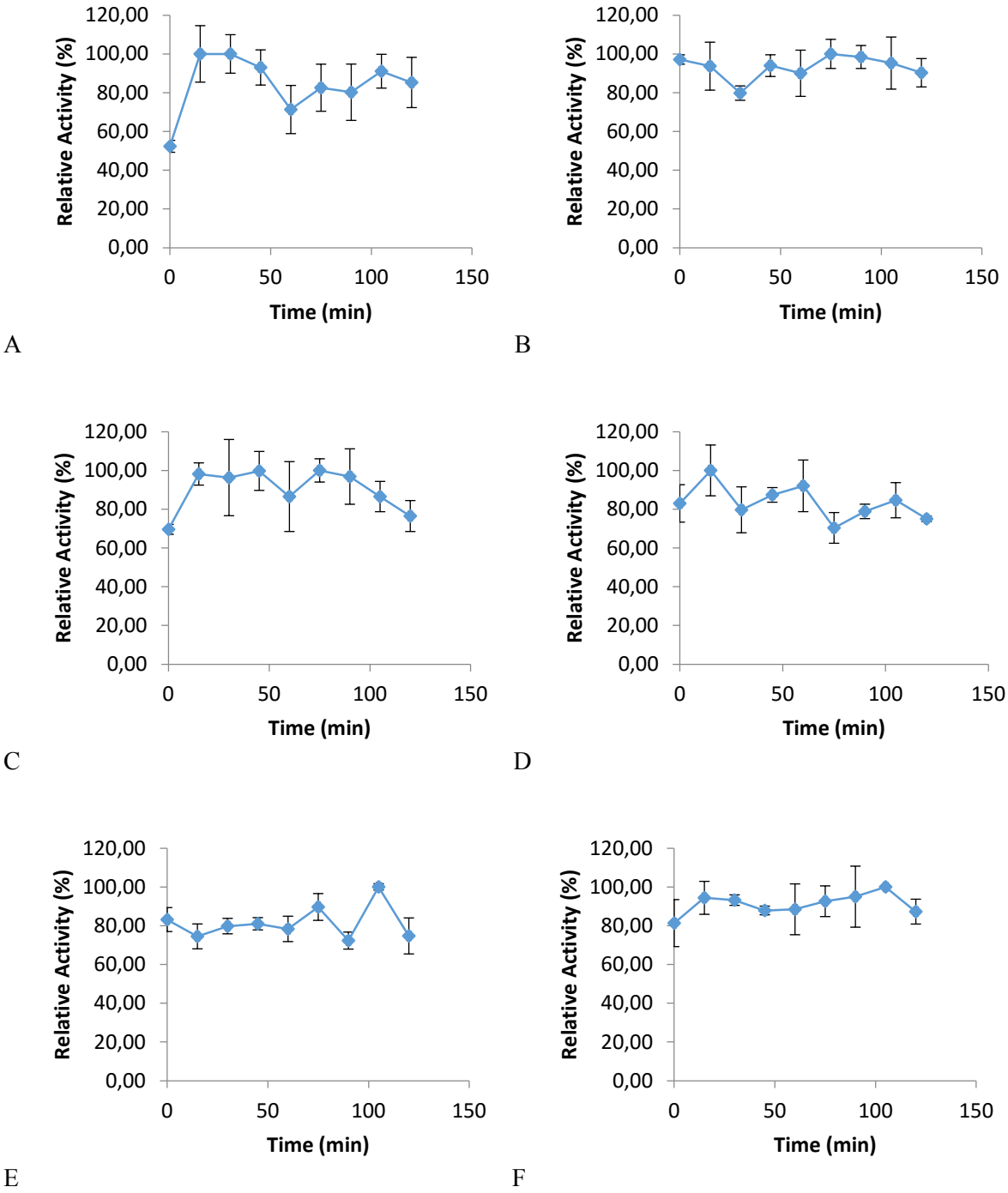


Figure 2.3: pH Stability profile of endoglucanase from crude culture filtrate of *A. fumigatus* at A. pH 4, B. pH5, C. pH 6, D. pH 7, E. pH 8, F. pH 9

The effect of temperature on the activity of the crude extract was determined from 50-90°C at the optimum pH of 6. Again the enzyme performed very well over the range of temperatures investigated with optimum activity at 50°C at 9.2 nkat.ml^{-1} and very steady activity between 8.2-8.8 nkat.ml^{-1} between 60-90°C (Figure 2.4). The EG thermostability was investigated under thermophilic conditions, at 50°C and above over a 2 hour period and proved to be very stable (Figure 2.5 A-E).

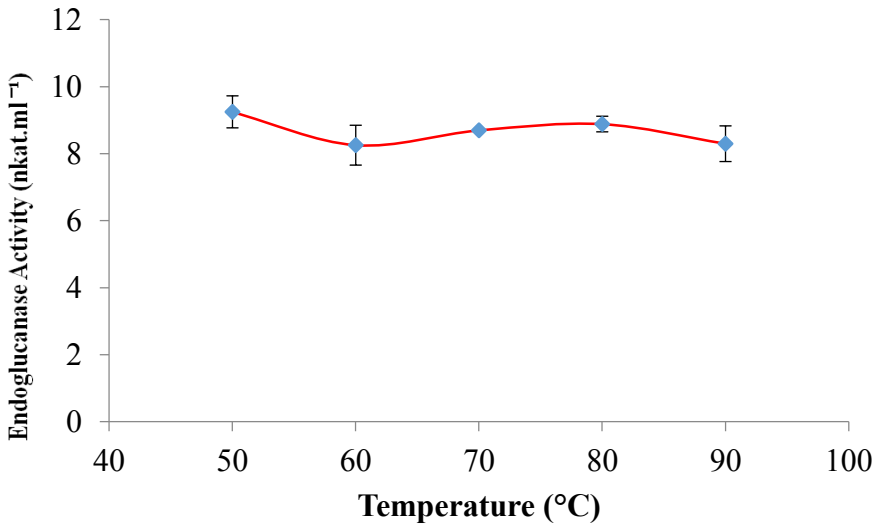


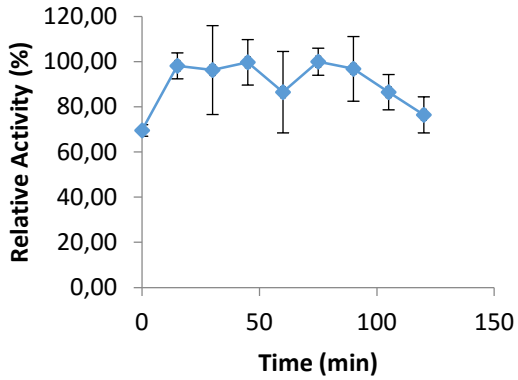
Figure 2.4: Effect of temperature on endoglucanase activity from crude culture filtrate of *A. fumigatus*

The highest EG activity was observed at 50°C, much like many other fungal EGs which possess temperature optima between 50-60°C (Tao *et al*., 2010). There are, however, EGs that do not follow this norm, displaying far greater thermal stability. Such is the case with *A. fumigatus* FBSPE-05 which displays optimal EG activity at 65°C (Grigorevski-Lima *et al*., 2009), *A. terreus* M11 EG which has optimum activity at 70°C (Gao *et al*., 2008) and *A. fumigatus* SMN1 with optimum activities at 65.5°C and 64.3°C (Saqib *et al*., 2010). Differences in temperature optima may occur as a result of differing amino acid compositions which may confer greater stability at higher temperatures. Saqib and colleagues (2010) also

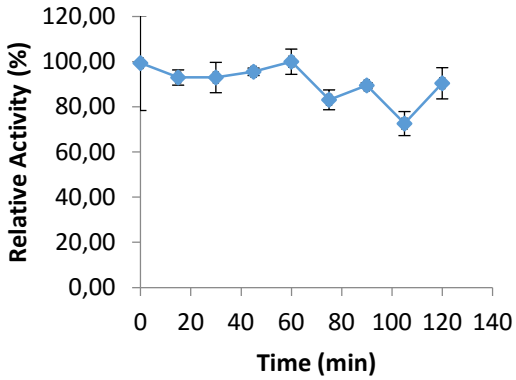
demonstrated that structural differences may arise in the same proteins when produced under different fermentation conditions that ultimately affect thermal activity and stability as a result of carbohydrate moieties that are non-covalently attached to the enzymes.

The crude enzyme displayed remarkable thermostability retaining almost 100% activity after 90 minutes at 50°C and retaining between 70-90% after 120 minutes at 70 to 90°C. The stability observed at optimum temperature is unsurprising as *A. fumigatus* FBSPE-05 also retained 100% activity after 2 hours at 50°C (Grigorevski-Lima *et al*., 2009), while *A. terreus* retained 99% activity after 150 minutes at 50°C (Narra *et al*., 2014). However, there appeared to be no significant loss in activity over the higher range of temperatures investigated which is uncommon among endoglucanases characterized thus far. Incubation of EGs at temperatures higher than their optima for extended periods of time has often resulted in a drastic decline in their relative activities. The EG produced by *A. aculeatus* lost 50% of its initial activity after 30 minutes incubation at 90°C (Naika *et al*., 2007), the EG from *A. terreus* suffered rapid inactivation at 70°C retaining only 10% activity after 30 minutes (Narra *et al*., 2014) and *A. glaucus* XC9 retained 60% activity after an hour's incubation at 60°C (Tao *et al*., 2010).

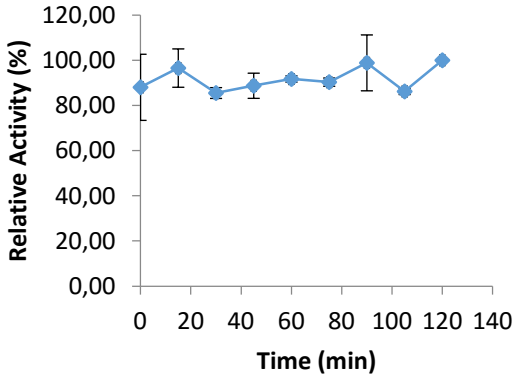
Although the EG activity appeared to be extremely stable over the pH and temperature ranges tested, the actual activity of the enzyme was extremely low between 2 – 3 nkat.ml^{-1} and this was cause for concern as the DNS assay that is employed to evaluate EG activity tends to overestimate the activity of hydrolytic enzymes (Gusakov *et al*., 2011). At such low levels of activity, combined with a low amount of protein, the accuracy of the DNS assay could not be trusted with certainty and thus, in line with the scope of this study, purification of the EG/s were attempted such that a higher concentration of enzyme might be obtained and used for more reliable characterization.



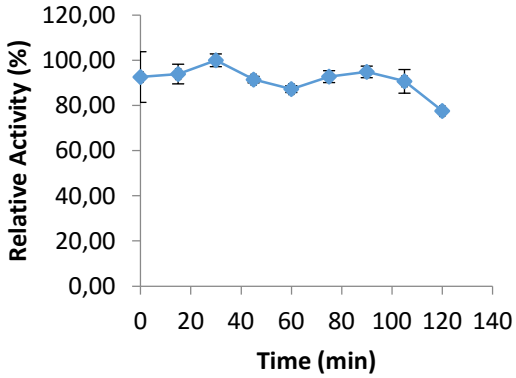
A



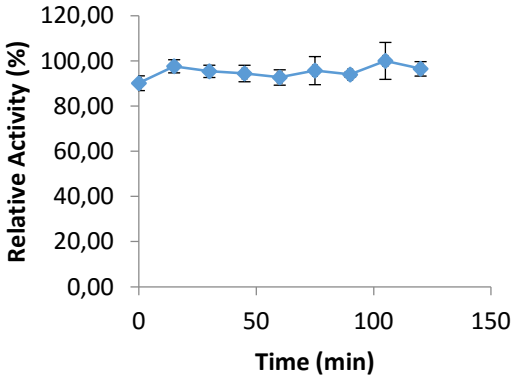
B



C



D



E

Figure 2.5: Thermostability profile of endoglucanase from crude culture filtrate of *A. fumigatus* at A. 50 °C, B. 60 °C, C. 70 °C, D. 80 °C, E. 90 °C

2.3.4 SDS-PAGE and zymogram analysis

SDS-PAGE and zymogram analysis were performed to determine the molecular weight and number of EGs present. SDS-PAGE revealed the presence of at least 7 bands ranging in size from 100 to 25 kDa (Figures 2.6 and 2.7). SDS-PAGE analysis of the crude cellulase of *A. fumigatus* Z5 revealed that the strain secretes many more proteins (17 proteins were visualized) into the medium than observed in this study. The proteins spanned a wide range, from 90 kDa to 10 kDa with eight of them displaying EG activity following zymogram analysis (Liu *et al*., 2011). Six of the proteins presenting EG were between 50 and 18 kDa which are where many of the *A. fumguatus* protein in this study was located. It was unfortunate however, that zymogram analysis was unsuccessful following renaturation and as a result those proteins possessing EG activity could not be identified.

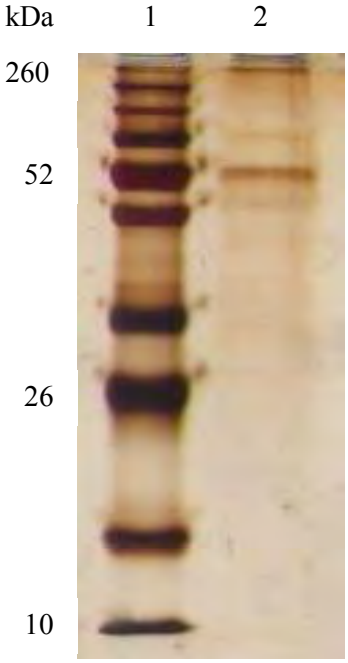


Figure 2.6: Silver stained 12% SDS-PAGE of day 9 unconcentrated crude culture filtrate of *A. fumigatus* (Lane 1: Marker: Spectra Multicolour Broad Range Protein Ladder; 2: Day 9 *A. fumigatus* culture filtrate)

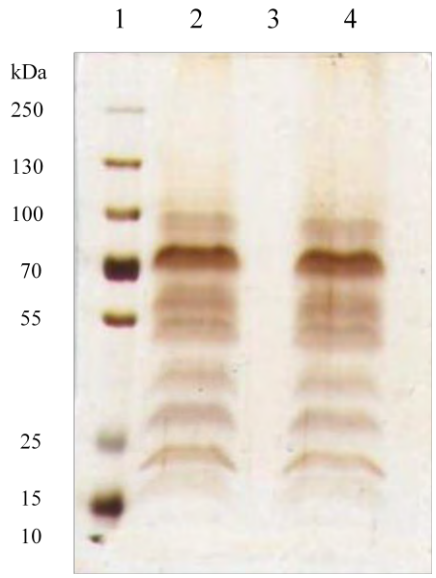


Figure 2.7: Silver stained 12% SDS-PAGE of 10× concentrated crude enzymes (1: Marker: PageRuler Plus Prestained Protein Ladder; 2 and 4: 10X concentrated sample of *A. fumigatus* culture filtrate)

Sample preparation for SDS-PAGE involved the addition of β-mercaptoethanol and heating of the sample nearing boiling. β-mercaptoethanol is a powerful reducing agent which cleaves disulfide bonds which helps in the linearization of peptides while exposure to extreme heat also aids in the efficient denaturation of proteins thus allowing them to separate based on peptide chain length with no influence from secondary structures. It was evident by the lack of EG activity on CMC substrate plates that the proteins did not return to their native or any active form following renaturation attempts.

Hence native-PAGE and zymogram analysis was performed. Samples were not exposed to SDS, β-mercaptoethanol or to heat denaturation. CMC was incorporated into the PAGE gel which allowed for uncomplicated detection of EG activity. The zymogram in Figure 2.8 used to confirm EG activity, shows a smear of high molecular weight bands displaying EG activity. The exact molecular weight of the EG could not be determined as the protein ladder used contains denatured peptides. EGs from *Aspergillus* spp. range from as low as 12.5 kDa to as high as 112 kDa (de Vries and Visser, 2001). The zymogram below is indicative of the

presence of multiple high molecular weight EGs. Zymogram analysis of *A. fumigatus* FBSPE-05 culture supernatant presented bands at 27.3, 35.3, 56.4, 73.4, 82.2 and 94.9 kDa (Grigorevski-Lima *et al*., 2009). Cellulase and EG multiplicity may occur as a result of post-transcription modification, post-translational modifications or as a result of the presence of multiple genes (Maheshwari *et al*., 2000). *T. reesei* is known is cellulase expression also triggers protease expression and as a result, and thus multiple forms of cellulases may also be observed due to proteolytic cleavage (Juturu and Wu, 2014).

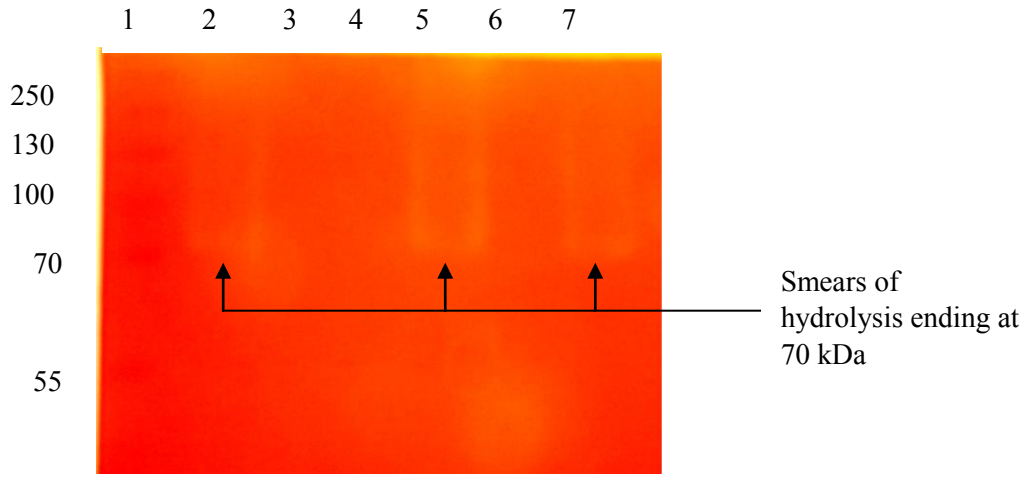


Figure 2.8: Zymogram analysis of Native PAGE (1: Marker: PageRuler Plus Prestained Protein Ladder; 2, 5 and 7: 10× concentrated sample of *A. fumigatus* culture filtrate)

2.3.5 Ammonium sulphate fractionation

In an attempt to concentrate and purify the EG enzymes for characterization, the crude extract was produced in bulk and subjected to ammonium sulphate precipitation. Table 2.2 details the recovery and yields following ammonium sulphate precipitation.

Table 2.2: Ammonium sulphate precipitation of crude culture filtrate of *A. fumigatus*

Ammonium sulphate (%)	**Protein**			**Endoglucanase Activity**			**Enzyme recovered (%)**
	Volume (ml)	Protein (µg/ml)	Total Protein (µg)	Activity (IU/ml)	Total Enzyme (IU)	Specific Activity (U/µg)	
0	300	0.011	3.306	0.536	160.863	48.656	
20	10	0.055	0.554	0.230	2.305	4.162	0.086
40	10	0.071	0.710	0.294	2.942	4.142	0.110
60	10	0.004	0.041	0.205	2.047	50.150	0.076
70	5	0.011	0.054	0.245	1.223	22.765	0.046
80	5	0.025	0.125	0.299	1.497	11.960	0.056
90	5	0.042	0.210	0.387	1.934	9.230	0.072
100	5	0.003	0.017	0.346	1.731	101.796	0.065

The total protein present in the crude sample was extremely low, 3.3 µg with a specific EG activity of 48.6 U/µg. Low levels of EG were recovered at low ammonium sulphate concentrations (20-40%) and increased at 60-70%. However, the recovery of EG was extremely low with a saturation point at 60% ammonium sulphate (EG recovery was less than 1%). Nearly 50% of the total protein was not recovered after 90% ammonium sulphate was added and it is likely that they still reside in the supernatant. They could not be recovered as it even though 100% ammonium saturation was performed ammonium sulphate approaches its limit of solubility after 90% and further saturation is generally not recommended (Duong-Ly and Gabelli, 2014). The low initial concentration of protein as well as the multiplicity of EGs may have contributed to the low recovery the target enzyme.

The isozymes may possess differing numbers of hydrophobic amino acids which determines the ammonium sulphate concentration at which the protein leaves a solution (Sadasivam and Manickam, 1996; Nigam and Ayyagari, 2007). A review of thermophilic fungi and their enzymes revealed that some EGs, such as those isolated from *Humicola grisea* var. *thermoidea* and *Chaetomium thermophile* var. *dissitum* are unable to hydrolyse cellulose in

the absence of CBHs (Maheshwari *et al*., 2000). Thus, if such synergism is required for the activity of the EG in question, and if the components of the cellulase system of *A. fumigatus* precipitate in different fractions, it could account for the low activity observed. Another possible reason as to why there would be low activity is if protein samples did not dialyse properly, the ammonium sulphate may have interfered with the DNS assay reaction in some way. However, other than reading online research forums (Research Gate) of researchers with similar issues, there is no documented support for this assumption. Farinas and colleagues (2011) reported that at lower concentrations of ammonium sulphate (40%), EGs did not precipitate; however, they did at higher concentrations (80%). Precipitation at high ammonium sulphate concentrations could pose problems as crude supernatants tend to contain undesirable components such as undefined polysaccharides, pigments and other interfering proteins that could precipitate together with EG fractions and negatively impact on the activity of the target enzyme (Herculano *et al*., 2012). Due to the difficulty experienced during ammonium sulphate precipitation, IEF fractionation was attempted to yield a partially pure and concentrated EG, however, this technique did not yield any results of significance.

2.4 Conclusion and way forward

The preliminary characterization of the crude EG of *A. fumigatus* D9 (FFP) indicates that the enzyme is mildly acidothermophilic. The traditional hydrolysis of cellulose involves acid treatment at high temperatures, followed by enzymatic hydrolysis (Gao *et al*., 2008). The EG investigated in this study may not be suitable for application in biofuel from lignocellulosic biomass when its properties compared to other Aspergilli (de Vries and Visser, 2001) as well as other fungi (Payne *et al*., 2015). In comparison to *Trichoderma* sp., *A. niger* produces much lower levels of EGs and *A. fumigatus* D9 (FFP) seems to follow the same trend (Narra *et al*., 2014).

At this point, after careful consideration, taking into account the extremely low protein concentration, and the vast amount of research and data available on cellulases and EGs from *Aspergillus* species, it was decided that the focus of the study be shifted toward another thermophilic fungus. The fungus in question would be *Phialophora alba*, a thermophilic fungus isolated from *Eucalyptus* woodchips that is known to produce highly active thermostable xylanases. In order to potentially avoid the same road blocks experienced with purification of the EG from *A. fumigatus* a molecular and bioinformatic approach was employed. The latter part of this dissertation will thus focus on the cloning and expression of an EG gene from *P. alba*.

CHAPTER 3

Carbohydrate-active enzyme analysis of *Phialophora alba* isolated from *Eucalyptus* spp. woodchips

3.1 Introduction

Several studies have demonstrated that the saprophytic, parasitic and/or plant pathogenic nature of fungi makes them an important reservoir of carbohydrate-active enzymes (CAZymes) which are responsible for the biosynthesis, breakdown or modification of polysaccharides and their derivatives which make up plant cell walls (Ospina-Giraldo *et al*., 2010; Zhao *et al*., 2014; Brouwer *et al*., 2014). CAZymes may be grouped into five functional classes: glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and the latest addition to the CAZymes database, enzymes possessing auxiliary activities (AA) (Lombard *et al*., 2014; http://www.cazy.org).

Glycoside hydrolases cleave the glycosidic bond between carbohydrates or between carbohydrates and a non-carbohydrate substituent such as protein or lipids while GTs are involved in the biosynthesis of di-, oligo- and polysaccharides by catalysing the transfer of sugar moieties and forming glycosidic bonds (Lombard *et al*., 2014). Glycoside hydrolases such as cellulases and xylanases have been extensively studied in fungi for industrial exploitation as discussed at length in Chapter 1. Polysaccharide lyases degrade pectin and glycosaminoglycans while CEs catalyse the de-*O* or de-*N*-acylation of substituted sugars (Zhao *et al*., 2013). Auxiliary activity enzymes are redox enzymes that include lytic polysaccharide mono-oxygenases and ligninolytic enzymes (Veneault-Fourrey *et al*., 2014).

CAZymes can make-up 1-5% of an organisms coding sequences (Lombard *et al*., 2014). Recent studies have taken to analysing the collection of CAZymes encoded in an organism's

genome, referred to as the "CAZome". The information garnered from CAZome analysis has been used to investigate the metabolic potential of organisms, in particular, fungal plant pathogens and parasites, as CAZymes involved in cell-wall degradation are seen as pathogenicity factors (Ospina-Giraldo *et al*., 2010; Zhao *et al*., 2013; Brouwer *et al*., 2014).

With the explosion in genome sequencing efforts over the past decade, there has been a concurrent surge in the prediction of protein functions by computational means (Dassa *et al*., 2012; Häkkinen *et al*., 2012). The functional annotation of genes relies on the existence of current functional information *via* homology searches (Levasseur *et al*., 2014). This is not a straightforward task, as when it comes to CAZymes, they share low sequence identity and thus prediction of enzymatic activity cannot be based on their amino acid sequence alone (Henrissat and Davies, 1997). The CAZy Database provides a tremendous amount of information specifically on enzymes involved in the building and breakdown of complex carbohydrates and glycoconjugates (Cantarel *et al*., 2009). The CAZymes in the database are classified based on both their amino acid sequence and three-dimensional structure, thus allowing for differentiation of enzymes that possess different functions and belong to the same family, or of enzymes possessing the same function yet belonging to different families (Busk and Lange, 2013). A typical example is that of enzymes possessing endoglucanase (EG) activity; they have representatives in 17 different GH families (Juturu and Wu, 2014).

Functional prediction of GHs could be performed by aligning amino acid sequences of a GH family and deducing phylogenetic relationships, however, apart from the low sequence identity of proteins within a family, enzymes with the same function may have developed as a result of convergent evolution in different ancestors and would thus make the prediction unreliable; skewing the correlation between enzyme activity and phylogeny (Henrissat and Davies, 1997; Busk and Lange, 2013). A much more specific and sensitive approach for the

annotation of CAZymes is proposed by Park *et al*. (2010) which involves similarity searches against nonredundant sequences in the CAZy database much like the algorithm used by the basic local sequence alignment tool (BLAST) together with annotating CAZy family and protein family (Pfam) domains since CAZymes are usually modular proteins.

With plant biomass being an attractive renewable feedstock for the production of biofuel and the polysaccharide degradative ability of fungi, it is clear to see the need for CAZome analysis of saprophyts such as white- and brown-rot fungi, as well as species isolated from unique ecological niches. A brief overview of the CAZyme genes encoded in the genome of important CAZyme producers such as thermophiles *Thielavia terrestris*, *Myceliopthora thermophila* and the paradigm cellulase producer *Trichoderma reesei*, can be found in Table 3.1. The study of the CAZyme repertoire of plant biomass degrading fungi will greatly aid in the development of industrial enzyme cocktails for efficient degradation of plant biomass (Benoit *et al*., 2015).

Table 3.1: Distribution and comparison of the number of putative CAZyme genes between important fungal CAZyme producers

Species	GH	GT	PL	CE	AA	CBM	Total	Reference
Thielavia terrestris	212	91	4	28	58	80	473	http://www.cazy.org
Myceliophthora thermophila	195	87	8	28	50	47	415	http://www.cazy.org
Penicillium chrysogenum	225	103	9	22	22	51	432	http://www.cazy.org
Aspergillus nidulans	264	92	21	31	33	44	485	http://www.cazy.org
Trichoderma reesei	174	94	4	19	NR	41	332	Xie *et al*. (2014); Payne *et al*. (2015)

NR: Not reported; GH: Glycoside hydrolase; GT: Glycosyl transferase; PL: Polysaccharide lyase; Carbohydrate esterase; AA: Auxillary activity enzymes; CBM: Carbohydrate binding domain

Table 3.1 reveals a trend that in most organisms, GHs are the most highly represented class of CAZymes. However, not all GHs are involved polysaccharide degradation as will be discussed later in this chapter. There has been about 30 CAZymes that have been characterized from *T. reesei*, however, with CAZome analysis, there has been predictions of 44 uncharacterized enzymes that are involved in the degradation of biomass (Häkkinen *et al*., 2012). Thus, CAZome analysis is powerful bioinformatic method that reveals information on an organism's metabolic potential and nutritional lifestyle.

Phialophora alba, a thermophilic ascomycete fungus isolated from *Eucalyptus* woodchips and has been shown to possess highly active thermostable xylanases (Mosina, 2013). It was found that maximum endoglucanase activity is produced on Day 4 with two isozymes being produced with pH optima at 4 and 9 with temperature optim at 60°C (Dweba, 2013; Mbandlwa, 2013) Research and information on *Phialophora* species in general is scarce. To our knowledge, *P. alba* is only the second *Phialophora* sp. known to produce thermophilic CAZymes, and this is the first known report on the genome of *P. alba*. *Thielavia terrestris* and *Myceliopthora thermophila* are among the most well studied thermophiles with regard to their cellulolytic activity and thermostable enzymes and will thus be featured prominently in this discussion along with other well known CAZyme producers (Berka *et al*., 2011).

The aim of this phase of study was to data-mine the genome of *P. alba* and identify genes that encode EGs and other CAZymes of interest and those that could be utilised in industrial applications.

3.2 Materials and Methods

3.2.1 DNA extraction and sequencing

P. alba was grown in 100 ml malt extract broth (Sigma-Aldrich, St Louis, USA) *via* shake flask culture for 7 days at 30°C and 200 rpm. Fungal biomass was collected aseptically and

filtered through No. 1 Whatman filter paper. Fungal biomass was thereafter used to extract genomic DNA. Three genomic DNA isolation avenues had to be explored as DNA isolation from this fungus proved to be troublesome. The following methods/kits were tested in order to determine which method or kit would deliver the highest yield and best quality genomic DNA: (i) UltraClean Microbial DNA Isolation kit (MO BIO Laboratories, Carlesbad, CA, USA) according to the manufacturers instructions; (ii) Trizol reagent (Sigma-Aldrich, St Louis, USA) according to the manufacturers instructions; (iii) Soil Microbe DNA Miniprep kit (Zymo Research, Irvine, CA, USA) according to the manufacturers instructions; and (iv) the fungal DNA isolation protocol modified from Miller *et al*. (1999). The harvested fungal biomass was flash frozen with liquid nitrogen and ground to a fine powder with a mortar and pestle. One millilitre of fungal extraction buffer (see Appendix B for recipe) and 100 µl of 10% SDS were added to ~200 mg of the fungal biomass powder in a sterile 1.5 ml microfuge tube and incubated at 65°C for one hour with inversions every 20 minutes. Samples were thereafter centrifuged at 12 000 rpm for 10 minutes (Eppendorf Centrifuge 5430R, USA). The supernatant was collected and mixed with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) with gentle shaking. The mixture was thereafter centrifuged for 10 minutes at maximum speed. The aqueous layer was collected and subjected to a second round of phenol:chloroform:isoamylalcohol (25:24:1) extraction. The aqueous layers were transferred to clean microfuge tubes and subjected to an equal volume chloroform:isoamylalcohol (24:1) extraction and centrifuged as described above. Aqueous layers were treated with 0.6 volumes of isopropanol for 10 minutes at 4 °C and centrifuged at 12 000*g* for 10 minutes. The pelleted DNA was washed twice with 700 µl ice-cold 70% ethanol and dried at room temperature and thereafter resuspended in 20 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 8).

Whole genome sequencing was performed at Inqaba Biotech on the Illumina Mi-Seq Platform. The genome was assembled at Inqaba Biotech using the CLC Genomics Workbench while all further analysis was performed by the author of this dissertation using the CLC Main Workbench and online bioinformatics tools as described below.

3.2.2 Data-mining for CAZymes

Putative open reading frames (ORFs) were predicted using the CLC Main Workbench version 6.7.1 and the default ORF finding settings (Start codon: AUG; Both strands; Genetic code: Standard; Minimum length of codons: 100). All ORFs were extracted and compared to the CAZy Reference Dataset using the CAZymes Analysis Toolkit (CAT) web server opting for a strict E-value threshold of 10^{-5} (Park *et al*., 2010; Lombard *et al*., 2014). The results returned were manually curated based on Pfam domains. A strict selection criterion was applied where only those ORFs that returned hits for Pfam domains for both query and subject were carried forward to investigate the carbohydrate degradative capabilities of the fungus.

3.3 Results and Discussion

3.3.1 Isolation of genomic DNA from P. alba

Genomic DNA isolation using commercial fungal DNA extraction kits (ZR and MOBIO kits) resulted in low yields between 18.5 – 23.9 ng/µl and 260/280 absorbance ratios as low as 1.3 indicating impure DNA. Further processing could not be carried out with the Trizol reagent due to the precipitation of cell debris together with the DNA. Studies have shown that molecular techniques such as polymerase chain reaction and restriction analysis are inhibited by contaminating polysaccharides from fungal cell walls (Al-Samarrai and Schmid, 2000) thus an alternative DNA isolation protocol had be sought out. The modified method by Miller *et al*. (1999), proved to yield high concentrations of DNA from *P. alba*, ranging between 1-3

µg/µl with 260/280 and 260/230 ratios above 1.8 indicating pure DNA. All further DNA extractions from *P. alba* were carried out using this modified protocol.

The extraction of pure, high molecular weight genomic DNA from fungi is widely considered to be rather difficult (Al-Samarrai and Schmid, 2000). There are a number of protocols available, however, they tend to be biased toward specific morphological forms and groups of fungi and are thus not versatile enough to be applied to different types of fungi (González-Mendoza *et al*., 2010). Isolation of good quality and high yield of genomic DNA from *P. alba* for the purposes of whole genome sequencing proved to be extremely problematic even with the use of commercial DNA isolation kits. Thus a considerable amount of time was spent trying to obtain high quality genomic from *P. alba*.

3.3.2 Genome summary

The *P. alba* genome was assembled into 3 762 contigs with a N50 of 108 823. The complete sequenced genome was 32 132 264 bp (32.1 Mb) with a G + C content of 48.4 %. The genome size is smaller than those of *T. terrestris* and *M. thermophila* of 36.9 and 38.7 Mb respectively, with only slightly higher G + C contents of 54.7 and 51.4%, respectively (Berka *et al*., 2011). A reduced genome size is indicative of thermophilic adaptation; it is thought that transposable elements do not favour a thermophilic lifestyle and thus elements such as transposases and reverse transcriptases are often lacking in thermophilic fungi (van Noort *et al*., 2013). The thermophilic fungus and xylanase superproducer, *Thermomyces lanuginosus* has a considerably smaller genome of 23.3 Mb and G + C content of 52.1% (Mchunu *et al*., 2013). It is also hypothesised that G + C content has a positive relation correlation to thermophilicity, however, there are usually other intrinsic and extrinsic factors that lead one to question this correlation (Zheng and Wu, 2010). The total number of open reading frames (ORFs) found were 29 144 which were analysed using the CAZYmes Analysis Toolkit (CAT) and revealed the presence of 452 putative CAZyme genes.

3.3.3 The carbohydrate-active enzymes of P. alba

The overall number of putative CAZymes in *P. abla* is very similar to other well-known Ascomycete CAZyme producers (Xie *et al*., 2014; Payne *et al*., 2015; http://www.cazy.org); however, there appears to be marked differences in the distribution of the enzyme classes. The number of GHs in *P. alba* (198) are similar to those in *M. thermophila* (195) and higher than that in *Trichoderma reesei* (174), while being much lower than those found in *T. terrestris* (212), *Penicillium chrysogenum* (225) and *Aspergillus nidulans* (264). *P. alba* appears to be lacking in the classes of PLs (1), AAs (29), and CBMs (29) when compared to other fungi, while having a higher number of CEs (48).

It should be noted that the distribution of CAZymes among the enzyme classes will not give an accurate description of the lignocellulosic degradative ability of an organism, but it is rather the distribution of CAZymes among distinct families within a class (Fernandez-Fueyo *et al*.. 2012). Differences in the familial distribution of CAZymes will have an impact on how a fungus may acquire its nutrients i.e. parasitically or saprobically. Table 3.2 details the distribution of CAZyme genes among the different enzyme families (GHs, PLs and CEs) as well as CBMs.

Table 3.2: Putative carbohydrate-active enzymes present in the genome of *Phialophora alba*

CAZymes Family	Known Activity	Number of ORFs
GH 1	β-glucosidase (EC 3.2.1.21)	6
GH 3	β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37)	11
GH 5	endo-β-1,4-glucanase / cellulase; glucan endo-1,6-β-glucosidase (EC 3.2.1.75)	2
GH 10	endo-1,4-β-xylanase; endo-1,3-β-xylanase; tomatinase; xylan endotransglycosylase	2
GH 13	α-amylase (EC 3.2.1.10); oligo-α-glucosidase	9
GH 15	glucoamylase (EC 3.2.1.3)	5
GH 16	xyloglucan: xyloglucosyltransferase; keratan-sulfate endo-1,4-β-galactosidase; endo-1,3-β-glucanase; endo-1,3(4)-β-glucanase; licheninase; β-agarase; κ-carrageenase; xyloglucanase; endo-β-1,3-galactanase; β-porphyranase; hyaluronidase; endo-β-1,4-galactosidase; chitin β-1,6-glucanosyltransferase; endo-β-1,4-galactosidase	10
GH 17	glucan endo-1,3-β-glucosidase; glucan 1,3-β-glucosidase; licheninase; ABA-specific β-glucosidase; β-1,3-glucanosyltransglycosylase	1
GH 18	Chitinase (EC 3.2.1.14)	15
GH 20	β-hexosaminidase; lacto-N-biosidase; β-1,6-N-acetylglucosaminidase); β-6-SO3-N-acetylglucosaminidase	1
GH 25	Lysozyme	3
GH28	rhamnogalacturonan α-1,2-galacturonohydrolase (EC 3.2.1.173)	1
GH 29	α-L-fucosidase	2
GH 30	[reducing end] β-xylosidase (EC 3.2.1.-)	1
GH 31	α-1,3-glucosidase (EC 3.2.1.84)	8
GH 32	exo-inulinase (EC 3.2.1.80)	2
GH 35	β-galactosidase; exo-β-glucosaminidase; exo-β-1,4-galactanase; β-1,3-galactosidase	1

GH 36	α-galactosidase (EC 3.2.1.22)	6
GH 37	α,α-trehalase (EC 3.2.1.28)	2
GH 38	α-mannosidase; mannosyl-oligosaccharide α-1,3-1,6-mannosidase	1
GH 43	β-xylosidase; α-L-arabinofuranosidase; arabinanase; xylanase; galactan 1,3-β-galactosidase; α-1,2-L-arabinofuranosidase; exo-α-1,5-L-arabinofuranosidase; [inverting] exo-α-1,5-L-arabinanase; β-1,3-xylosidase	10
GH 47	α-mannosidase (EC 3.2.1.113)	19
GH 51	α-L-arabinofuranosidase (EC 3.2.1.55)	2
GH 53	endo-β-1,4-galactanase (EC 3.2.1.89)	2
GH 55	exo-β-1,3-glucanase; endo-β-1,3-glucanase	5
GH 63	processing α-glucosidase (EC 3.2.1.106)	3
GH 65	α,α-trehalase; maltose phosphorylase; trehalose phosphorylase; kojibiose phosphorylase; trehalose-6-phosphate phosphorylase; nigerose phosphorylase; 3-O-α-glucopyranosyl-L-rhamnose phosphorylase; 2-O-α-glucopyranosylglycerol: phosphate β-glucosyltransferase	1
GH 67	α-glucuronidase (EC 3.2.1.139)	1
GH 71	α-1,3-glucanase	9
GH 72	β-1,3-glucanosyltransglycosylase	11
GH 75	chitosanase	1
GH 76	α-1,6-mannanase	18
GH 78	α-L-rhamnosidase	1
GH 81	endo-β-1,3-glucanase	1
GH 88	d-4,5-unsaturated β-glucuronyl hydrolase	1
GH 89	α-N-acetylglucosaminidase	1
GH 92	mannosyl-oligosaccharide α-1,2-mannosidase; mannosyl-oligosaccharide α-1,3-mannosidase; mannosyl-oligosaccharide α-1,6-mannosidase; α-mannosidase; α-1,2-mannosidase; α-1,3-mannosidase; α-1,4-mannosidase; mannosyl-1-phosphodiester α-1,P-mannosidase	12
GH 105	d-4,5-unsaturated β-glucuronyl hydrolase	1

GH 109	α-N-acetylgalactosaminidase	1
GH 125	exo-α-1,6-mannosidase	1
GH 127	β-L-arabinofuranosidase	1
GH 128	β-1,3-glucanase	2
GH 132	Activity on β-1,3-glucan; activity on laminarioligosaccharides	5
CE 1	acetyl xylan esterase; cinnamoyl esterase; feruloyl esterase; carboxylesterase; S-formylglutathione hydrolase; diacylglycerol *O*-acyltransferase; trehalose 6-*O*-mycolyltransferase	1
CE 4	Polysaccharide deacetylase	9
CE 5	Cutin hydrolase	2
CE 10	Acetylcholinesterase (EC 3.1.1.7); brefeldin A esterase (EC 3.1.1.-)	37
PL 7	poly(β-mannuronate) lyase / M-specific alginate lyase; α-L-guluronate lyase / G-specific alginate lyase; poly-(MG)-lyase / MG-specific alginate lyase	1
CBM 1	Cellulose binding	1
CBM 3	Cellulose binding	1
CBM 12	Chitin binding	1
CBM 13	Mannose or xylan binding – specificity not established	4
CBM 18	Chitin binding	4
CBM 20	Starch binding	10
CBM 21	Starch binding	2
CBM 24	α-1,3-glucan (mutan)-binding	1
CBM 32	Galactose, lactose, polygalacturonic acid, (β-D-galactosyl-1,4-β-D-N-acetylglucosamine	1
CBM 48	Glycogen binding	4

Activities in red indicate those hits that provided positive identification of a specific Pfam domain or E.C number i.e. those having functional predictions.

Following the trend of many other fungi, such as those listed in Table 3.1 as well as fungi such as *Phytopthora* spp. (93), *Laccaria bicolori* (Veneault-Fourrey *et al*., 2014) and many others, the GH class of enzymes were the most highly represented with the 198 putative genes in *P. alba* being distributed in 43 families. CEs, PLs and CBMs each had 49, 1 and 29

genes arranged in 4, 1 and 10 families, respectively. *Phytopthora ultimum* possess 183 GHs distributed in 28 families and had 6, 29 and 57 PLs, CEs and CBMs, respectively, distributed among 5, 3 and 7 families respectively (Brouwer *et al*., 2014). The CAZymes listed in Table 3.2 however, are not all involved in the breakdown of polysaccharides. Some CAZymes may be involved in protein glycosylation and energy storage and recovery (Zhao *et al*., 2014). Figure 3.1 Illustrates this fact by broadly distinguishing between the CAZymes based on their substrates.

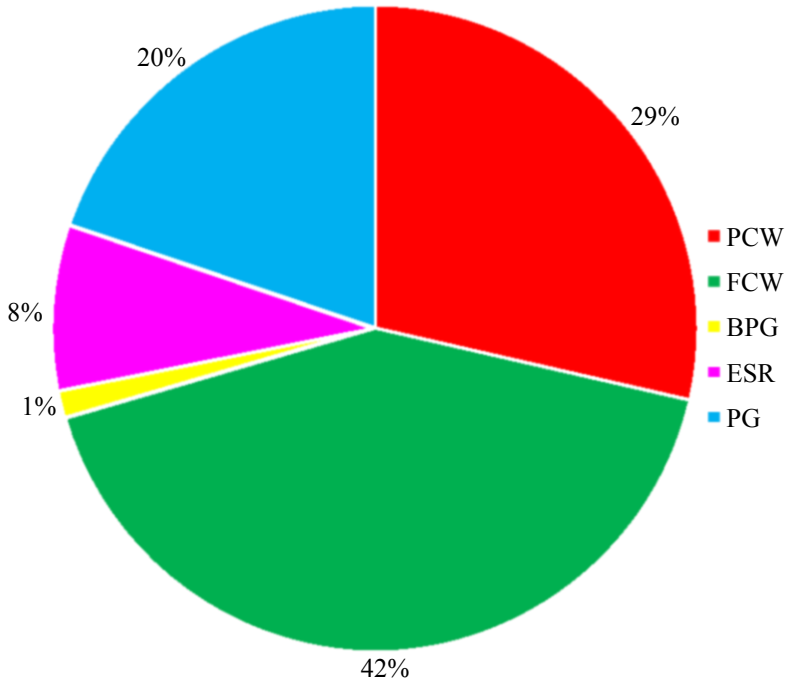


Figure 3.1 Functional distinction of putative CAZymes from *P. alba* genome (PCW: plant cell wall; FCW: fungal cell wall; BPG: bacterial peptidoglycan, ESR: energy storage and recovery; PG: protein glycosylation). Total count excludes PLs, CBMs and those CAZymes whose exact function has not yet been determined (GHs: 92, 109, 125, 127, 128, 132 and CE10)

Twenty nine percent of the of the CAZymes in *P. alba* are not directly involved in the degradation of polysaccharides for nutrient acquisition but rather have cellular functions such as GH 8, 38, 47, 63 and 92 being involved in protein glycosylation and GH 15, 31, 32, 37 and 65 being involved in energy storage and recovery. The majority of GHs appear to have

activity on fungal (42%) and plant cell walls (29%). The of percentage GH genes related to plant polysaccharide degradation is much lower than those observed for *Aspergillus* spp. which is between 58-66% (Benoit *et al*., 2015). Figure 3.2 better illustrates the relationship between the CAZymes of *P. alba* and their substrate specificities.

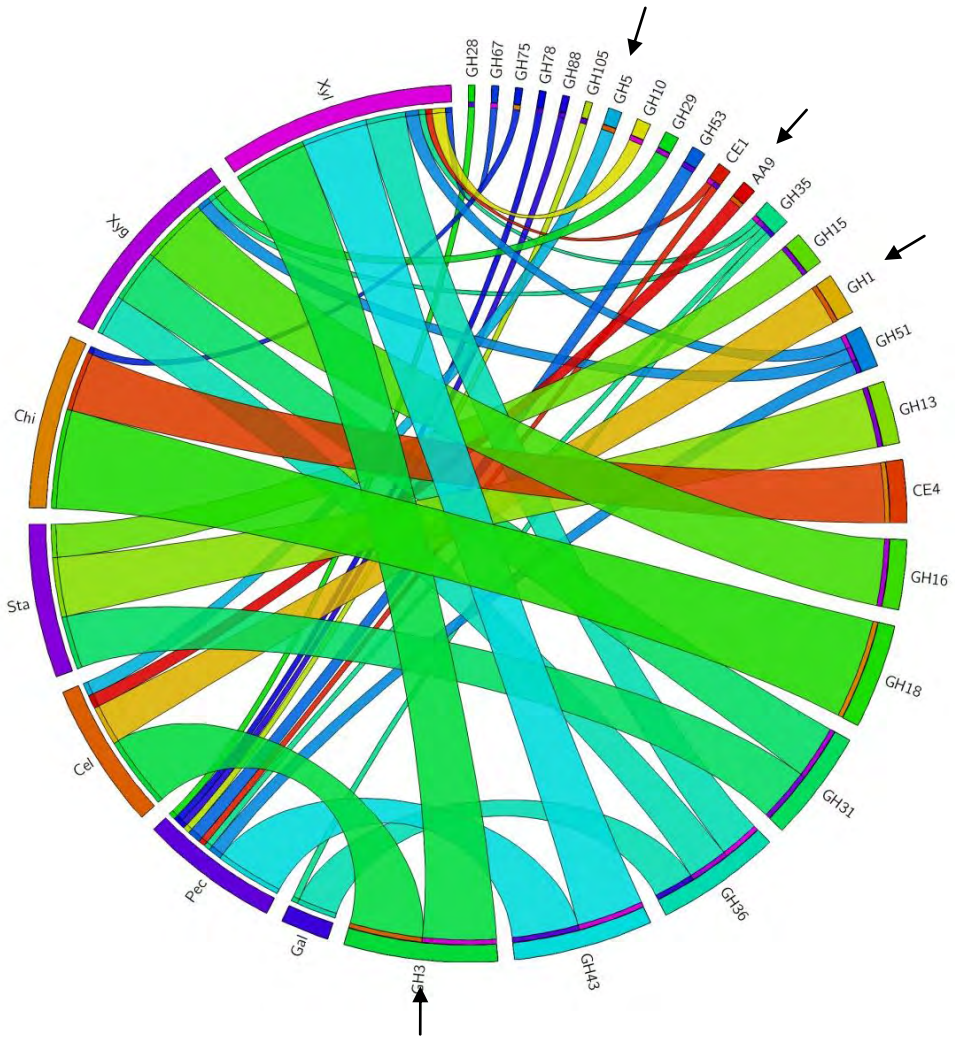


Figure 3.2 Chord diagram illustrating the substrate specificities of carbohydrate active enzyme families from *P. alba* (Gal: Galactomannan; Pec: Pectin; Cel: Cellulose; Sta: Starch; Chi: Chitin; Xyg: Xyloglucan; Xyl: Xylan) (Particular attention [black arrows] is drawn to those families of GHs that potentially hydrolyse cellulose)

Figure 3.2 not only illustrates the relative abundance of enzymes involved in the degradation of a specific substrate (ribbon width correlates with the number of ORFs per enzyme [Table 3.2]), but also the ability of an enzyme or enzyme family to be active on multiple substrates.

Examples include GH35 which is able to act on xylan, xyloglucan, galactomannan and pectin while GH31 can act on both starch and xyloglucan. It is clear to see a greater diversity and abundance of enzymes involved in the degradation of plant cell wall polysaccharides such as xylan, xyloglucan and pectin as opposed to those involved in cellulose degradation. Subsequent sections will focus on specific substrates and the ability of *P. alba* to degrade them.

3.3.4 Cellulose degrading capacity

Although there appears to be a large number of enzymes involved in cellulose degradation (Figure 3.2), the majority of cellulose active enzymes in *P. alba* belong to GH families 1 and 3 which are β-glucosidases that cleave cellobiose to glucose. It is important to note that β-glucosidases serve other purposes beside cellulose degradation such as host-pathogen interactions, and cell signalling (Krisch *et al*., 2010; Battaglia *et al*., 2011). Thus, high numbers of β-glucosidase in a genome may obscure the count of the number of genes related to cellulose degradation (Benoit *et al*., 2015). It is likely that *P. alba* is a potential opportunist that benefits from degarders that possess both EGs and β-glucosidases as opportunists tend to only possess β-glucosidases (Do *et al*., 2014).

P. alba appears to be a poor cellulose degrader as it only possesses a single EG belonging to GH family 5. This is supported by a concurrent study on the lignocellulolytic arsenal of *P. alba* carried out in our research group by Mbandlwa (2013), who observed poor EG activity on Congo red plate screening assays. The degradation of cellulose is brought about by the synergistic efforts of EGs, CBHs and β-glucosidases. Of the 29 GH families that are known to be involved in the degradation of plant biomass, GH families 1, 3, 5, 6, 7, 12, 45 and 74 are among the important GHs involved in cellulose degradation (Zhao *et al*., 2014; Payne *et al*., 2015). Aside from GH families 1, 3 and 5, *P. alba* is lacking in the other families that are required for efficient cellulose degradation. EGs belonging to families 5, 7, 12 and 45 are

found in greater numbers in fungi such as *Fusarium graminearum*, *Podospora anserine* and *Aspergillus oryzae* at 20, 22 and 20, respectively (van den Brink and de Vries, 2011).

It is also interesting to note that *P. alba* is deficient in CBHs belonging to GH families 6 and 7 in a similar fashion to *Rhizopus oryzae* (Battaglia *et al*., 2011). Fungi such as *Aspergillus niger*, *Trichoderma reesei* and *Neurospora crassa* have 4, 3 and 8 CBHs, respectively (van den Brink and de Vries, 2011). This apparent reduction in the set of cellulose active enzymes in *P. alba* is further reinforced by the presence of only 2 putative CBMs (CBM 1 and CBM 3) with cellulose binding capabilities.

The lack of a complete set of cellulose degrading enzymes may be indicative of an organisms adaptation to a specific mode of cellulose degradation. *T. terrestris* and *M. thermophila* may be considered all purpose decomposers in terms of their ability to degrade plant cell walls, and have similar complements of GHs to *T. reesei*, however, there is an expansion of the GH 61 family in these two thermophiles (Berka *et al*., 2011). GH 61 was originally classified as a weak EG but has subsequently been reclassified as AA9, a copper-dependant lytic monooxygenase that cleaves cellulose chains at various carbons (http://www.cazy.org). *P. alba* possess only two AA9 family members but has 16 AA3 members as compared to 12 and 7 in *T. terrestris* and *M. thermophila,* respectively. AA3 family enzymes are classed as cellobiose dehydrogenases, of which the most widely accepted activity is hydroxyl radical cleavage of cellulose (Langston *et al*., 2011). *Pycnoporus cinnabarinus* possesses a complete a of ligninolytic enzymes of which cellobiose dehydrogenase is an important part of as carbohydrates and lignin are closely connected into plant cell walls (Levasseur *et al*., 2014). Thus, an apparent deficiency in cellulose degrading GH families does not preclude *P. alba* from possessing cellulolytic activity, provided the expression of these auxiliary activity enzymes are efficiently induced when required.

3.3.5 Hemicellulose degrading capacity

Endoxylanases are largely represented in GH families 10 and 11 with GH10 having a higher substrate specificity than GH11 (Zhao *et al*., 2014). *P. alba* possesses only two GH10 as opposed to *T. terrestris* and *M. thermophila* which have six GH10 and five GH11 and four GH10 and eight GH11, respectively. It is interesting to note that Mosina (2013) reported *P. alba* as being a highly efficient xylan degrader that produces six xylanase isozymes as determined by zymogram analysis. This is not an uncommon phenomenon, as there are many cases in which there is a discrepancy between the number of xylanase genes present and the number of xylanases expressed, as is the case with *Penicillium oxalicum* GZ-2 which expresses eight active xylanases that are encoded by six different xylanase genes (Liao *et al*., 2015). This multiplicity may be brought about by the differential splicing of primary xylanase mRNA transcripts (post-translational modification) or as a result of proteolytic cleavage or differential glycosylation (post-translational modifications) (Juturu and Wu, 2014).

Other enzymes required for complete xylan and hemicellulose degradation include α-L-arabinofuranosidase with representatives in GH families 3, 10, 43, 51, 54 and 62 and β-xylosidase with representatives in GH families 3, 43 (Veneault-Fourrey *et al*., 2014). Efficient xylan degradation is also dependent on the activity of acetylxylan esterases (CE 1-7) (van den Brink and de Vries, 2011). It has been reported that fungal α-L-arabinofuranosidase are found mainly in GH families 51 and 54 (Zhao *et al*., 2014). *P. alba* lacks GH54 representatives and has only two α-L-arabinofuranosidases in family GH51 following a similar trend of over 47 fungi studied by Zhao and colleagues (2014). However, *P. alba* may possess α-L-arabinofuranosidase representatives in GH43 and possibly in GH3. β-Xylosidase appears to be well represented in *P. alba* with positively identified genes in GH3 and potential representation in GH43. *P. alba* possess 10 potential acetyl esterases belonging to CE families 1 and 4.

Xyloglucanases are not abundant in fungi (Segato *et al*., 2014), however, *P. alba* possess 10 GH16 genes which could potentially encode enzymes with specificity for xyloglucan. GH16 enzymes may be involved in the processing of the fungal cell wall or be involved in the hydrolysis of β-1,3-glucans (Veneault-Fourrey *et al*., 2014). Hartl *et al*. (2011) showed that an endo-β-1,3-glucanase from *A. fumigatus* is capable of hydrolysing β-1,3-glucan as well as carboxymethyl cellulose. There is also a patent out that describes an increase in the rate of cellulose degradation when CBHs and EGs are used in conjunction with fungal GH16 polypeptides (Corporation, 2013). The substrate specificity of the GH16 enzymes in *P. alba* can only be determined if tested experimentally. *P. alba* also possess five GH55 which together with a single GH81 representative display β-1,3-glucanase activity and is able to degrade the plant polysaccharide callose (a polymer of β-1,3-glucan) (Zhao *et al*., 2014). It is evident that the xylan degrading ability of *P. alba* reported by Mosina (2013) and Mbandlwa (2013) is well supported and it is clear to see that *P. alba* possesses a full complement of hemicellulose degrading enzymes.

3.3.6 Other plant polysaccharide degrading capabilities

Pectin is a complex heterogenous polysaccharide and is mainly composed of galacturonans and is degraded by a number of pectinases including pectate lyase, rhamnogalacturonases, pectin methylesterases and pectin acetylesterases (Abbott and Boraston, 2008). The *P. alba* genome contains 20 putative enzymes that are involved in pectin degradation with the majority of enzymes belonging to GH families 43 and CE4. *P. alba* does not possess pectin, pectate and rhamnogalacturonan lyases (PL: 1, 2, 3, 4, 9 or 11) but unlike *Laccaria bicolour* (Veneault-Fourrey *et al*., 2014), it does possess α-L-arabinofuranosidase (GH51) and α-L-rhamnosidase (GH78). *P. alba* appears to possess a diverse set of enzymes responsible for the degradation of branched pectins (albeit very few of each), however, it lacks in polygalacturonases which are required for degradation of the main pectin chain (Battaglia *et*

al., 2011). *P. cinnabarinus* also appears to be very limited with genes encoding pectinolytic enzymes, having very few GH28 members and no representative families having α-L-arabinofuranosidase activity (Levasseur *et al*., 2014). It can thus be concluded that *P. alba* is not an efficient degrader of pectin and is unlikely to be a plant pathogen as plant pathogens tend to have a higher amounts of pectin degrading enzymes (Brouwer *et al*., 2014; Zhao *et al*., 2014; Veneault-Fourrey *et al*., 2014).

P. alba has a considerable number of enzymes involved in the hydrolysis of starch (22) present, including amylases in GH families 13 (9), 15 (5) and 31 (8) and accompanied by 22 starch CBMs (CBM20 (10) and CBM21 (2)). *R. oryzae* only has a total 13 starch degrading enzymes belonging to GH families 13 (4), 15 (6) and 31 (3) with eight CBM21-containing proteins and no CBM20 modules being present (Battaglia *et al*., 2011). It is assumed that the presence of these enzymes are required for the metabolism of low levels of starch found in ground wood (Hori *et al*., 2013). One would be tempted to predict the type of wood that *P. alba* prefers to associate with seeing as it was isolated from woodchips. Plants differ with regard to their cell wall components especially in the proportions of hemicelluloses and pectin (King *et al*., 2011). However, there is no evidence that suggests that the diversity or number of enzymes present has any correlation to the preference of a fungus to any specific wood or plant species (Zhao *et al*., 2014).

3.3.7 Chitin degrading capacity

The high representation of members within GH family 18 in *P. alba* cannot be ignored and this may reveal alternate means by which this fungus obtains nutrients. Chitin is a major component of fungal cell walls (Xie *et al*., 2014) and it is possible that *P. alba* may be mildly mycoparasitic as it possess a high number of GH18 chitinases (15) in comparison to xylanases and cellulases; as well as a moderate number of CE4 (9) which may encode chitin

deacetylases which transform chitin into chitosan (Veneault-Fourrey *et al*., 2014). However, one cannot be certain of this mode of nutrient acquisition as chitinases have been shown to have diverse roles in a fungus' own cell wall remodelling, hyphal growth and autolysis (Gortari and Hours, 2008; Battaglia *et al*., 2011).

3.4 Conclusion

This genomic survey of the putative CAZymes of *P. alba* provides insight into this thermophile's lignocellulosic degradative abilities. It also highlights the differences with regard to the modes of cellulose and xylan degradation, i.e. with the aid of GHs or AAs. These results show that *P. alba* is much more efficient at hydrolysing xylan than cellulose and supports the biochemical characterization studies carried out by fellow members in this research group. Future studies should focus on transciptomic analysis and a more rigorous bioinformatic annotation of the genome of *P. alba* which could reveal more insight into the nutritional strategy of this fungus.

CHAPTER 4

Cloning and expression of the endoglucanase gene from *Phialophora alba* isolated from *Eucalyptus* spp. woodchips

4.1 Introduction

A number of fungi produce cellulolytic enzymes with the majority of cellulases being produced by *Aspergillus* and *Trichoderma* strains (Ja'afaru, 2013; Narra *et al*., 2014). However, most of their enzymes display optimal activity under mesophilic conditions with short-lived thermostability making them inefficient in their industrial applications (Zhao *et al*., 2012a). Enzyme thermostability is an essential property of EGs involved in the saccharification of lignocellulosic biomass as hydrolysis is much more efficient at elevated temperatures (Liu *et al*., 2011).

The ability of microbes to produce large quantities of enzymes by fermentation is a great advantage; however, cellulase yields are dependent on a number of factors such as temperature, pH, incubation times, carbon and nitrogen sources as well as cations (Gautam *et al*., 2011). Therefore, optimization of fermentation conditions is required to achieve acceptable yields. Furthermore, even though fungi are known to secrete cellulolytic enzymes, difficulties have been observed in downstream processes with a variety of fermentation methods used for the production of EGs and other cellulases (Bien-Cuong *et al*., 2009). In addition to this, the specific activity of cellulases are at least two times lower than that of other glycoside hydrolases and as such native cellulolytic organisms are required to produce cellulases in high titres in order to efficiently degrade cellulose, however, this places a metabolic burden on the organism (Wilson, 2008; Zhang and Lynd, 2004; Zhang and Lynd, 2005; Vinuselvi and Lee, 2012). Molecular biology provides a means of expressing these cellulases in a host strain that is capable of producing and secreting large amounts of

recombinant protein (Bien-Cuong *et al*., 2009). Heterologous expression of proteins is an attractive alternative allowing for controlled expression and ease of production and purification of EGs from large scale fermentations (Liu *et al.,* 2011).

Pichia pastoris is an expression host that is able to express high levels of foreign proteins even if they are toxic to the cell (Macauley-Patrick *et al*., 2005). It has been reported that *P. pastoris* is capable of expressing 2.31-2.62-fold higher amounts of α-amylase when compared to wild type expression in *Bacillus licheniformis* (Wang *et al*., 2014). Várnai and colleagues (2014) showed that upon optimizing fermentation condition of recombinant *Pichia* strains, recombinant EG yields as high as between 3-5 g/l can be obtained.

The present phase of work is focused on the cloning and expression of an EG from *P. alba*, a thermophilic fungus isolated from *Eucalyptus* woodchips which has been reported to produce carbohydrate active enzymes. A previous study by Mosina (2013) revealed that *P. alba* produces highly active thermostable xylanases, and a preliminary study on the crude extract of *P. alba* confirmed that it also produces EGs that could maintain relatively high activity under thermophilic conditions (Mbandlwa, 2013). To date there has only been one other report of CAZymes being produced by a *Phialophora* sp. *Phialophora* sp. G5, isolated from acidic waste water has been reported to produce a thermophilic/stable: (i) GH10 xylanase (Zhang *et al*., 2011); (ii) GH5 EG (Zhao *et al*., 2011); (iii) GH45 EG (Zhao *et al*., 2012b); (iv) GH6 CBH (Zhao *et al*., 2012b) and (v) GH7 β-1,4-glucanase (Zhao *et al*., 2012c).

In order to get a clear understanding of the decisions made during this study, one has to take into account all studies involving *P. alba* including organisms carbohydrate degradative capabilities as discussed in Chapter 3. Concurrent studies by Mbandlwa (2013) and Dweba (2013) focused on the characterization of the crude lignocellulosic enzymes, and the characterization and partial purification of the EG from *P. alba*, respectively. Mbandlwa

showed that *P. abla* possesses relatively weak EG activity *via* plate screening assays as well as secreting very low amounts (0.3 U/ml) under shake flask culture. Thus Dweba sought to partially purify and characterize the *P. alba* EG. Purification of the EG and putative isozymes proved to be difficult with 60% remaining in the supernatant following 80% ammonium sulphate precipitation and thus the supernatant was used for characterization. It was found that the EG had optimum activity at pH 4 and 60°C, retaining 70% and 60% activity at optimum pH and temperature, respectively, after 90 minutes incubation. Liu *et al*. (2011) encountered many difficulties during the production of EGs from *A. fumigatus* Z5 *via* fermentation and found that this limited scale up for the production of large quantities of the enzyme and thus chose to clone and express the EGs in a *Pichia* expression host to overcome these limitations. With low levels of EG being produced by *P. alba* and difficulties encountered with initial purification methods, cloning and heterologous expression seemed like a more feasible option than laborious fermentation optimizations.

4.2 Materials and Methods

4.2.1 Strains, vectors and media used

Vectors used include pJET1.2/blunt provided in the CloneJET PCR Cloning Kit (ThermoFisher Scientific, Massachusetts, USA) and pTZ57R/T provided in the InsTAclone PCR Cloning Kit (ThermoFisher Scientific, Massachusetts, USA) for cloning into *E. coli* DH5α and HB101. Expression vectors and hosts used were pPIC9 and pET-22b for expression in *Pichia pastoris* GS115 and *E. coli* BL21(DE3), respectively.

P. alba was cultivated in the following media: (i) malt extract (ME) broth (Sigma-Aldrich, St Louis, USA) for genomic DNA isolation and (ii) cellulase inducing medium 7 g $K_2HPO_4.H_2O$, 3.45 g $NaH_2PO_4.H_2O$, 2 g $(NH_4)_2SO_4$, 0.4 g Yeast extract, 0.005 g $CaCl_2$, 0.0125 g $MgSO_4.7H_2O$ per litre and adjusted to pH with HCl (Gomes *et al*., 2000)

supplemented with (a) 1% carboxymethyl cellulose (CMC), (b) avicel, (c) α-cellulose or (d) saw dust (*Eucalyptus nitens*) for RNA isolation and cDNA synthesis. Media used for transformations and expression in *E. coli* hosts include: Luria-Bertani (LB) broth and agar (1% Tryptone, 0.5% Yeast Extract, 1% NaCl) for cultivation of host strains and LB agar plates and broth supplemented with 100 μg/ml ampicillin (LB-amp) for the selection and growth of transformants. Expression in *E. coli* BL21(DE3) was induced by supplementing the appropriate media with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Media used for *Pichia* include: yeast peptone dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% dextrose (glucose)); minimal dextrose (MD) medium (1.34% yeast nitrogen base (YNB) (Sigma-Aldrich, St Louis, USA), 4×10^{-5}% biotin, 2% dextrose); minimum methanol (MM) medium (1.34% YNB, 4×10^{-5}% biotin, 0.5% methanol), buffered glycerol-complex (BMGY) or buffered methanol-complex (BMMY) (1% yeast extract, 2% peptone, 100 mM potassium phosphate (pH 6, 1.34% YNB, 4×10^{-5}% biotin, 1% glycerol or 0.5% methanol) and yeast peptone methanol (YPM) (1% yeast extract, 2% peptone, 1% methanol). Media was prepared according the instructions in the *Pichia* Expression Kit manual (Invitrogen, California, USA). Agar plates of the media listed where prepared by adding 12 g bacteriological agar per litre prior to autoclaving when required.

4.2.2 DNA isolation

P. alba was grown in 100 ml ME broth *via* shake flask culture for 7 days at 30°C and 200 rpm. Genomic DNA was isolated using the modified protocol from Miller *et al*. (1999) as described in Chapter 3. Plamid DNA was isolated using the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, Massachusetts, USA) according to the manufacturers instructions.

4.2.3 RNA isolation and cDNA synthesis

Due to the low amounts of secreted EG observed by Mbandlwa (2013), attempts were made to induce greater expression of EG genes for successful cDNA synthesis by cultivating *P. alba* in a variety of cellulose sources. To obtain complementary DNA, *P. alba* was grown in cellulase inducing medium at 50°C with constant agitation at 200 rpm for five to six days. Total RNA was isolated using the SV Total RNA Isolation System (Promega, Wisconsin, USA) and cDNA and second strand synthesis was carried out using the Maxima H Minus Double-Standed cDNA synthesis kit (ThermoFisher Scientific, Massachusetts, USA) with primers Oligo$(dT)_{18}$ or a gene specific primer set (Table 4.1) according to the manufacturers instructions. Total RNA was also subjected to poly(A) RNA enrichment with the NucleoTrap® mRNA spin columns (Macherey-Nagel, Düren, Germany) according to the manufacturers instructions. Total RNA, poly(A) mRNA and cDNA was stored at -80°C until required. cDNA libraries were constructed using vectors pJET1.2/blunt or pTZ57R/T or the cDNA was used as template DNA for PCR amplification of EG genes.

4.2.4 PCR amplification of endoglucanase gene

Genomic DNA or cDNA was used as the template for PCR amplification of an EG gene or fragment using the primers listed in Table 4.1. Primers set F10EgF and F10EgR; Eg-GH5-f and Eg-GH5-r and Eg5-F and Eg5-R were based on the GH family 5 EG of *Phialophora* G5 (Zhao *et al*., 2011) while the degenerate primer sets EgGH45-f and EgGH45-r and CBHII-f and CBHII-r for GH family 45 EG and GH family 6 CBH, respectively, were primers that were used to clone two novel thermostable cellulases from *Phialophora* G5 (Zhao *et al*., 2012b). PCR reactions were optimized for $MgCl_2$ concentrations and annealing temperatures when required. Optimization was performed with SuperTherm™ Taq DNA Polymerase while amplification for the purposes of cloning was performed with a high fidelity polymerase, either MyFi™ DNA Polymerase (Bioline, London, UK), Phusion High Fidelity

DNA Polymerase (New England Biolabs, Massachusetts, USA) or *Pfu* DNA Polymerase (ThermoFisher Scientific, Massachusetts, USA).

Table 4.1: List of primers used to amplify endoglucanase genes or fragments

Primers	Sequences (5′ → 3′)	Size (bp)
Primers used prior to whole genome sequencing		
F10EgF	GATCGAATTCATGATGCGCTCACTGTTACTTTC	33
F10EgR	CTAGAAGCTTTTACGAGGACCGGGCGAT	28
Eg-GH5-f	GCCGGNTTYGAYTTYGG	17
Eg-GH5-r	GTCGTGNGGYTCRTTCAT	18
Eg5-F	GCAGGGTTCGACTTCGGA	18
Eg5-R	GTCGTGCGGTTCGTTCAT	18
EgGH45-f	ACCCGYTAYTGGGAYTGYTG	20
EgGH45-r	GCACSCGTYAAARATNCCNACNCC	24
CBHII-f	TTACCAGAYAGRGAYTGYGC	20
CBHII-r	CACATTTGTCGCRAYNCCYC	20
Primers used after whole genome sequencing		
GH5_ORF_F	ATGGCGTCGCCATTGATATC	20
GH5_ORF_Flank_F	GGTAAGCCATAATGGCGTCG	20
GH5_ORF_R	TCAAGGACCCTGCGTGACCT	20
Exp_GH5_ORF_F	CGAT*GAATTC*ATGGCGTCGCCATTGATATC	30
Exp_GH5_ORF_R	ATGC*GCGGCCGC*TCAAGGACCCT	23
5′AOX1	GACTGGTTCCAATTGACAAGC	21
3′AOX1	GCAAATGGCATTCTGACATCC	21

Y = C/T, R = A/G, N = A/T/G/C; Nucleotides in italics indicate restriction sites

Following whole genome sequencing and data-mining as described in Chapter 3, primers GH5_ORF_F, GH5_ORF_Flank_F and GH5_ORF_R were constructed based on the EG gene identified. Exp_GH5_ORF_F and Exp_GH5_ORF_R were designed to contain restriction sites for *Eco*RI and *Not*I respectively, for ease of cloning into the multiple cloning sites of both pPIC9 and pET-22b expression vectors such that they would be in frame and under the control of their respective promoters for induction. Primers 5′AOX1 and 3′AOX1 were used to confirm integration of the gene into the *Pichia pastoris* GS115 genome.

4.2.5 Sequence and structure analysis

DNA sequencing of inserts in recombinant vectors was performed either at Inqaba Biotech or at the Central Analytics Facilities (CAF) at Stellenbosch University. Sequence assembly was performed using CLC Bio Genomic Workbench. The nucleotide sequence was analysed using BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify sequence similarities. Introns and exons were predicted using the online software FGENESH (http://www.softberry.com/berry.phtml). Protein sequence analysis was performed using the Pfam (http://pfam.xfam.org/) and InterProScan (http://www.ebi.ac.uk/interpro/interproscan.html) websites to determine protein family and domains. The tertiary structure of the catalytic domain was predicted online using Phyre^2 web portal (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The number of N-glycosylation site were predicted online using NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/).

4.2.6 Cloning and expression of endoglucanase gene

Restricted plasmids, fragments, and amplicons of interest were either extracted from agarose gels using the GeneJET Gel extraction kit (ThermoFisher Scientific, Massachusetts, USA) or purified using the GeneJET PCR Purification kit (ThermoFisher Scientific, Massachusetts, USA) or the DNA Clean and Concentrator™ kit (Zymo Research, Irvine, CA, USA). Fragments of interest were ligated into cloning vectors: pJET1.2/blunt or pTZ57R/T and expression vector: pPIC9 for expression in *P. pastoris* GS115 or pET-22b for expression in *E. coli* BL21(DE3). Ligations were performed using a vector to insert ration of 1:3 and incubation was at 4°C overnight.

4.2.6.1 Transformation of *E.coli* DH5α and *E. coli* HB101

Chemically competent *E.coli* DH5α was prepared by calcium chloride treatment as described by Sambrook and Russel (2000). Prior to expression in either yeast or bacterial host, the recombinant cloning and expression vectors were transformed into the chemically competent *E. coli* DH5α or HB101 *via* heat shock treatment. Positive clones were selected for on LB-amp plates. Clones harbouring recombinant vectors were confirmed by isolating plasmid using the GeneJET Plamid Miniprep Kit and restriction analysis or by colony PCR with gene specific primers.

4.2.6.2 Expression in *Pichia pastoris* GS115

Electrocompetent *P. pastoris* GS115 was prepared by treating cells with water and 1 M sorbitol as described in the *Pichia* Expression Kit manual. Recombinant pPIC9 was linearized with *Sal*I, *Bgl*II or *Not*I and used for transformation of *Pichia*. Transformation was carried out by mixing 10 µg of linearized plasmid with 80 µl of compentent *Pichia* cells and subjected to electroporation (Gene Pulser Xcell™ Electroporation System, Bio-Rad, California, USA) using the preset protocol for transformation of *Saccharomyces cerevisiae* as recommended by the *Pichia* Expression Kit manual. *Pichia* cells were plated on MD plates and incubated at 30°C until colonies appeared, usually after 48 hours. The recombinant *Pichia* phenotype (Mut^+ or Mut^s) was determined by comparing growth of the clone on minimal dextrose and minimal methanol plates. Phenotypes and integration of the gene into the genome were confirmed by PCR analysis with the AOX1 primer set.

Expression was carried out according to the expression protocol detailed in the *Pichia* Expression Kit manual for the Mut^+ phenotype. Expression was attempted using different media (MM, BMMY and YPM). Expression was also initially investigated at the recommended 0.5% methanol as well as 1-2% methanol as done by Pham *et al*. (2011). Five

hundred microlitre aliquots were taken every 24 hours and screened for EG activity and subjected to SDS-PAGE analysis. When required, aliquots were concentrated 10 fold *via* centrifugation through an Amicon Centifugal Filter (Millipore, Massachusetts, USA) with a molecular weight cut-off of 50 kDa (based on theoretical molecular weight).

4.2.6.3 Expression in *E. coli* BL21(DE3)

Chemically competent *E.coli* BL21(DE3) was prepared by calcium chloride treatment as described by Sambrook and Russel (2000). Recombinant pET-22b was transformed into chemically competent *E. coli* BL21(DE3) and positive clones were selected for on LB-amp agar plates. Clones harbouring recombinant vectors (Table 4.2) were confirmed by colony PCR with gene specific primers. The clones harbouring the gene were streaked onto LB-amp supplemented with 1% CMC plates and 1 mM IPTG and screened for EG activity after 24 hours by plate screening assays. Clones were also grown overnight in LB-amp broth and expression was induced with 1 mM IPTG for six hours. Thereafter the cells were ruptured *via* sonication and centrifuged at maximum speed for 2 minutes (Eppendorf Centrifuge 5430R, USA). The supernatant was collected and assayed for EG activity as well as subjected to SDS-PAGE analysis.

Table 4.2: Recombinant vectors and strains generated during the course of the study

Constructed Recombinant Vectors		
Recombinant vector	Plasmid	Insert
Plasmid G1	pTZ57R/T	*EgGH5* gene
Plasmid cDNA2	pTZ57R/T	1200 bp amplicon from PCR
Plasmid F1	pTZ57R/T	*EgGH5* gene + upstream flanking region
Plasmid F2	pTZ57R/T	*EgGH5* gene + upstream flank region
Plasmid F3	pTZ57R/T	*EgGH5* gene + upstream flanking region
Plasmid F4	pTZ57R/T	*EgGH5* gene + upstream flanking region
pPIC9_EgGH5[a]	pPIC9	*EgGH5* gene amplified from Plasmid G1
pPIC9_13_Eg_GH5	pPIC9	*EgGH5* gene
pET_B1	pET-22b	*EgGH5* gene
pET_B3	pET-22b	*EgGH5* gene
pET_B4	pET-22b	*EgGH5* gene
pET_B5	pET-22b	*EgGH5* gene
Constructed Recombinant Strains		
Recombinant strain	Host	Recombinant vector
*Pichia*_EgGH5_1[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_EgGH5_2[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_EgGH5_3[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_EgGH5_4[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_EgGH5_5[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_EgGH5_6[b]	*P. pastoris* GS115	pPIC9_EgGH5
*Pichia*_S1[c]	*P. pastoris* GS115	pPIC9_13_Eg_GH5
*Pichia*_S2[c]	*P. pastoris* GS115	pPIC9_13_Eg_GH5
*Pichia*_B2[c]	*P. pastoris* GS115	pPIC9_13_Eg_GH5
eB1[d]	*E. coli* BL21(DE3)	pET_B1
eB3[d]	*E. coli* BL21(DE3)	pET_B3
eB4[d]	*E. coli* BL21(DE3)	pET_B4
eB5[d]	*E. coli* BL21(DE3)	pET_B5

[a] Plasmid containing *EgGH5* gene with point mutations; [b]*Pichia* clones containing plasmid with gene that has point mutations; [c] *Pichia* clones containing recombinant plasmid linearized with either S: *Sal*I or B: *Bgl*II; [d] Recombinant *E. coli* BL21(DE3) clones.

4.2.7 Enzyme activity assays

EG activity was determined *via* the DNS assay with slight modification (Bailey *et al*., 1992) and Congo red plate screening assays (Teather and Wood, 1982) as described in Chapter 2.

Both assays were performed at 50°C. The Congo red plate screening assay was also performed at 30°C.

4.2.8 SDS-PAGE and Zymogram analysis

Protein concentration was determined with using the Bradford assay (Bradford, 1976). The molecular weight of the recombinant proteins were identified by SDS-PAGE and zymogram analysis as described in Chapter 2.

4.3 Results and Discussion

4.3.1 DNA, RNA and poly(A) mRNA isolation

Genomic DNA was isolated *via* the method of Miller *et al*. (1999), while RNA and poly-A-tailed mRNA were isolated using commercial isolation kits. The quantities and qualities of each preparation can be found in Table 4.3. These nucleic acids were used as template for PCR as well as for the construction of cDNA libraries.

Table 4.3 DNA, RNA and poly-A-tailed mRNA quantities and qualities

Nucleic acid	Quantity (ng/µl)	Quality (260/280 nm)
DNA prep 1	146.4	1.7
DNA prep 2	434.7	1.99
DNA prep 3	1089.9	1.54
RNA prep 1	530.6	1.93
RNA prep 2	446.1	1.99
RNA prep 3	581.3	1.88
Poly-A-tailed RNA prep 1	20.4	1.98
Poly-A-tailed RNA prep 2	13.1	1.89

Genomic DNA was extracted on three separate occasions. The third attempt was carried out using a greater mass of fungal mycelia and resulted in a high yield of DNA, however, the quality proved to be lower, possibily due to protein and fungal polysaccharide contamination. RNA isolation using the SV Total RNA Isolation System proved to deliver reproducible

results with moderate yield; however, subsequent isolation of mRNA resulted in very low yields.

Poly(A) mRNA was ligated to pJET1.2/blunt and transformed into *E. coli* HB101. Twenty three positive clones were streaked onto LB-amp-CMC plates and screened for EG activity *via* Congo red staining, however, none of the clones displayed zones of hydrolysis. The low yield of mRNA recovered is responsible for there being so few clones in the library. The sample processing for for total RNA isolation involves harsh physical treatment (grinding with liquid nitrogen) of the fungal biomass. Due to the fragile nature of mRNA, this could have contributed to the low yields obtained. It was then decided that focus should be placed on PCR amplification of an EG gene from *P. alba* rather than solely relying on cDNA/mRNA libraries.

4.3.2 PCR amplification of the endoglucanase gene from P. alba

Following induction on each medium, total RNA was isolated from fungal biomass from each culture medium and pooled; double-stranded cDNA was generated and used as template DNA for PCR with primer set F10EgF and F10EgR. The primer set used was based on the gene sequence of the GH5 EG from *Phialophora* sp. G5 and thus the expected size of the amplicon was ~1300 bp (Zhao *et al*., 2011). Two distinct bands of 1000 bp and 600 bp were produced (Figure 4.1). These bands were excised, purified and sequenced.

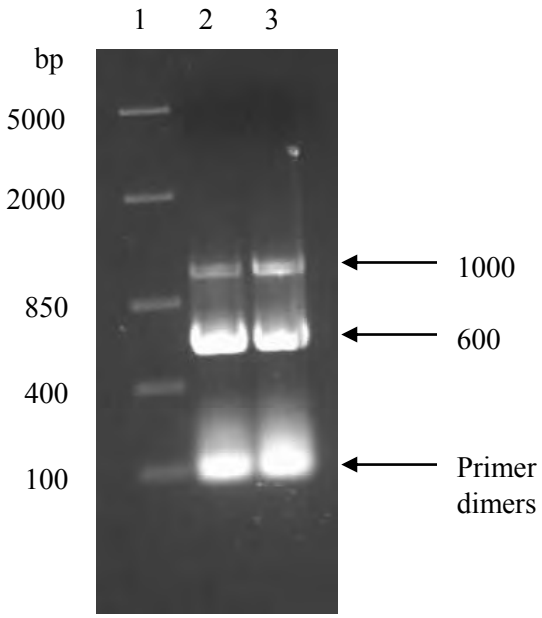


Figure 4.1: PCR amplification using primer set F10EgF and F10EgR and double-stranded cDNA as template (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: cDNA using dT_{18}; 3: cDNA using F10EgR)

However, in each instance whether cDNA synthesis was initiated with the Oligo$(dT)_{18}$ primer or gene specific primer F10EgR, the putative gene sequences did not display similarity to EGs or other glycoside hydrolases. A review of the physiology and secreted enzymes of thermophilic fungi generalize that crystalline forms of cellulose such as avicel and α-cellulose are superior inducers of cellulase enzymes over amorphous cellulose such as CMC, however, some fungi are capable of producing high cellulase as well as xylanase activity on hemicellulosic substrates (Maheshwari *et al*., 2000). Even though a variety of cellulose substrates (both crystalline and amorphous) were used it was believed that they were unable to induce high expression of cellulase genes. Multiple attempts at cloning an EG gene from genomic DNA were made.

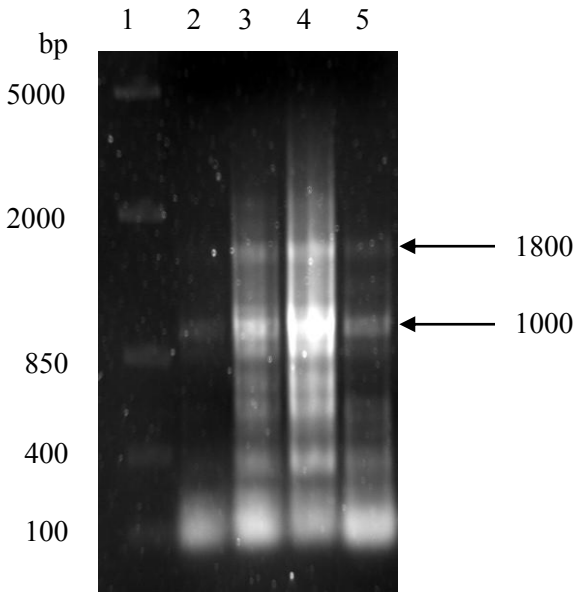


Figure 4.2: PCR amplification using primers F10EgF and EgF10R at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: 1.25 mM; 3: 2 mM; 4: 2.5 mM; 5: 3 mM)

PCR amplification using the same primer set with genomic DNA produced a range of amplicons from ~2000 bp to 400 bp with a major band appearing at ~1000 bp (Figure 4.2). This band, together with a 1800 bp band was excised and cloned into *E. coli* DH5α. Plasmid DNA was isolated and the inserts were sequenced, however, BLAST analysis revealed that both these fragments were not similar to EGs or related glycoside hydrolases either. Degenerate primers for family 5 (expected amplicon size: 390 bp) (Figure 4.3), primers based on the cellulase domain of the EG from *Phialophora* sp. G5 (expected amplicon size: 400 bp) (Figure 4.4) and degenerate family 45 glycosyl hydrolases primers (expected amplicon size: 600 bp) (Figures 4.5 and 4.6) were also used to PCR amplify putative EG genes from *P. alba* genomic DNA and cDNA.

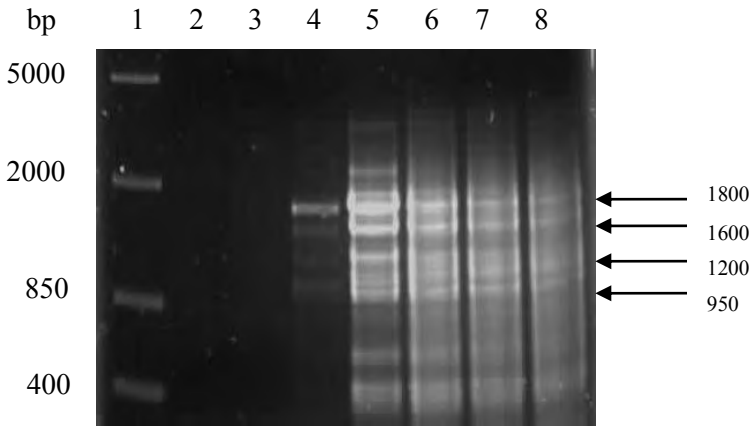


Figure 4.3: PCR amplification using degenerate primers Eg-GH5-f and Eg-GH5-r at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: No template control; 3: 0 mM; 4: 1 mM; 5: 1.5 mM; 6: 2 mM; 7: 2.5 mM; 8: 3 mM)

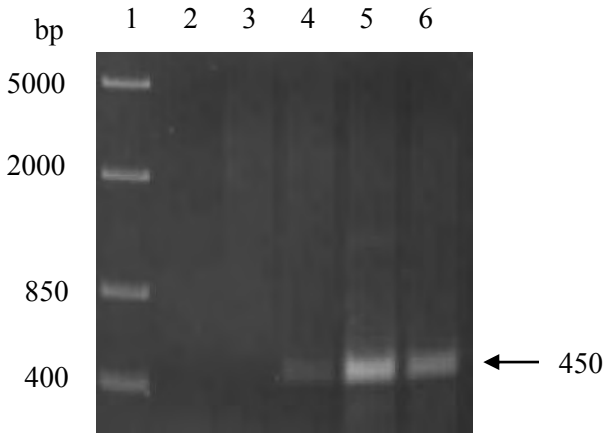


Figure 4.4: PCR amplification using cellulase core primers Eg5-F and Eg5-R at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: No template control; 3: 1.5 mM; 4: 2 mM; 5: 2.5 mM; 6: 3 mM)

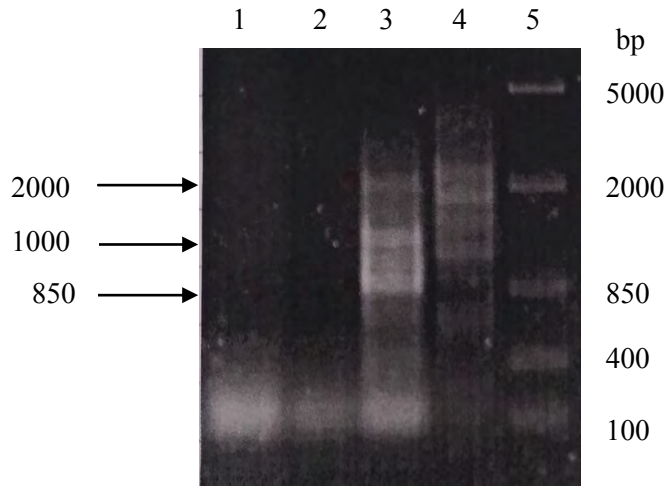


Figure 4.5: PCR amplification using degenerate GH45 primers and degenerate CBHII primers (Lane 1: No template control GH45; 2: No template control CBHII; 3: GH45(1.25 mM); 4: CBHII (1.5 mM); 5: Marker (bp): FastRuler Middle Range DNA Ladder)

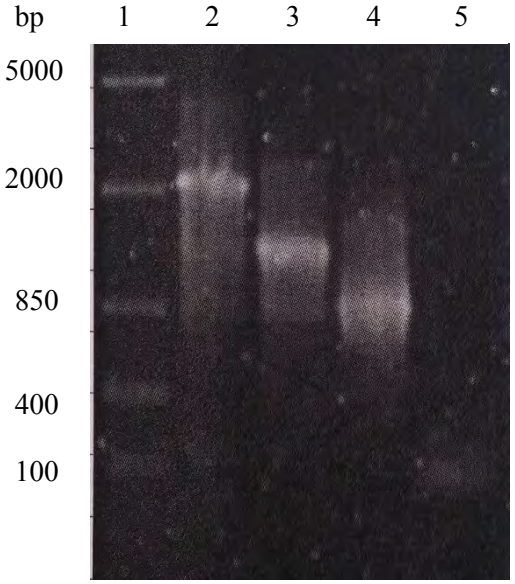


Figure 4.6: PCR re-amplification of excised bands from initial amplification with degenerate GH45 primers (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder; 2: ~2000 bp excised band; 3: ~1000 bp excised band; 4: ~850 bp excised band; 5: No template control)

All the fragments (indicated by arrows in each figure) were cloned and sequenced; however, none showed similarity to EGs following BLAST analysis. The enzymes belonging to GH family 5 are divided into eight subfamilies and are highly divergent with regard to their function and protein sequences as they share less than 25% sequence identity (Hilge *et al*.,

1998). This variation in sequences could explain the difficulty experienced in attempting to amplify GH5 family genes.

At this point it was decided that the whole genome of the fungus be sequenced and subsequently data-mined for the EG gene and other CAZymes. The details of the sequenced *P. alba* genome has already been discussed at length in Chapter 3. A single open reading frame (ORF) (Appendix C: Figure C1) which matched a GH5 cellulase gene from the thermophile *Thielavia terrestris* NRRL 8126 was selected for the continuation of the study. Specific PCR primers (GH5_ORF primer sets) were designed based on the sequence of the ORF and used to amplify the gene of interest (Figure 4.7). Family 45 GHs were not present in the *P. alba* genome and thus explains why the BLAST analysis of cloned fragments did not reveal a hit for GH45 EGs.

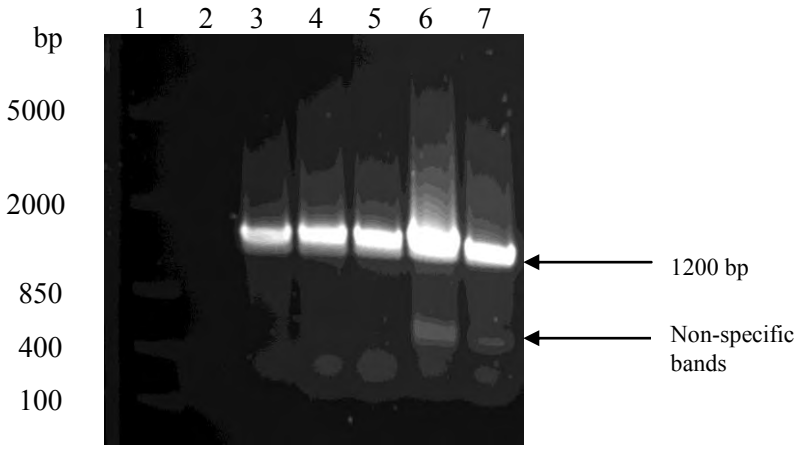


Figure 4.7: PCR amplification of endoglucanase gene from *P. alba* using GH5_ORF primers (G) and GH5_ORF_Flank primers (F) at different $MgCl_2$ concentrations (Lane 1: Marker (bp): FastRuler Middle Range DNA Ladder, 2: G 1 mM, 3: G 1.5 mM, 4: G 2mM, 5: F 1 mM, 6: 1.5 mM, 7: F 2 mM)

PCR amplification of the EG gene was carried out using the primer sets GH5_ORF_F and GH5_ORF_R primers (G) and GH5_ORF_Flank and GH5_ORF_R primers (F) over a range of $MgCl_2$ concentrations. The expected size of the amplicon was 1251 bp. A $MgCl_2$ concentration of 1 mM was insufficient to produce any bands of interest using the GH5_ORF

primers (Lane 2), however, amplification was successful with GH5_ORF_R and GH5_ORF_Flank_F (Lane 5) as well as at higher $MgCl_2$ concentration. As the $MgCl_2$ concentration increased, non-specific amplification was observed when using the GH5_ORF_R and GH5_ORF_Flank_F primers even though a greater yield in target amplicon was observed. As a result, all further amplifications of the target gene were performed using 1.5 mM $MgCl_2$ with the GH5_ORF primers. cDNA was also used as template DNA in PCR reactions to obtain a full length transcribed gene. The amplicons was excised, purified, ligated to plasmid pTZ57R/T and transformed into *E. coli* DH5α. Positive clones were screened for the gene of interest by isolating plasmid DNA and restriction analysis with *Xba*I and *Bam*HI (Figure 4.8).

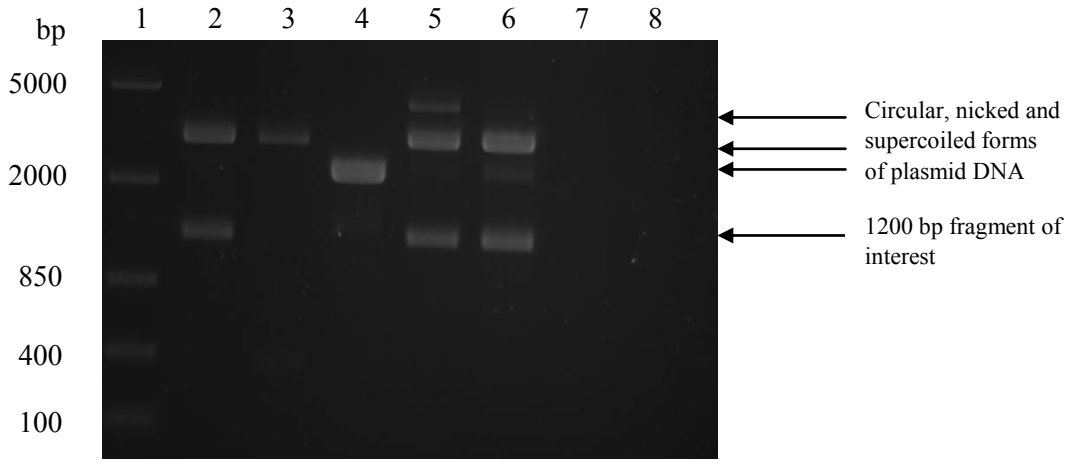


Figure 4.8: Restriction analysis of selected positive clones (Lane 1: Marker: FastRuler Middle Range DNA Ladder, 2: Plasmid G1, 3: Plasmid cDNA1, 4 Plasmid cDNA2, 5: Plasmid F1, 6: Plasmid F2, 7: Plasmid F3, 8: Plasmid F4) (G and F refer to the primers sets discussed previously, cDNA refers to cDNA being used as template DNA)

The inserts for Plasmids G1, cDNA2, F1 and F2 were sequenced and it was found that plasmids G1, F1 and F2 all possessed the same insert and was identical in sequence to the parent ORF. Multiple attempts at sequencing the insert in Plasmid cDNA2 were made, however, for some unknown reason the sequencing reaction always failed. New primers were

synthesised based on the Eg5_ORF primers to include restrictions sites *EcoR*I and *Not*I on the forward and reverse primers, respectively, for subsequent cloning into expression vector pPIC9 and pET-22b.

4.3.3 Sequence analysis of endoglucanase gene from P. alba

The 1251 bp gene had 68%, 64% and 72% sequence similarity to GH5 EGs and cellulases from *Rasamonia emersonii*, *Thielavia terrestris* and *Myceliophthora thermophila*, respectively, all three fungi are known thermophiles (Berka *et al*., 2011; Houbraken *et al*., 2012). The open reading frame of the *P. alba* EG gene comprised of a single exon with no introns. The open reading frame encodes a polypeptide of 416 amino acids (Figure 4.9) with a calculated molecular mass of 45.7 kDa and a calculated *p*I of 4.94. InterPro and Pfam protein analysis revealed that the polypeptide belonged to the GH family 5 with a catalytic domain for cellulase activity (PF00150) with no carbohydrate binding domain present. The tertiary structures of the *P. alba* EG as well as the EG from *M. thermophila* (for the sake of comparison) was predicted using the Phyre^2 web portal (Figure 4.10). The structure of the catalytic domain of the EG from *P. alba* is indicative of a GH5 family enzyme as they all share the $(\beta/\alpha)_8$-barrel fold (Tseng *et al*., 2011). It is also thought that variations in the surface loops (indicated in white and blue in Figure 4.10) may account for the different substrate specificities (Tseng *et al*., 2011).

A number of conserved amino acids were identified in EgGH5 including Gly76, Arg81, Asn216, Glu217, Glu330 and Trp368 which were conserved among other GH5 EGs when aligned. Among these conserved amino acids were Glu217 and Glu330 which are the putative catalytic residues and are analogous to the acid/base catalytic and nucleophile glutamic acid residues (Glu226 and Glu338) of EgG5 from *Phialophora* sp. G5 (Zhao *et al*., 2011).

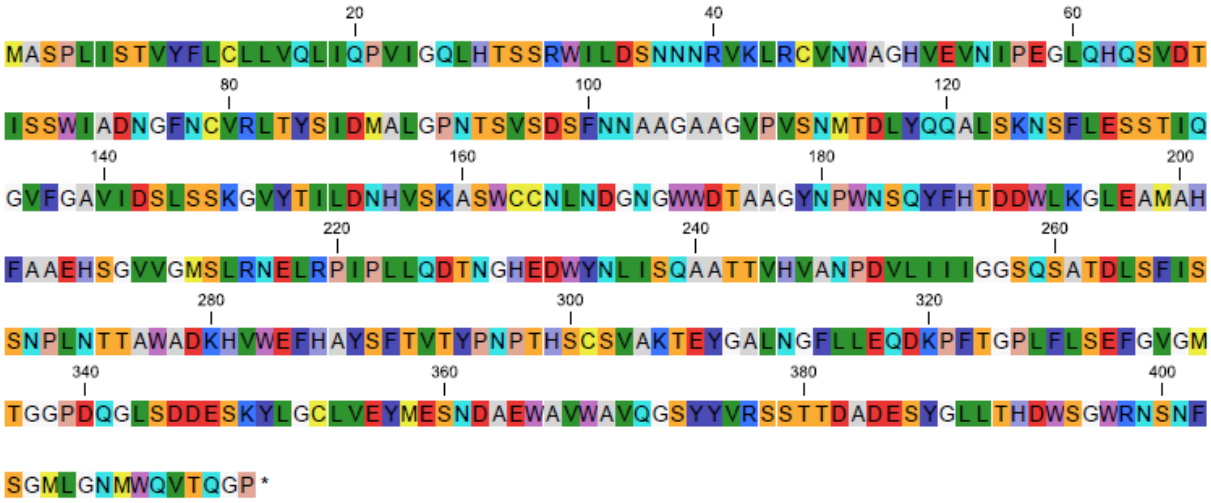


Figure 4.9: The 416 amino acid long sequence of the *P. alba* GH5 endoglucanase translated from the nucleotide sequence

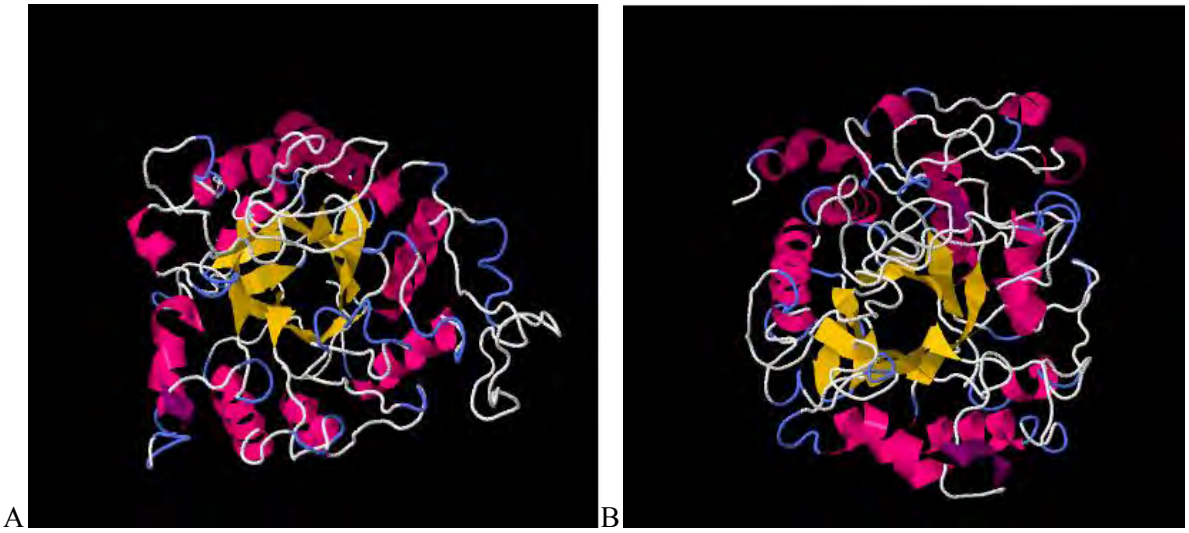


Figure 4.10: The tertiary structures of EgGH5 from *P. alba* (A) and the GH5 protein from *M. thermophila* (B) (models predicted and rendered using the Phyre[2] web portal)

4.3.4 Expression in P. pastoris

The *EgGH5* gene insert was reamplified from Plasmid G1 with the Exp_Eg5_ORF primers, ligated into pPIC9 and transformed into *P. pastoris* GS115. *Pichia* clones with successful integration of the pPIC9_EgGH5 recombinant vector into their genome were screened *via*

PCR and their phenotype (Mut^+ or Mut^s) was determined by plating the clones on minimal dextrose and methanol plates. All clones screened were of the genotype Mut^+, indicated by similar growth on both minimal dextrose and minimal methanol plates.

Six *Pichia* clones harbouring the *EgGH5* gene were selected for time course analysis in BMMY media for the production of the EG. However, no EG activity was detected and low levels of protein were secreted into the culture filtrates. It was thought that proteases secreted into the medium by *P. pastoris* degraded the recombinant protein and thus the experiment was repeated using an unbuffered medium (MM) in an effort to inactive neutral proteases. However, no activity was detected. *Pichia*, like *Saccharomyces cerevisiae* is able to glycosylate recombinant proteins, albeit, to a lesser extent (*Pichia* Expression Kit Manual) thus hyperglycosylation was ruled out as a possible reason for enzyme inactivity as there are only four possible sites for N-linked glycosylation sites on EgGH5 as predicted by the NetNGlyc server.

Sequencing of the recombinant pPIC9_Eg_GH5 revealed that the gene was cloned in frame within the plasmid. The AOX1 primers were used to amplify the integrated EG from the *Pichia* clones and sequencing confirmed that the gene integrated into the host genome in the correct reading frame. However, upon aligning the sequence with the original ORF, there appeared to be two point mutations leading to amino acid changes at positions 85 (Serine to Proline) and 122 (Lysine to Proline) (Figure 4.11). Proline is unique among amino acids as its side chain cyclises back onto the amino acid backbone and thus lacks an amide proton required for participation in α-helix stabilization (Li *et al*., 1996). For this reason, and among others, proline is a potent destabilizer of protein secondary structures. Analysis using Phyre^2 revealed that position 86 is highly sensitive to mutation. It also revealed that position 85 is adjacent to an amino acid (Ile86) that is directly involved in an α-helix. It was thus concluded

that this point mutation may have been responsible for the lack of expression (or secretion) and activity by destabilizing the protein's structure.

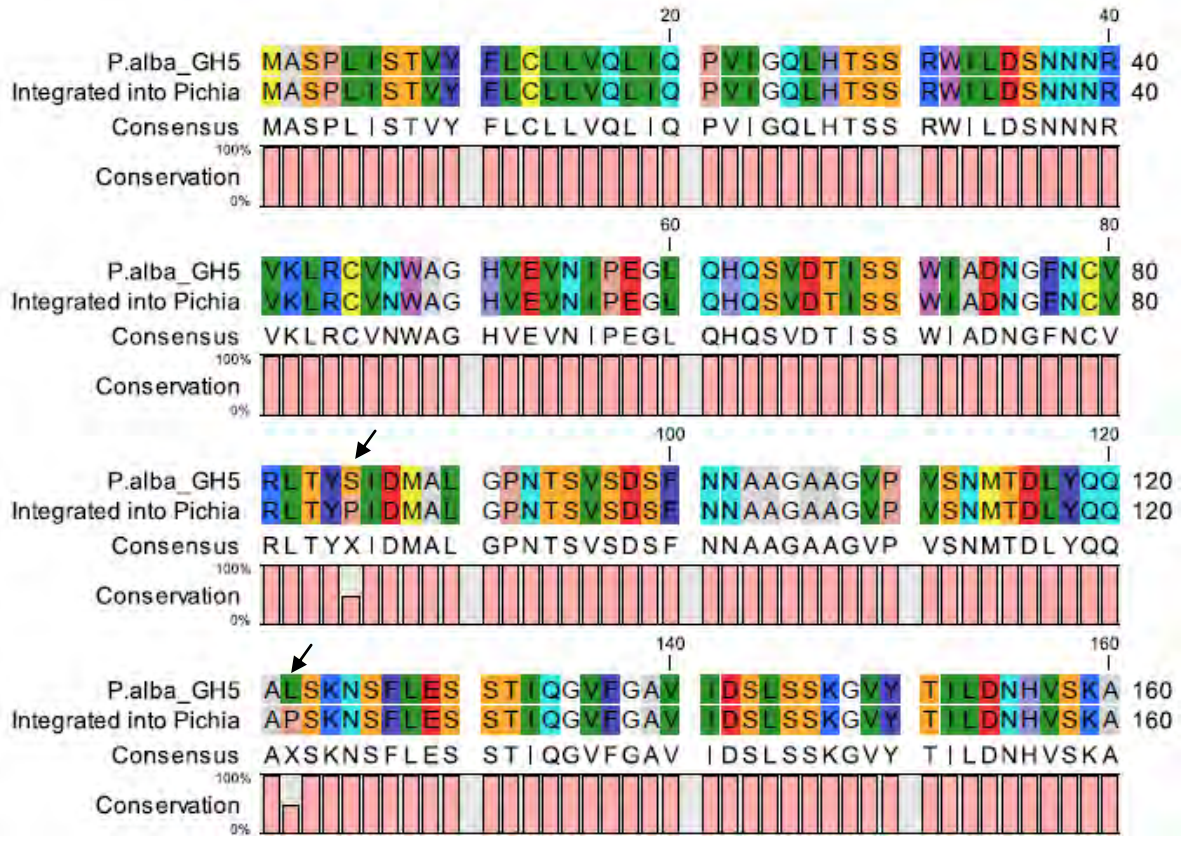


Figure 4.11: Section of sequence alignment (1-160) of translated native GH5 endoglucanase gene (*P. alba*_GH5) and endoglucanase gene integrated into *Pichia* genome (amino acid changes indicated by black arrows)

The gene was thus reamplified with a high fidelity polymerase (*pfu*) and cloned into pPIC9 to produce the recombinant vector pPIC9_13_Eg_GH5. The recombinant vector was linearized in three separate restriction reactions using either *Sal*I, *Bgl*II or *Not*I and transformed once more into *P. pastoris* GS115. Three different restriction enzymes were used to linearize the vector as this would result in integration of the plasmid at different locations in the *Pichia* genome. Depending on the site of integration, expression of the protein could be better (*Pichia* Expression Kit Manual). Ten positive clones from each successful transformation were streaked onto MM plates supplemented with 1% CMC and 1% avicel to screen for EG

activity. A total of 30 clones were screened for functional EG activity (Figure 4.12). Though the figure below shows clearing, these are artefacts caused by the *Pichia* colonies preventing the Congo red stain from staining the plate effectively. Selected clones were re-grown (4 per plate) and screened to confirm that these were indeed false positive results.

Figure 4.12: Congo red plate screening assay (CMC) of *Pichia* clones from transformations performed with recombinant vector linearlized with *Sal*I, *Bgl*II and *Not*I

The *Pichia* clones (30) were screened for the presence of the EG. Three clones with gene integration (Figure 4.13) (B2, S1 and S2) were selected for time course analysis. Phenotype screening revealed that they were all Mut^+. However, once again, EG activity was not detected over the period of induction. Protein levels were extremely low and could not be detected with the Bradford assay and thus the protein could not be standardized for SDS-PAGE analysis. However, silver staining did reveal the presence of a protein within the region of ~47 kDa present in the supernatant of the *Pichia* clones containing the insert. It was thus decided to re-grow and induce expression in a larger volume and concentrate the samples to screen for activity.

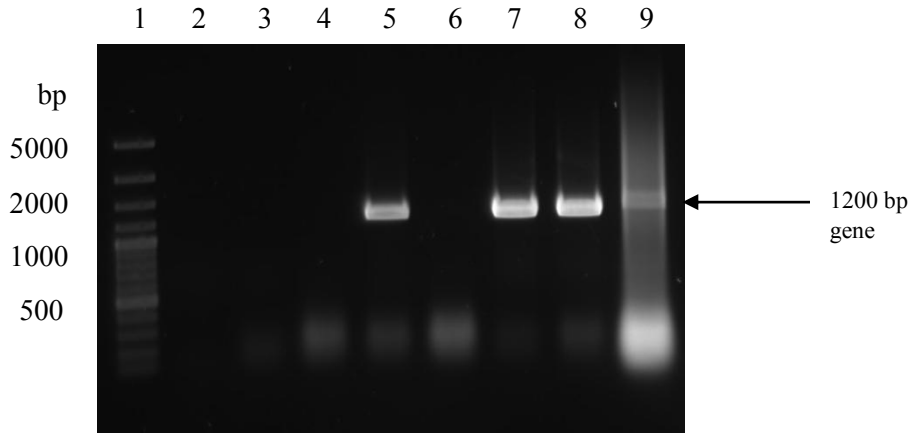


Figure 4.13: PCR amplification of endoglucanase gene from *Pichia* clones using GH5_ORF primers (Lane 1: Marker: GeneRuler 100 bp Plus DNA Ladder, 2: No template control, 3: *Bgl*II no insert linear pPIC9, 4: *Sal*I no insert linear pPIC9, 5: B2, 6:B6, 7: S1, 8: S2, 9:S3) (B and S: *Bgl*II and *Sal*I linearized vector clones)

The induction was carried out in 50 ml of media as opposed to 10 ml. Larger samples (4 ml) were collected each day and concentrated 10 fold. Unfortunately, no EG activity was observed after performing the DNS assay as well as after spotting the supernatant onto CMC plates and staining with Congo red. *Pichia* clones were also ruptured and assayed, however, no EG activity was detected. There was, however, induction and expression of proteins as confirmed by the Bradford assay and SDS-PAGE (protein standardized to 0.323 µg/ml).

There was a steady increase in the production and secretion of proteins after induction with methanol. *Pichia* clone S1 secreted the highest amount of protein, 6.3 µg/ml after 4 days of induction (Figure 4.14). *Pichia* clones S2 and B2 secreted similar amounts of protein after 4 days (3.0 and 2.9 µg/ml, respectively). There was no significant increase in the amount of protein secreted by the control strain as the extracellular protein concentration remained below 1 µg/ml.

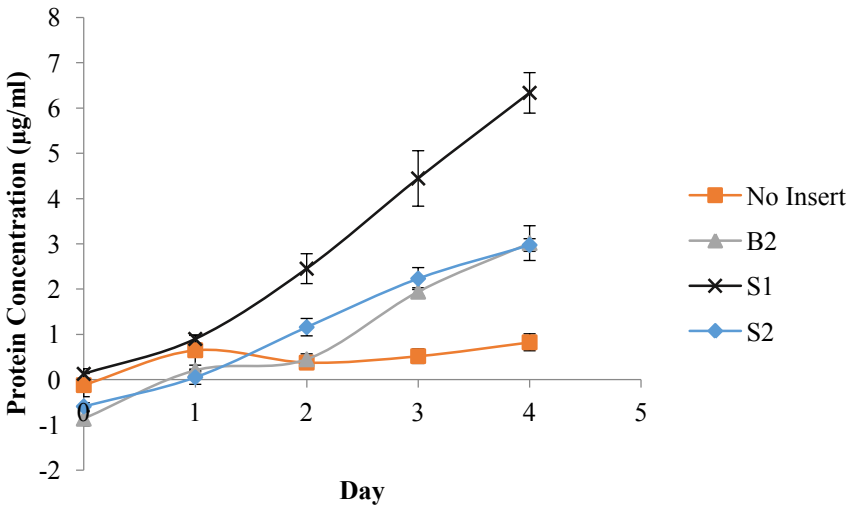


Figure 4.14: Induction of recombinant protein expression by *Pichia* clones with 0.5% methanol over a period of 4 days

Pichia makes for an ideal expression host as it does not secrete many native proteins as is evident from lanes 2-4 displaying only two native proteins at ~60 kDa and ~46 kDa (Figure 4.15). The recombinant strains however, secreted a host of different proteins ranging from ~100 kDa to 20 kDa. Both recombinant clones produced roughly the same protein banding pattern with bands of interest at ~48 kDa. This 48 kDa band was quite prominent on day two appeared fainter on day four. The theoretical size of the protein is 45.7 kDa, however *Pichia pastoris* GS115 is known to glycosylate proteins which can result in the protein having a higher molecular weight. Pham *et al*. (2011) reported that *P. pastoris GS115* produced a recombinant β-glucanase that was 7 kDa greater than the wild-type protein produced by *A. niger* as a result of glycosylation. As mentioned previously there were four possible sites predicted for N-linked glycosylation on the EG at positions 93, 113, 296 and 401. Any one of these sites may have glycosylated by *Pichia* resulting in a larger protein. The core glycan that is added to proteins *via* N-linked glycosylation is composed of 14 saccharides ($Glc_3Man_9GlcNAc_2$) resulting in a 2 kDa mass increase for a single core glycan (Zhong and Somers, 2012).

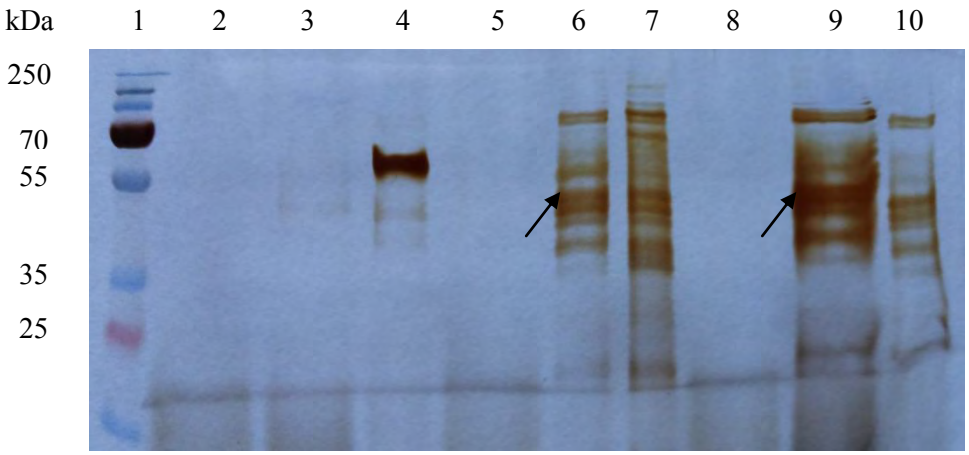


Figure 4.15: Silver stained SDS-PAGE gel of recombinant proteins expressed by *Pichia* clones after induction with methanol (Lane 1: Marker: PageRuler Plus Prestained Protein Ladder, Lanes: 2-4: *Pichia* pPIC9-no insert Day 0, 2, 4; Lanes: 5-7: *Pichia* pPIC9-S1 Day 0, 2, 4; Lanes: 8-10: *Pichia* pPIC9-B2 Day 0, 2, 4)

Zymogram analysis confirmed the lack of EG activity was as no substrate hydrolysis zones were evident in the gel (Appendix A: Figure A5). One can only speculate as to why no activity was observed even though the protein was successfully expressed. *P. pastoris* is known as an excellent expression host with numerous cellulolytic enzymes being successfully expressed and characterized (Macauley-Patrick *et al*., 2005; Quay *et al*., 2011; Li *et al*., 2011; Krainer *et al*., 2012; Jimenez *et al*., 2014), however, not every protein can be produced or expressed in high titres in *P. pastoris*, particularly if they are membrane-attached, prone to protease degradation or expressed as oligomers (Ahmad *et al*., 2014). It is highly unlikely that the EG from *P. alba* is a membrane associated protein EGs are typically proteins that are secreted. It is possible that the EG functions as a dimer as is the case with xylanases Xyl I and Xyl II from *Talaromyces emersonii* (Maheshwari *et al*., 2000) and the EG E2 from *Thermomonospora fusca* (McGinnis and Wilson, 1993). Expression in *P. pastoris* may not allow for proper formation of the dimer and thus the enzyme cannot attain an active form. It was reported by Poidevin and coworkers (2013) that through all their

efforts and trials, a GH family 6 EG from *Podospora anserine*, *Pa*GH6D, could not be cloned and expressed in *P. pastoris* for unknown reasons.

A likely reason as to the lack of activity is that the cloned gene encoding the EG, actually encodes a non-functional protein or GH5 having a different substrate specificity to what was returned after the Pfam search. Support for this notion is that although BLAST analysis shows identity with other thermophilic EGs, those sequences are also based on data derived from CAZome analysis of fungi and thus there is no functional data or studies available for them. The problem with the identification of cellulases *via* computational means is that current algorithms used to identify cellulase domains are not able to recognize and differentiate between members of the same family and thus the results returned cannot be accepted with certainty until physical experiments are performed (Sukharnikov *et al*., 2011). The *EgGH5* gene may very well encode a GH family 5 protein with substrate specificity other than cellulose.

Phialophora sp. G5 is a mesophilic fungus that is able to produce thermophilic cellulases (Zhao *et al*., 2012a). *P. alba*, being a thermophilic fungus must have had to undergo genomic changes in order to survive and thrive under conditions of high temperature. The GH5 EG from *P. alba* unlike the GH5 EG from *Phialophora* G5, does not possess a carbohydrate binding domain. Zhao and colleagues (2011) showed that removal of the CBM from the EG of *Phialophora* sp. G5 resulted in greater thermostability of the enzyme. *P. alba* may have selected for a CBM-truncated version of *EgGH5*, however, the deletion may have inadvertently resulted in an EG with dimished activity if additional parts of the gene were lost.

CAZome analysis of *P. alba* revealed that the fungus is deficient in a complete set of enzymes required for cellulose hydrolysis and is instead highly capable of xylan and possibly

chitin degradation. *Rhizopus oryzae* has adapted at the genomic level to its environment as it is unable to degrade complex polysaccharides such as cellulose and xylan but is quite capable of degrading chitin, chitosan and β-1,3-glucan (Battaglia *et al*., 2011). It is possible that *P. abla* has evolved and adopted a highly specialized approach to carbohydrate degradation relying mostly on hemicellulosic substrates while using auxiliary enzymes for the degradation of cellulose and may rely on other organisms in the environment for the liberation of nutrients from cellulose.

4.3.5 Expression in E. coli BL21(DE3)

Despite the mounting evidence suggesting the cloning of an inactive protein, a final attempt at expressing an active EG for the purposes of characterization was made. The EG gene was re-amplified, ligated into expression vector pET-22b and transformed into *E. coli* BL21(DE3). Five clones were screened by colony PCR for the presence of the 1200 bp insert (Figure 4.16). Of the five clones screened, four of them, eB1, eB3, eB4 and eB5 contained the recombinant plasmids based on restriction analysis. The clones harbouring the insert were streaked onto LB-amp-CMC plates containing 1 mM IPTG and screened for EG activity after 24 hours by Congo red plate staining assay, however, no CMC hydrolysis was observed. The five clones were also grown overnight in LB broth and expression was induced with 1 mM IPTG for six hours. The cells were thereafter ruptured by sonication and the cell content was tested for EG by spotting on CMC plates prepared in phosphate buffer (pH 5). The EG positive control displayed a very pale halo of clearing around the spot (indicated by the arrow) indicative of EG activity while the intracellular proteins of the clones were negative for EG activity (Figure 4.17). The DNS assay also confirmed lack of activity.

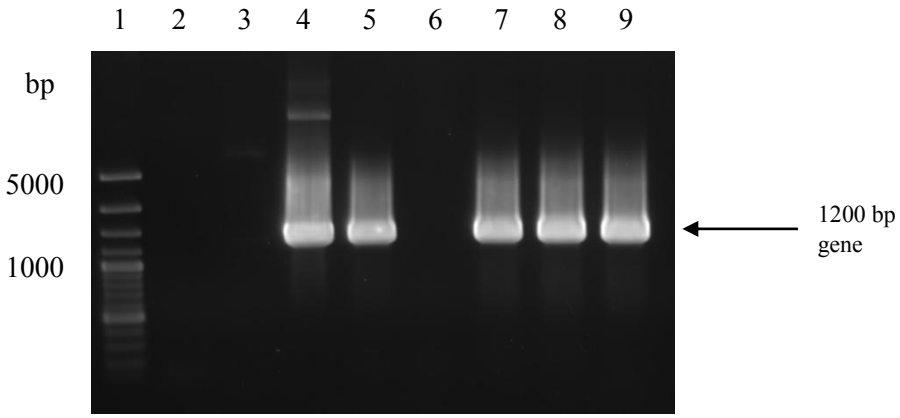


Figure 4.16: Colony PCR of *E. coli* BL21 (DE3) clones (Lane 1: Marker: GeneRuler 100 bp Plus DNA Ladder, 2: No template control, 3: pET-22b-no insert, 4: pPIC9_13_Eg_GH5, 5-9: Clones eB1 – eB5)

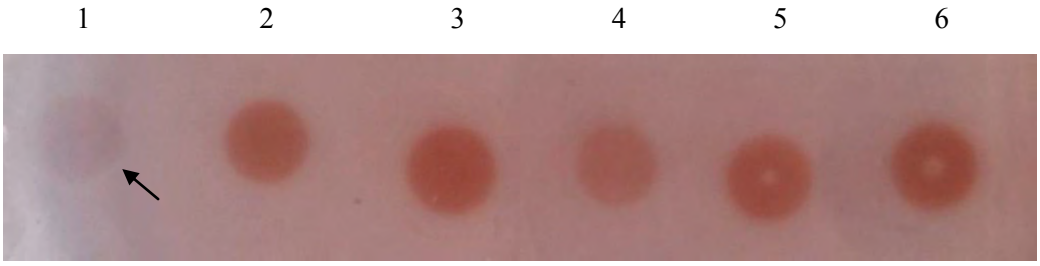


Figure 4.17: Congo red plate screening assay of intracellular proteins of *E. coli* BL21 (DE3) clones (Pos: Fungal extract known to possess endoglucanase activity, Spots 2 – 6: eB1 – eB5)

The intracellular proteins from the positive clones (eB1, 3, 4 and 5) were standardized to 1 µg/ml and subjected to SDS-PAGE analysis. The *E. coli* BL21(DE3) control strain containing the expression vector with no insert produced a host of intracellular proteins. It was observed that all four clones harboring the recombinant pET-22b vector produced a ~46 kDa band (indicated by arrow) (Figure 4.18). There appeared to be overexpression of another protein at ~55 kDa, however, it is unlikely that this is the protein of interest as it matches the banding pattern of the control strain.

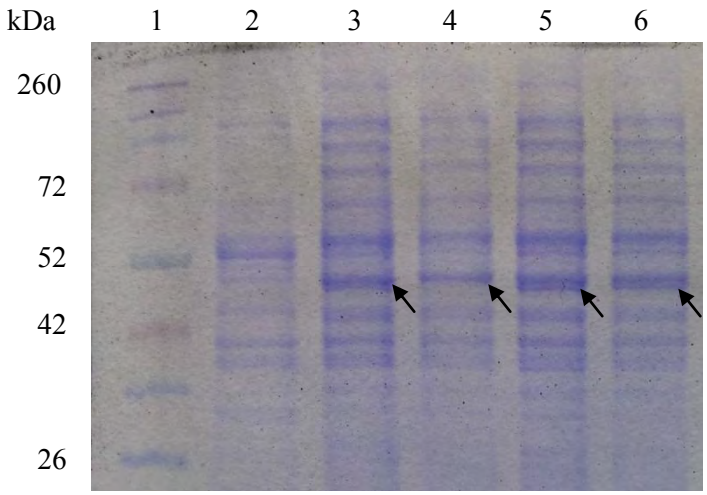


Figure 4.18: Coomassie stain of SDS-PAGE of intracellular proteins from recombinant *E. coli* BL21(DE3) clones following 6 hours induction with 1 mM IPTG (Lane 1: Marker: Spectra Multicolor Broad Range Protein Ladder, 2: *E. coli* BL21(DE3) with pET-22b-no insert, 3: eB1, 4: eB3, 5: eB4, 6: eB5)

There is no doubt that *E. coli* is an expression host of choice for a variety of reasons, however, the expression of eukaryotic proteins can often be problematic, resulting in inefficient folding due to erroneous disulfide bond formation, the formation of inclusion bodies, protein inactivity, codon bias, lack of glycosylation and poor host growth (Rosano and Ceccarelli, 2014). In this instance, aside from the aggregation of the recombinant protein in inclusion bodies, any one of these reasons may be responsible for the lack of activity; however, it is strongly believed that gene that was cloned from *P. alba* either codes for an inactive protein or for a GH5 enzyme with a substrate specificity other than cellulose. The substrate specificity of GH5 enzymes are extremely diverse. The main objective of this study was to identify an EG that could be used for application to hydrolyse cellulose from lignocellulosic biomass for bioethanol production. Hence, activity of the putative EG cloned was not tested on alternate substrates to identity its specificity.

4.4 Conclusion

Numerous challenges were encounted throughout this study as detailed above. After troubleshooting along each step and expressing the gene in both eukaryotic and prokaryotic

systems, it was found that the cloned gene did not produce a protein with EG activity. Thus, the protein was unable to be characterized. It is believed that the problem may lie with the current computational models used to identify enzymes with cellulase activity. This problem could be curbed in the future by creating databases and prediction models that not only take into account protein sequences and catalytic domains but also enzyme biochemical data and accessory domain architecture such that a more robust identification is obtained (Sukharnikov *et al*., 2011). The CAZy database itself still lacks an inhouse similarity search feature and they themselves acknowledge that they are a long way from reliable substrate and product prediction for CAZymes encoded in a genome (Cantarel *et al*., 2009; Lombard *et al*., 2014). It is believed that only by increasing the output of experimental data will the functional predictions of CAZymes become more reliable (Lombard *et al*., 2014).

CHAPTER 5

Concluding Remarks

Due to their use in a wide variety of applications such as use in detergents, juice clarification, cotton processing, paper recycling and as animal feed additives, cellulases are the third largest industrial enzyme on a global scale (Wilson, 2009). The most popular application however, that is currently being investigated is the use cellulases for the saccharification of lignocellulosic biomass, in particular, waste materials such as sugarcane bagasse, corncob and rice straw (Kuhad *et al*., 2011). The monosaccharides that have been liberated would thereafter be converted to ethanol to produce renewable biofuel. The production of bioethanol from lignocellulosic biomass has come to a point where newer, more robust, thermostable cellulases are required for efficient production (Yennamalli *et al*., 2013).

In the current study, fungal isolates were screened for thermostable/thermophilic endoglucanases that would be used to liberate sugars from sugarcane bagasse for bioethanol production. *A. fumigatus* D9 displayed the highest endoglucanase activity of the isolates screened, however, the activity was still very low in comparison to other EGs reported in literature. Attempts to purify this enzyme also proved to be extremely difficult. Taking into consideration the difficulty of purification as well as physiochemical properties of the crude endoglucanase, the focus of the study shifted to that of the thermophilic fungus, *P. alba*.

Multiple attempts at cloning an endoglucanase gene from *P. alba* were made in vain until finally, the genome of the fungus was sequenced and a specific gene was identified. The gene was successfully cloned and expressed in both *P. pastoris* and *E. coli*. However, in both instances the recombinant protein did not display endoglucanase activity. Bioinformatic analysis and datamining of CAZymes from the *P. alba* genome revealed that the fungus is not an efficient degrader of cellulose and it is possible that the gene that was cloned codes for an

inactive protein or for an GH5 enzymes with a substrate specificity other than cellulose. Even though Pfam domain searches revealed the catalytic domain to be active on cellulose, it is possible that due to the low proportion of biochemical data to sequence data, this prediction may have been incorrect as the substrate specificity of GH5 enzymes are extremely diverse. A more rigorous and reliable database and computational algorithm that takes into account protein sequences, structural architecture and biochemical characterization will need to be developed in order to efficiently predict the functions of GH enzymes. Vlasenko *et al*. (2010) performed a phylogenetic analysis of GH5 enzymes in the CAZy database and found that many of the subfamilies consist solely or mostly of mannanases. They also noted that of the seven GH5 endoglucanases they assayed, six of them possessed significant specificity toward mannan and galactomannan, however, these substrate specificities were in addition to cellulose. CAZome analysis of the *P. alba* genome sheds light on the nutritional strategy of this fungus and reveals its preference toward xylan and chitin rather than cellulose. Furthermore, the analysis did not show preference for mannan or galactomannan. Future studies should focus on transciptomic analysis together with bioinformatic analysis of the *P. alba* genome and the metagenome of microbes found on *Eucalyptus* spp. woodchips that could reveal more insight into how *P. alba* fits into this microcosm.

REFERENCES

Abbott, D. W. and A. B. Boraston. 2008. Structural biology of pectin degradation by Enterobacteriaceae. Microbiology and Molecular Biology Reviews. **72:**301-316.

Ahmad M., M. Hirz, H. Pichler and H. Schwab. 2014. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Applied Microbiology and Biotechnology. **98:**5301-5317.

Al-Samarrai, T. H. and J. Schmid. 2000. A simple method for extraction of fungal genomic DNA. Letters in Applied Microbiology. **30:**53-56.

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

Anish, R., M. S. Rahman and M. Rao. 2007. Application of cellulases from an alkalothermophilic *Thermomonospora* sp. in biopolishing of denims. Biotechnology and Bioengineering. **96:**48-56.

Antoni, D., V. V. Zverlov and W. H. Schwarz. 2007. Biofuels from microbes. Applied Microbiology and Biotechnology. **77:**23-35.

Asha, B. M., M. Revathi, A. Yadav and N. Sakthivel. 2012. Purification and characterization of a thermophilic cellulase from a novel cellulolytic strain, *Paenibacillus barcinonensis*. Journal of Microbiology and Biotechnology. **11:**1501-1509.

Bailey, M., P. Biely and K. Poutanen. 1992. Interlaboratory testing of methods for assay of xylanase activity. Journal of Biotechnology. **23:**257-270.

Bajaj, B. K., H. Pangrota, M. A. Wani, P. Sharma and A. Sharma. 2009. Partial purification and characterization of a highly thermostable and pH stable endoglucanase from a newly isolated *Bacillus* strain M-9. Indian Journal of Chemical Technology. **16:**382-387.

Balat, M. 2011. Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review. Energy Conversion and Management. **52:**858-875.

Battaglia, E., I. Benoit, J. van den Brink, A. Wiebenga, P. M. Coutinho, B. Henrissat and R. P. de Vries. 2011. Carbohydrate-active enzymes from the zygomycete fungus *Rhizopus oryzae*: a highly specialized approach to carbohydrate degradation depicted at genome level. BMC Genomics **12:**38-50.

Bayer, E. A., H. Chanzy, R. Lamed and Y. Shoham. 1998. Cellulose, cellulases and cellulosomes. Current Opinion in Structural Biology. **8:**548-557.

Benoit, I., H. Culleton, M. Zhou, M. DiFalco, G. Aguilar-Osorio, E. Battaglia, O. Bouzid, C. P. J. M. Brouwer, H. B. O. El-Bushari, P. M. Coutinho, B. S. Gruben, K. S. Hildén, J. Houbraken, L. A. J. Barboza, A. Levasseur, E. Majoor, M. R. Mäkelä, H. -M Narang, B. Trejo-Aguilar, J. van den Brink, P. A. vanKuyk, A. Wiebenga, V. McKie, B. McCleary, A. Tsang, B. Henrissat and R. P. de Vries. 2015. Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. Biotechnology for Biofuels. **8:**107-121.

Berka, R. M., I. V. Grigoriev, R. Otillar, A. Salamov, J. Grimwood, I. Reid, N. Ishmael, T. John, C. Darmond, M. -C. Moisan, B. Henrissat, P. M. Coutinho, V. Lombard, D. O. Natvig, E. Lindquist, J. Schmutz, S. Lucas, P. Harris, J. Powlowski, A. Bellemare, D. Taylor, G. Butler, R. P. de Vries, I. E. Allijn, J. van den Brink, S. Ushinsky, R. Storms, A. J. Powell, I. T. Paulsen, L. D. H. Elbourne, S. E. Baker, J. Magnuson, S. LaBoissiere, A. J. Clutterbuck, D. Martinez, M. Wogulis, A. L. de Leon, M. W. Rey and A. Tsang. 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. Nature Biotechnology. **29:**922-927.

Bhat, M. K. 2000. Cellulases and related enzymes in biotechnology. Biotechnology Advances. **18:**355-83.

Bhat, M. K. and S. Bhat. 1997. Cellulose degrading enzymes and their potential industrial applications. Biotechnology Advances. **15:**583-620.

Bien-Cuong, D., D. Thi-Thu, J.–G. Berrin, D. Haltrich, T. Kim-Anh, J.–C. Sigoillot and M. Yamabhai. 2009. Cloning, expression in *Pichia pastoris*, and characterization of a thermostable GH5 mannan endo-1,4-β-mannosidase from *Aspergillus niger* BK01. Microbial Cell Factories. **8:**doi:10.1186/1475-2859-8-59.

Bradford, M. M. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry. **72:**248-252.

Brodeur, G., E. Yau , K. Badal, J. Collier, K. B. Ramachandran and S. Ramakrishnan. 2011. Chemical and physicochemical pretreatment of lignocellulosic biomass: a review. Enzyme Research. **2011:**1-17.

Brouwer, H., P. M. Coutinho, B. Henrissat and R. P. de Vries. 2014. Carbohydrate-related enzymes of important *Phytophthora* plant pathogens. Fungal Genetics and Biology. **72:**192-200.

Brown Jr., R. M. 1999. Cellulose structure and biosynthesis. Pure and Applied Chemistry. **71:**767-775.

Brown Jr., R. M. 2003. Cellulose structure and biosynthesis: What is in store for the 21st century?. Journal of Polymer Science. Part A Polymer Chemistry. **42:**487-495.

Burgess, R. R. 2009. Chapter 20 Protein Precipitation Techniques Methods, *In*: Enzymology, 1st ed. Elsevier Inc.

Busk, P. K. and L. Lange. 2013. Function-based classification of carbohydrate-active enzymes by recognition of short, conserved peptide motifs. Applied and Environmental Microbiology. **79:**3380-3391.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research. **37:**D233-D238.

Cantrell, K. B., T. Ducey, K. S. So and O. G. Hunt. 2008. Livestock waste-to-biogenery generation opportunities. Bioresourse Technology. **99:**7941-7953.

Carbohydrate active enzymes database. http://www.cazy.org/. Accessed: 15/03/16.

Cardona, C., J. Quintero and I. C. Paz. 2010. Production of bioethanol from sugarcane bagasse: Status and perspectives. Bioresourse Technology. **101:**4754-66.

Chen, H. and W. Qui. 2010. Key technologies for bioethanol production from lignocellulose. Biotechnology Advances. **28:**556-562.

Cockerill, S. and C. Martin. 2008. Are biofuels sustainable? The EU perspective. Biotechnology for Biofuels **1:**1-9.

Corporation, I.E. 2013. Cellulose-degrading enzyme composition comprising GH16. WO2013177714 A1.

Dassa, B., I. Borovok, R. Lamed, B. Henrissat, P. Coutinho, C. L. Hemme, Y. Huang, J Zhou and E. A. Bayer. 2012. Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. BMC Genomics **13:**210-223.

Dave, B. R., A. P. Sudhir and R. B. Subramanian. 2015. Purification and properties of an endoglucanase from *Thermoascus aurantiacus*. Biotechnology Reports **6:**85-90.

Davies, G. and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. Structure. **3:**853-859.

de Vries, R. P. and J. Visser. 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. Microbiology and Molecular Biology Reviews. **65:**497-522.

Demain, A. L. and P. Vaishnav. 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances. **27**:297-306.

Demirbras, A. 2005. Bioethanol from cellulosic materials: A renewable motor fuel from biomass. Energy Sources. **27:**327-337.

Den Haan, R., J. E. Mcbride, D. C. la Grange, L. R. Lynd and W. H. van Zyl. 2007b. Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. Enzyme and Microbial Technology. **40:**1291-1299.

Den Haan, R., S. H. Rose, L. R. Lynd and W. H. van Zyl WH. 2007a. Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metabolic Engineering. **9:**87-94.

Din, N., H. G. Damude, N. R. Gilkes, R. C. Miller, R. A. J. Warren and D. G Kilburn. 1994. C1–Cx revisited: intramolecular synergism in a cellulase. Proceedings of the National Academy of Sciences. **91:**11383-11387.

Din, N., N. R. Gilkes, B. Tekant, R. C. Miller, R. Anthony Jr, R. A. J. Warren and G. D. Kilburn. 1991. Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Nature Biotechnology. **9:**1096-1099.

Divne, C., J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. K. Knowles, T. T. Teeri and T. A. Jones. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science **265:**524-528.

Do, T. H., T. T. Nguyen, T. N. Nguyen, Q. G. Le, C. Nguyen, K. Kimura and N. H. Truong. 2014. Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. Journal of Bioscience and Bioengineering. **118:**665-671.

Doolotkeldieva, T. D and S. T. Bobusheva. 2011. Screening of wild-type fungal isolates for cellulolytic activity. Microbiology Insights. **4:**1-10.

Duong-Ly, K. C. and S. B. Gabelli. 2014. Salting out of proteins using ammonium sulfate precipitation laboratory methods *In:* Enzymology: Protein Part C, 1st ed. Elsevier Inc.

Durrant, L. R. 1996. Biodegradation of lignocellulosic materials by soil fungi isolated under anaerobic conditions. International Biodeterioration & Biodegradation **37:**189-195.

Dutta, T., R. Sahoo, R. Sengupta, S. S. Ray, A. Bhattacharjee and S. Ghosh. 2008. Novel cellulases from an extremophilic filamentous fungi *Penicillium citrinum*: production and characterization. Journal of Industrial Microbiology and Biotechnology. **35:**275-282.

Dweba, C. 2013. Partial purification and characterization of endoglucanases from *Phialophora alba*. Honours Dissertation, University of KwaZulu-Natal.

Elakkiya, P. and V. Muralikrishnan. 2014. Cellulase production and purification of mutant strain *Trichoderma viride*. International Journal of Current Microbiology and Applied Sciences. **3:**720-727.

Elshafei, A. M., M. M. Hassan, B. M. Haroun and O. M. Abdel-fatah. 2009. Purification and properties of an endoglucanase of *Aspergillus terreus* DSM 826. Journal of Basic Mircobiology. **49:**426-432.

Elshahed, M. S. 2010. Microbiological aspects of biofuel production: Current status and future direction. Journal of Advances Research. **1:**103-111.

Eriksson, K. E. and T. M. Wood. 1985. *In*: Biosynthesis and Biodegradation of Wood Components (T. Higuchi ed.). 469 – 503. Academic Press, New York.

Fan, J., Q. Yang, X. Chen, Z. Liu, Y. Wang and D. Wang. 2012. Cloning, expression , and characterization of endoglucanase gene egIV from *Trichoderma viride* AS 3.3711. Journal of Microbiology and Biotechnology. **22:**390-399.

Farinas, C. S., L. M. Scarpelini, E. Miranda and V. B. Neto. 2011. Evaluation of operational parameters on the precipitation of endoglucanase and xylanase produced by solid state fermentation of *Aspergillus niger*. Brazilian Journal of Chemical Engineering. **28:**17-26.

Fernandes, S. and P. Murray. 2010. Metabolic engineering for improved microbial pentose fermentation. Bioengineered Bugs. **1:**424-428.

Fernandez-fueyo, E., F. J. Ruiz-dueñas, P. Ferreira, D. Floudas, D. S. Hibbett, U. Kües, C. Hori, K. Igarashi, M. Samejima, B. W. Held and K. W. Barry. 2012. Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. Proceesings of the National Academy of Sciences of the USA. 109:5458-5463.

Fickers, P. 2014. *Pichia pastoris*: a workhorse for recombinant protein production. Current Research in Microbiology and Biotechnology. **2:**354-363.

Fortman, J. L., S. Chhabra, A. Mukhopadhyay, H. Chou, T. S. Lee, E. Steen and J. D. Keasling. 2008. Biofuel alternatives to ethanol: pumping the microbial well. Trends in Biotechnology. **26:**375-381.

Gao J, Weng H, Zhu D, Yuan M, Guan F, Xi Y. 2008. Production and characterization of cellulolytic enzymes from the thermoacidophilic fungal *Aspergillus terreus* M11 under solid-state cultivation of corn stover. Bioresourse Technology. **99:**7623-9.

Gautam, S. P., P. S. Bundela, A. K. Pandey, J. Khan, M. K. Awasthi and S. Sarsaiya. 2011. Optimization for the production of cellulase enzyme from municipal solid waste residue by two novel cellulolytic fungi. Biotechnology Research International. **2011:** doi:10.4061/2011/810425.

Ghattyvenkatakrishna, P. K., E. M. Alekozai, G. T. Beckham, R. Schulz, M. F. Crowley, E. C. Uberbacher and X. Cheng. 2013. Initial recognition of a cellodextrin chain in the cellulose-binding tunnel may affect cellobiohydrolase directional specificity. Biophysical Journal. **104:**904-912.

Gilna, V. V and K. M. Khaleel. 2011. Cellulase enzyme activity of *Aspergillus fumigatus* from mangrove soil on lignocellulosic substrate. Recent Research in Science and Technology. **3:**132-134.

Goedegebuur, F., T. Fowler, J. Phillips, P. van der Kley, P. van Solingen, L. Dankmeyer and S. D. Power. 2002. Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. Current Genetics. **41:**89-98.

Gomes, I., J. Gomes, D. J. Gomes and W. Steiner. 2000. Simultaneous production of high activities of thermostable endoglucanase and β-glucosidase by the wild thermophilic fungus *Thermoascus aurantiacus*. Applied Microbiology and Biotechnology. **53:**461-468.

González-Mendoza, D., R. Argumedo-Delira, A. Morales-Trejo, A. Pulido-Herrera, L. Cervantes-Díaz, O. Grimaldo-Juarez, A. Alarcón. 2010. A rapid method for isolation of total DNA from pathogenic filamentous plant fungi. Genetics and Molecular Research **9:**162-166.

Gortari, M. C. and R. A. Hours. 2008. Fungal chitinases and their biological role in the antagonism onto nematode eggs. A review. Mycological Progress. **7:**221-238.

Govender, L. 2012. Season variation and the effects on quality woodchips. Masters Dissertation. University of KwaZulu-Natal.

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

Grassick, A., P. G. Murray, R. Thompson, C. M. Collins, L. Byrnes, G. Birrane, T. M. Higgins and M. G. Tuohy. 2004. Three-dimensional structure of a thermostable native cellobiohydrolase, CBHIB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. European Journal of Biochemistry. **271:**4495-4506.

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

Grigorevski-Lima, A. L., F. N. M. Da Vinha, D. T. Souza, A. S. R. Bispo, E. P. S. Bon, R. R. R. Coelho and R. P Nascimento. 2009. *Aspergillus fumigatus* thermophilic and acidophilic endoglucanases. Applied Biochemistry and Biotechnology. **155:**321-9.

Gusakov, A. V., E. G. Kondratyeva and A. P. Sinitsyn. 2011. Comparison of two methods for assaying reducing sugars in the determination of carbohydrase activities. International Journal of Analytical Chemistry. doi:10.1155/2011/283658

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

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

Häkkinen, M., M. Arvas, M. Oja, N. Aro, M. Penttilä, M. Saloheimo and T. M. Pakula. 2012. Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. Microbial Cell Factories. **11:**134-146.

Hallenbeck, P. C. and D. Ghosh. 2009. Advances in fermentative biohydrogen production: the way forward? Trends in Biotechnology. **27:**287-297.

Hamelinck, C.N., G. van Hooijdonk and A. P. Faaij. 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. Biomass and Bioenergy. **28:**384-410.

Hartl, L., A. Gastebois, V. Aimanianda and J. P. Latgé. 2011. Characterization of the GPI-anchored endo β-1,3-glucanase Eng2 of *Aspergillus fumigatus*. Fungal Genetics and Biology. **48:**185-191.

Hendriks, T. W. M. and G. Zeeman. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresource Technology. **100:**10-8.

Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochemical Journal. **280:**309-316.

Henrissat, B. and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid similarities. Biochemical Journal. **293:**781-788.

Henrissat, B. and G. Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. Current Opinions in Structural Biology. **7:**637-644.

Herculano, P. N., T. S. Porto, M. H. C. Maciel, K. A. Moreira, C. M. Souza-Motta and A. L. F. Porto. 2012. Partitioning and purification of the cellulolytic complex produced by *Aspergillus japonicus* URM5620 using PEG–citrate in an aqueous two-phase system. Fluid Phase Equilibria. **335:**8-13.

Hilge, M., S. M. Gloor, W. Rypniewski, O. Sauer, T. D. Heightman, W. Zimmermann, K. Winterhalter and K. Piontek. 1998. High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca* - substrate specificity in glycosyl hydrolase family 5. Structure. **6:**1433-1444.

Hölker, U. and J. Lenz. 2005. Solid-state fermentation — are there any biotechnological advantages? Current Opinions in Microbiology. **8:**301-306.

Hong, J., H. Tamaki, K. Yamamoto and H. Kumagai. 2003. Cloning of a gene encoding a thermo-stable endo-beta-1,4-glucanase from *Thermoascus aurantiacus* and its expression in yeast. Biotechnology Letters. **25:**657-61.

Hori, C., J. Gaskell, K. Igarashi, M. Samejima, D. Hibbett, B. Henrissat and D. Cullen. 2013. Genomewide analysis of polysaccharides degrading enzymes in 11 white- and brown-rot *Polyporales* provides insight into mechanisms of wood decay. Mycologia. **105:**1412-1427.

Houbraken, J., H. Spierenburg and J. C. Frisvad. 2012. *Rasamsonia*, a new genus comprising thermotolerant and thermophilic *Talaromyces* and *Geosmithia* species. Antonie Van Leeuwenhoek. **101:**403-421.

Huang, X.-M., Q. Yang, Z.-H. Liu, J.-X. Fan, X.-L. Chen, J.-Z. Song and Y. Wang. 2010. Cloning and heterologous expression of a novel endoglucanase gene *egVIII* from *Trichoderma viride* in *Saccharomyces cerevisiae*. Applied Biochemistry and Biotechnology. **162:**103-115.

Huy, N.D., C. L. Nguyen, H. Park, N. H. Loc, M. Choi, D. Kim, J. Seo and S. Park. 2015. Characterization of a novel manganese dependent endoglucanase belongs in GH family 5 from *Phanerochaete chrysosporium*. Journal of Bioscience and Bioengineering. **121:**154-159. ISBN: 978-953-51-0008-9. InTech. DOI:10.5772/22268.

Ja'afaru, M. I. 2013. Screening of fungi isolated from environmental samples for xylanase and cellulase production. ISRN Microbiology. http://dx.doi.org/10.1155/2013/283423.

Jeffries, T. W. 1994. Biodegradation of lignin and hemicelluloses. Biochemistry. **8:**233-277.

Jian, S., R. R. Sharma-Shivappa, M. Chinn and N. Howell. 2008. Effect of microbial pretreatment on enzymatic hydrolysis and fermentation of cotton stalks for ethanol production. Biomass Bioenergy. **33:**88-96

Jimenez, A. V., H. Wang and B. D. Siegfried. 2014. Expression and characterization of a recombinant endoglucanase from western corn rootworm in *Pichia pastoris*. Journal of Insect Research. **14:**DOI: 10.1093/jisesa/ieu104.

Juturu, V. and J. C. Wu. 2014. Microbial cellulases: Engineering, production and applications. Renewable and Sustainable Energy Review. **33:**188-203.

Kaar, W. E., C. V. Gutierrez and C. M. Kinoshita. 1998. Steam explosion of sugarcane bagasse as a pretreatment for conversion to bioethanol. Biomass and Bioenergy. **14:**277-287.

Karnchanatat, A., A. Petsom, P. Sangvanich, J. Piapukiew, A. J. S. Whalley, C. D. Reynolds, G. M. Gadd and P. Sihanonth. 2008. A novel thermostable endoglucanase from the wood-decaying fungus *Daldinia eschscholzii* (Ehrenb.:Fr.) Rehm. Enzyme and Microbial Technology. **42:**404-413.

Khosla, C. 2006. Microbial synthesis of biodiesel. Global Climate & Energy Project. Stanford University.

King, B. C., K. D. Waxman, N. V. Nenni, L. P. Walker, G. C. Bergstrom and D. M. Gibson. 2011. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnology for Biofuels. **4:**4-18.

Kleywegt, G. J., J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Stâhlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri and T. A. Jones. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 A resolution, and a comparison with related enzymes. Journal of Molecular Biology. **272:**383-97.

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

Krainer, F. W., C. Dietzsch, T. Hajek, C. Herwig, O. Spadiut and A. Glieder. 2012. Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microbial Cell Factories. **11:**22-36.

Krisch, J., M. Takó, T. Papp and C. Vágvölgyi. 2010. Characteristics and potential use of β -glucosidases from Zygomycetes. Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology. **2010:**891-896.

Kuhad, R. C., R. Gupta and A. Singh. 2011. Microbial cellulases and their industrial applications. Enzyme Research. **2011:**doi:10.4061/2011/280696.

la Grange, D.C., R. den Haan and W. H. van Zyl. 2010. Engineering cellulolytic ability into bioprocessing organisms. Applied Microbiology and Biotechnology. **87:**1195-1208.

Laemmli, U. 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature. **22:**680-685.

Lambertz, C., M. Garvey, J. Klinger, D. Heesel, H. Klose, R. Fischer and U. Commandeur. 2014. Challenges and advances in the heterologous expression of cellulolytic enzymes : a review. Biotechnology for Biofuels. **7:**135-145.

Langston, J. A., T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko and M. D. Sweeney. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. Applied and Environmental Microbiology. **77:**7007-7015.

Lee, I., B. R. Evans, L. M. Lane and J. Woodward. 1996. Substrate–enzyme interactions in cellulase systems. Bioresourse Technology. **58:**163-169.

Levasseur, A., A. Lomascolo, O. Chabrol, F. J. Ruiz-Dueñas, E. Boukhris-Uzan, F. Piumi, U. Kües, A. F. J. Ram, C. Murat, M. Haon, I. Benoit, Y. Arfi, D. Chevret, E. Drula, M. J. Kwon, P. Gouret, L. Lesage-Meessen, V. Lombard, J. Mariette, C. Noirot, J. Park, A. Patyshakuliyeva, J. C. Sigoillot, A. Wiebenga, H. A. B. Wösten, F. Martin, P. M. Coutinho, R. P. de Vries, A. T. Martínez, C. Klopp, P. Pontarotti, B. Henrissat and E. Record. 2014. The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. BMC Genomics. **15:**486-510.

Li W., W.-W. Zhang, M.-M Yang and Y.-L. Chen. 2008. Cloning of the thermostable cellulase gene from newly isolated *Bacillus subtilis* and its expression in *Escherichia coli*. Molecular Biotechnology. **40:**195-201.

Li, D. C., M. Lu, Y. L. Li and J. Lu. 2003. Purification and characterization of an endocellulase from the thermophilic fungus *Chaetomium thermophilum* CT2. Enzyme and Microbial Technology. **3:**932-937.

Li, D.-C., A.-N. Li and A. C. Papageorgiou. 2011. Cellulases from thermophilic fungi: recent insights and biotechnological potential. Enzyme Research. doi:10.4061/2011/308730

Li, J., C. Tang, H. Shi H and M. Wu. 2011. Cloning and optimized expression of a neutral endoglucanase gene (*ncel5A*) from *Volvariella volvacea* WX32 in *Pichia pastoris*. Journal of Bioscience and Bioengineering. **111:**537-540.

Li, S. C., N. K. Goto, K. A. Williams and C. M. Deber. 1996. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. Proceedings of the National Academy of Sciences of the USA. **93:**6676-6681.

Li, X., R. Kondo and K. Sakai. 2002. Biodegradation of sugarcane bagasse with marine fungus *Phlebia* sp. MG-60. Journal of Wood Science. **48:**159-162.

Liao, H., H. Zheng, S. Li, Z. Wei, X. Mei, H. Ma, Q. Shen and Y. Xu. 2015. Functional diversity and properties of multiple xylanases from *Penicillium oxalicum* GZ-2. Scientific Reports. **5:**DOI: 10.1038/srep12631.

Lindenmuth, B. E. and K. McDonald. 2011. Production and characterization of *Acidothermus cellulolyticus* endoglucanase in *Pichia pastoris*. Protein Expression and Purification. **77:**153-158.

Liu, D., R. Zhang, X. Yang, H. Wu, D. Xu, Z. Tang and Q. Shen. 2011. Thermostable cellulase production of *Aspergillus fumigatus* Z5 under solid-state fermentation and its application in degradation of agricultural wastes. International Biodeterioration and Biodegradation. **65:**717-725.

Liu, D., R. Zhang, X. Yang, Y. Xu, Z. Tang, W. Tian and Q. Shen. 2011. Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi *Aspergillus fumigatus* Z5 isolated from compost. Protein Expression and Purification. **79:**176-86.

Lombard, V., H. G. Ramulu, E. Drula, P. M. Coutinho and B. Henrissat. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. **42:**D490-D495.

Lynd, L. R., P. J. Weimer, W. H. van Zyl and I.S. Pretorius. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiology and Molecular Biology Reviews. **66:**506-577.

Lynd, L. R., W. H. van Zyl, J. E. McBride and M. Laser. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current Opinions in Biotechnology. **16:**577-583.

Macauley-Patrick, S., M. L. Fazenda, B. McNeil and L. M. Harvey. 2005. Heterologous protein production using the *Pichia pastoris* expression system. Yeast. **22:**249-270.

Maheshwari, R., G. Bharadwaj and M. K. Bhat. 2000. Thermophilic fungi: Their physiology and enzymes. Microbiology and Molecular Biology Reviews. **64:**461-488.

Maki, M., K. T. Leung and W. Qin. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. International Journal of Biological Sciences. **5:**500-516.

Margeot, A., B. Hahn-Hägerdal, M. Edlund, R. Slade and F. Monot. 2009. New improvements for lignocellulosic ethanol. Current Opinion in Biotechnology. **20:**372-380.

Martín, C., Y. González, T. Fernández and A. B. Thomsen. 2006. Investigation of cellulose convertibility and ethanolic fermentation of sugarcane bagasse prettreatment by wet oxidation and steam explosion. Journal of Chemical Technology and Biotechnology. **81:**1669-1677.

Mathew, G. M., R. K Sukumaran, R. R. Singhania and A. Pandey. 2008. Progress in research on fungal cellulases for lignocellulose degradation. Journal of Scientific and Industrial Research. **67:**898-907.

Mbandlwa, P. 2013. Characterization of the lignocellulosic arsenal from *Phialophora alba*. Honours Dissertation, University of KwaZulu-Natal.

McGinnis, K. and D. B. Wilson. 1993. Disulfide arrangement and chemical modification of beta-1,4-endoglucanase E2 from *Thermomonospora fusca*. Biochemistry. **32:**8151-8156.

Mchunu, N. P., K. Permaul, A. Y. A. Rahman, J. A. Saito, S. Singh and M. Alam. 2013. Xylanase superproducer: Genome sequence of a compost-loving thermophilic fungus, *Thermomyces lanuginosus* Strain SSBP. Genome Announcements. **1:**4-5.

Menon, V. and M. Rao. 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. Progress in Energy and Combustion Science. **38:**522-550.

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

Mosier, N., C. Wyman, B. Dale, R. Elander, Y. Y. Lee, M. Holtzapple and M. Ladisch. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresource Technology. **96:**673-86.

Mosina, N. L. 2013. Partial purification and characterization of a *Phialophora alba* xylanase and its application to pretreated sugarcane bagasse. Masters Dissertation, University of KwaZulu-Natal.

Muanya, C. 2009. Microbes offer climate friendly solution to energy crisis. The Guardian, 30 January 2009.

Munoz, I. G., W. Ubhayasekera, H. Henriksson, I. Szabo, G. Pettersson, G. Johansson, S. L. Mowbray and J. Stahlberg. 2001. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. Journal of Molecular Biology. **314:**1097-1111.

Murray, P. G., A. Grassick, C. D. Laffey, M. M. Cuffe, T. Higgins, A. V. Savage, A. Planas and M. G. Tuohy. 2001. Isolation and characterization of a thermostable endo-β-glucanase active on 1,3-1,4-β-D-glucans from the aerobic fungus *Talaromyces emersonii* CBS 814.70. Enzyme and Microbial Technology. **29:**90-98.

Naika, G. S., P. Kaul and V. Prakash. 2007. Purification and characterization of a new endoglucanase from *Aspergillus aculeatus*. Journal of Agricultural and Food Chemistry. **55:**7566-72.

Narra, M., G. Dixit, J. Divecha, K. Kumar, D. Madamwar and A. R. Shah. 2014. Production, purification and characterization of a novel GH 12 family endoglucanase from *Aspergillus terreus* and its application in enzymatic degradation of delignified rice straw. International Biodeterioration and Biodegradation. **88:**150-161.

Nigam A. and A. Ayyagari A. 2007. To estimate DNA by diphenylamine method. *In:* Lab Manual in Biochemistry, Immunology and Biotechnology. McGraw-Hill Education, India, 86-87.

O'Sullivan A. C. 1997. Cellulose: The structure slowly unravels. Cellulose. **4:**173-207.

Onsori, H., M. R. Zamani, M. Motallebi and N. Zarghami. 2005. Identification of over producer strain of endo- 1,4-glucanase in *Aspergillus* species : Characterization of crude carboxymethyl cellulase. Journal of Biotechnology. **4:**26-30.

Onyike, E., R. Auta and A. J. Nok. 2008. Isolation, partial purification and characterization of endoglucanase using corn cobs as carbon source. Nigerian Journal of Biochemistry and Molecular Biology. **23:**1-11.

Ospina-giraldo, M. D., J. G. Griffith, E. W. Laird and C. Mingora. 2010. The CAZyome of *Phytophthora* spp .: A comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. BMC Genomics. **11:**525-541.

Padilha, I. Q. M., L. C. T. Carvalho, P. V. S. Dias, T. C. S. L. Grisi, F. L. Honorato da Silva, S. F. M. Santos and D. A. M. Araújo. 2015. Production and characterization of thermophilic carboxymethyl cellulase synthesized by *Bacillus* sp. growing on sugarcane bagasse in submerged fermentation. Brazilian Journal of Chemical Engineering. **32:**35-42.

Panagiotou, G., D. Kekos, B. J. Macris and P. Christakopoulos. 2003. Production of cellulolytic and xylanolytic enzymes by *Fusarium oxysporum* grown on corn stover in solid state fermentation. Industrial Crops and Products. **18:**37-45.

Park, B. H., T. V. Karpinets, M. H. Syed, M. R. Leuze, E. C. Uberbacher. 2010. CAZymes Analysis Toolkit (cat): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. **20:**1574-1584.

Parkkinen, T., A. Koivula, J. Vehmaanpera and J. Rouvinen. 2008. Crystal structures of *Melanocarpus albomyces* cellobiohydrolase Ce17B in complex with cello-oligomers show high flexibility in the substrate binding. Protein Science. **17:**1383-1394.

Parry, N.J., D.E. Beever, E. Owen, W. Nerinckx, M. Claeyssens, J. Van Beeumen and M.K. Bhat. 2002. Biochemical characterization and mode of action of a thermostable endoglucanase purified from *Thermoascus aurantiacus*. Archives of Biochemistry and Biophysics. **404:**243-253.

Payne, C. M., B. C. Knott, H. B. Mayes, H. Hansson, M. E. Himmel, M. Sandgren, J Ståhlberg and G. T. Beckham. 2015. Fungal Cellulases. Chemical Reviews. **115:**1308-1448.

Percival Zhang, Y. –H., M. E. Himmel and J. R. Mielenz. 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnology Advances. **24:**452-81.

Pham, T. H., D. T. Quyen , N. M. Nghiem and T. D. Vu. 2011. Cloning, expression, purification, and properties of an endoglucanase gene (Glycosyl Hydrolase Family 12) from *Aspergillus niger* VTCC-F021 in *Pichia pastoris*. Journal of Microbiology and Biotechnology. **21:**1012-1020.

Poidevin, L., J. Feliu, A. Doan, J. G. Berrin, M. Bey, P. M. Coutinho, B. Henrissat, E. Record and S. Heiss-Blanquet. 2013. Insights into exo- and endoglucanase activities of family 6 glycoside hydrolases from *Podospora anserina*. Applied Environmental Microbiology. **79:**4220-4229.

Pol, D., R. S. Laxman and M. Rao. 2012. Purification and biochemical characterization of endoglucanase from *Penicillium pinophilum* MS 20. Indian Journal of Biochemistry and Biophysics. **49:**189-194.

Quay, D. H. X., F. D. A. Bakar, A. Rabu, M. Said, R. M. Illias, N. M. Mahadi and O. Hassan. 2011. Overexpression , purification and characterization of the *Aspergillus niger* endoglucanase , EglA , in *Pichia pastoris*. Journal of Biotechnology. **10:**2101-2111.

Quay, D. H. X., F. D. A. Bakar, A. Rabu, M. Said, R. M. Illias, N. M. Mahadi and O. Hassan. 2011. Overexpression , purification and characterization of the *Aspergillus niger* endoglucanase, EglA, in *Pichia pastoris*. African Journal of Biotechnology. **10:**2101-2111.

Rabinovich, M. L., M. S. Melnick and A. V. Bolobova. 2002. The structure and mechanism of action of cellulolytic enzymes. Biochemistry. **67:**1026-1050.

Reczye, K., Zs. Szengyel. R. Eklund and G. Zacchi. 1996. Cellulase production by *T. reesei* Bioresource Technology. **57:**25-30.

Rittmann, B. E. 2008. Opportunities for renewable bioenergy using microorganisms. Biotechnology and Bioengineering. **100:**203-212.

Rosano G. L. and E. Ceccarelli. 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Frontiers in Microbiology. **5:**1-17.

Sadasivam* S**. **and A. *Manickam. *1996.* Biochemical methods. 2nd edition, New Age International Ltd. Publisher, New Delhi. 179-186.

Sales, M. R., R. B. de Moura and M. F da Silva. 2011. Cellulase and xylanase production in *Aspergillus* species. Annals of Microbiology. **61:**917-924.

Saloheimo, M., M. Paloheimo, S. Hakola, J. Pere, B. Swanson, E. Nyyssonen, A. Bhatia, M. Ward and M. Penttila. 2002. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. European Journal of Biochemistry. **269:**4202-4211.

Sambrook J and D. W. Russel. 2000. Molecular Cloning, 3-Volume Set : A Laboratory Manual. Cold Spring Harboc Lab Press. New York. 3rd Edition.

Sánchez, C. 2009. Lignocellulosic residues: Biodegradation and bioconversion by fungi. Biotechnology Advances. **27:**185-194.

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

Saqib, A. A. N., M. Hassan, N. F. Khan and S. Baig. 2010. Thermostability of crude endoglucanase from *Aspergillus fumigatus* grown under solid state fermentation (SSF) and submerged fermentation (SmF). Process Biochemistry. **45:**641-646.

Segato, F., A. R. L. Damásio, R. C. de Lucas, F. M. Squina and R. A. Prade. 2014. Genomics review of holocellulose deconstruction by *Aspergilli*. Microbiology and Molecular Biology Reviews. **78:**588-613.

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

Shanmugapriya K, Saravana PS, Manoharan M. 2012. Isolation, screening and partial purification of cellulase from cellulase producing bacteria. International Journal of Advanced Biotechnology and Research. **3:**509-514.

Singh, A. and K. Hayashi. 1995. Microbial cellulases: Protein architecture, molecular properties and biosynthesis. Advances in Applied Microbiology. **40:**1-44.

Singh, R., A. J. Varma, R. S. Laxman and M. Rao. 2009. Hydrolysis of cellulose derived from steam exploded bagasse by *Penicillium* cellulases: Comparison with commercial cellulase. Bioresource Technology. **100:**6679-6681.

Singhania, R. R., A. K. Patel, C. R Soccol and A. Pandey. 2009. Recent advances in solid-state fermentation. Biochemical Engineering Journal. **44:**13-18.

Sorrell S, Speirs J, Bentley R, Brandt A, Miller R. 2010. Global oil depletion: A review of the evidence. Energy Policy **38**:5290-5295.

Starr, C and R. Taggart. 2004. Ecosystems. *In*: Biology: The unity and diversity of life. 10th Edition. Thomson Learning. United States of America.

Sukharnikov, L. O., B. J. Cantwell, M. Podar and I. B. Zhulin. 2011. Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. Trends in Biotechnology. **29:**473-9.

Sukumaran, R. K., R. R. Singhania and A. Pandey. 2005. Microbial cellulases – Production, applications and challenges. Journal of Scientific and Industrial Research. **64:**832-844.

Sul, O.–J., J.–H. Kim, S.–J. Park, Y.-J Son, B.-R Park, D. K. Chung, C.–S. Jeong and I.–S. Han. 2004. Characterization and molecular cloning of a novel endoglucanase from *Trichoderma* sp. C-4. Applied Microbiology and Biotechnology. **66:**63-70.

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

Taherzadeh, M. J. and K. Karimi. 2008. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. International Journal of Molecular Sciences. **9:**917-924.

Taniguchi, M., H. Suzuki, D. Watanabe, K. Sakai, K. Hoshino, and T. Tanaka. 2005 Evaluation of pretreatment with *Pleurotus ostreatus* for enzymatic hydrolysis of rice straw. Journal of Bioscience and Bioengineering. **100:**637-43.

Tao, Y.–M., X.–Z. Zhu, J.–Z. Huang, S.–J. Ma, X.–B. Wu, M.–N. Long and Q.-X Chen. 2010. Purification and properties of endoglucanase from a sugar cane bagasse hydrolyzing strain, *Aspergillus glaucus* XC9. Journal of Agriculture and Food Chemistry **58:**6126-30.

Teather, R. M. and P. J. Wood. 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Applied Environmental Microbiology. **43:**777-780.

Tseng, C. W., T. P. Ko, R. T. Guo, J. W. Huang, H. C. Wang, C. H. Huang, Y. S. Cheng, A. H. J. Wang and J. R. Liu JR. 2011. Substrate binding of a GH5 endoglucanase from the ruminal fungus *Piromyces rhizinflata*. Acta Crystallographica Section F: Structural Biology and Crystallization Communications. **67:**1189-1194.

Umamaheswari, M., M. Jayakumari, K. Maheswari, M. Subashree, P. Mala, T. Sevanthi and T. Manikadan. 2010. Bioethanol production from cellulosic materials. Asian Journal of Science and Technology. **1:**5-11.

van den Brink, J. and R. P. de Vries. 2011. Fungal enzyme sets for plant polysaccharide degradation. Applied Microbiology and Biotechnology. **91:**1477-92.

van Noort, V., B. Bradatsch, M. Arumugam, S. Amlacher, G. Bange, C. Creevey, S. Falk, D. R. Mende, I. Sinning, E. Hurt and P. Bork P. 2013. Consistent mutational paths predict eukaryotic thermostability. BMC Evolutionary Biology **13:**7-20.

van Zyl, W.H., R. den Haan and D. C. la Grange. 2011. Developing organisms for consolidated bioprocessing of biomass to ethanol, Biofuel production-recent developments and prospects, Dr. Marco Aurelio Dos Santos Bernardes (Ed.), ISBN: 978-953-307-478-8.

Várnai, A., C. Tang, O. Bengtsson, A. Atterton, G. Mathiesen and V. G. H. Eijsink. 2014. Expression of endoglucanases in *Pichia pastoris* under control of the GAP promoter. Microbial Cell Factories. **13:**DOI: 10.1186/1475-2859-13-57.

Veneault-Fourrey, C., C. Commun, A. Kohler, E. Morin, R. Balestrini, J. Plett, E. Danchin, P. Coutinho, A. Wiebenga, R. P. de Vries, B. Henrissat and F. Martin. 2014. Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. Fungal Genetics and Biology. **72:**168-181.

Viikari, L., M. Alapuranen, T. Puranen, J. Vehmaanperä and M. Siika-aho. 2007. Thermostable enzymes in lignocelluloses hydrolysis. Advances in Biochemical Engineering and Biotechnology. **108:**121-145.

Vinuselvi, P. and S. K. Lee. 2012. Heterologous expression and extracellular secretion of cellulases in recombinant microbes. Bioethanol, Prof. M. A. P. Lima (Ed.).

Vlasenko, E., M. Schülein, J. Cherry and F. Xu. 2010. Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. Bioresourse Technology. **101:**2405-2411.

Wackett, L. P. 2008. Biomass to fuels via microbial transformations. Current Opinion in Chemical Biology. **12:**187-193.

Wang J.–R., Y.–Y. Li, D.–Ni. Liu, J.–S. Liu, P. Li, L.–Z. Chen and S.–D. Xu. 2014. Codon optimization significantly improves the expression level of α-amylase gene from *Bacillus licheniformis* in *Pichia pastoris*. Biomedical Research International. DOI: 248680.

Wang, Y., X. Wang, R. Tang, S. Yu, B. Zheng and Y. Feng. 2010. A novel thermostable cellulase from *Fervidobacterium nodosum*. Journal of Molecular Catalysis B: Enzymatic. **66:**294-301.

Wilson, B. D. 2009. Cellulases and biofuels. Current Opinions in Biotechnology. **20:**295-299.

Wilson, D. B. (2008). Three microbial strategies for plant cell wall degradation. Annals of the New York Academy of Sciences. **1125:**289-297.

Woodward, J., K. A. Affholter, K. K. Noles, N. T. Troy and S. F. Gaslightwala. 1992. Does cellobiohydrolase II core protein from *Trichoderma reesei* disperse cellulose macrofibrils? Enzyme and Microbial Technology. **14:**625-630.

Xie, B.–B., Q.–L. Qin, M. Shi, L.–L. Chen, Y.–L. Shu, Y. Luo, X.–W. Wang, J.–C. Rong, Z.–T. Gong, D. Li, C.–Y. Sun, G.–M. Liu, X.–W. Dong, X.–H. Pang, F. Huang, W. Liu, X.–L. Chen, B.–C. Zhou, Y.–Z. Zhang and X.–Y. Song. 2014. Comparative genomics provide insights into evolution of *Trichoderma* nutrition style. Genome Biology and Evolution. **6:**379-390.

Xu, Q., A. Singh and M. E. Himmel. 2009. Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. Current Opinions in Biotechnology. **20:**364-371.

Yang, D., H. Weng, M. Wang, W. Xu, Y. Li and H. Yang. 2010. Cloning and expression of a novel thermostable cellulase from newly isolated *Bacillus subtilis* strain I15. Molecular Biology Reports. **37:**1923-1929.

Ye, X. Y., T. B. Ng and K. J. Cheng. 2001. Purification and characterization of a cellulase from the ruminal fungus *Orpinomyces joyonii* cloned in *Escherichia coli*. The International Journal of Biochemistry and Cell Biology. **33:**87-94.

Yennamalli R. M., A. J. Rader, A. J. Kenny, J. D. Wolt and T. Z. Sen. 2013. Endoglucanases: insights into thermostability for biofuel application. Biotechnology for Biofuels. **6:**136-145.

Zhang, F., P. Shi, Y. Bai, H. Luo, H. Huang, P. Yang, L. Miao and B. Yao. 2011. An acid and highly thermostable xylanase from *Phialophora* sp. G5. Applied Microbiology and Biotechnology. **89:** 1851-1858.

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

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

Zhang, Y.-H. P. and L. R. Lynd. 2005. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proceedings of the National Academy. Academy of Sciences of the United States of America. **102:**7321-7325.

Zhang, Y.-H., P., M. E. Himmel and J. R. Mielenz. 2006. Outlook for cellulase improvement: Screening and selection strategies. Biotechnology Advances. **24:**452-481.

Zhang, X.-Z. and Y.-H. P. Zhang. 2013. Cellulases: Characteristics, sources, production and application. Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymers. **First Edition**. Edited by: Shang-Tian Yang, Hesham A. El-Enshasy, and Nuttha Thongchul. John Wiley and Sons, Inc.

Zhao, J., P. Shi, H. Huang, Z. Li, T. Yuan, P. Yang, H. Luo, Y. Bai and B. Yao. 2011. A novel thermoacidophilic and thermostable endo-β-1,4-glucanase from *Phialophora* sp. G5: its thermostability influenced by a distinct β-sheet and the carbohydrate-binding module. Applied Microbiology and Biotechnology. **95:**947-955.

Zhao, J., P. Shi, T. Yuan, H. Huang, Z. Li, K. Meng, P. Yang and B. Yao. 2012c. Purification, gene cloning and characterization of an acidic β-1,4-glucanase from *Phialophora* sp. G5 with potential applications in the brewing and feed industries. Journal of Bioscience and Bioengineering. **144:**379-384.

Zhao, J., P. Shi, Y. Bai, H. Huang, H. Luo, H. Zhang, D. Xu, Y. Wang and B. Yao. 2012a. A thermophilic cellulase complex from *Phialophora* sp. G5 showing high capacity in cellulose hydrolysis. Applied Biochemistry and Biotechnology. **166:**952-60.

Zhao, J., P. Shi, Z. Li, P. Yang, H. Luo, Y. Bai, Y. Wang and B. Yao. 2012b. Two neutral thermostable cellulases from *Phialophora* sp. G5 act synergistically in the hydrolysis of filter paper. Bioresource Technology. **121:**404-10.

Zhao, X., L. Zi, F. Bai, H. Lin and X. Hao. 2011. Bioethanol from Lignocellulosic Biomass. Advances in Biochemical Engineering/Biotechnology. 128:25-51.

Zhao, Z., H. Liu, C. Wang and J. Xu. 2014. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics **15:**6-21.

Zheng, H. and H. Wu. 2010. Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. BMC Bioinformatics **11:**7-17.

Zheng, Y., Z. Pan and R. Zhang. 2009. Overview of biomass pretreatment for cellulosic ethanol production. International Journal of Agricultural and Biological Engineering. **2:**51-68.

Zhong, X. and W. Somers. 2012. Recent advances in glycosylation modifications in the context of therapeutic glycoproteins. *In*: Integrative Proteomics, Dr. Hon-Chiu Leung (Ed.), ISBN: 978-953-51-0070-6, InTech, DOI: 10.5772/29708.

APPENDIX

Appendix A

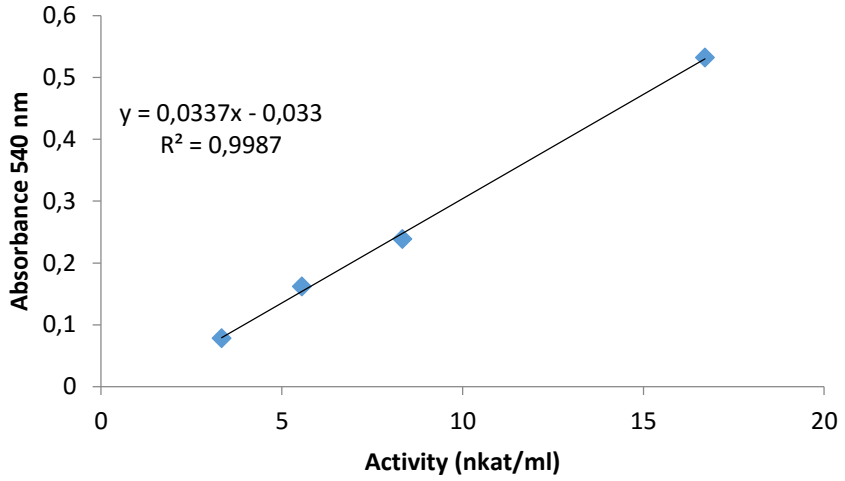


Figure A1: Glucose standard curve for the determination of endoglucanase activity during screening phase

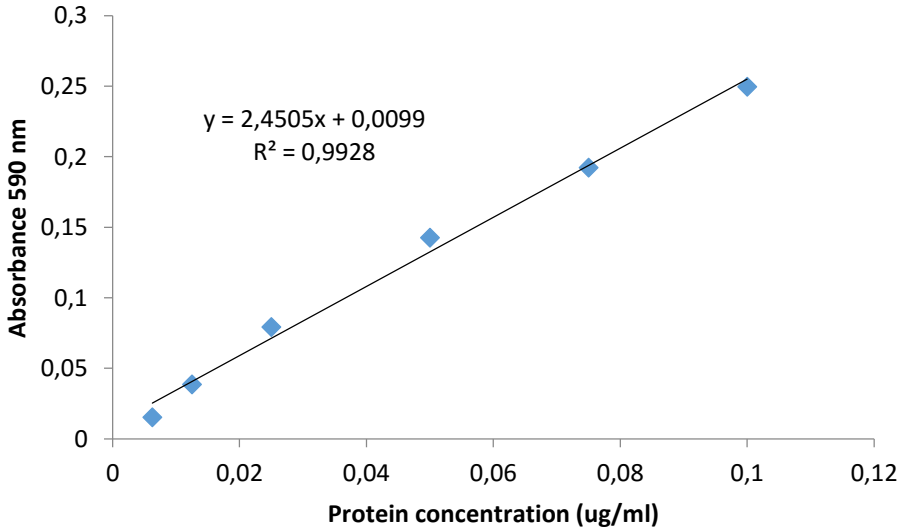


Figure A2: BSA standard curve used for the determination protein concentration *via* the Bradford assay

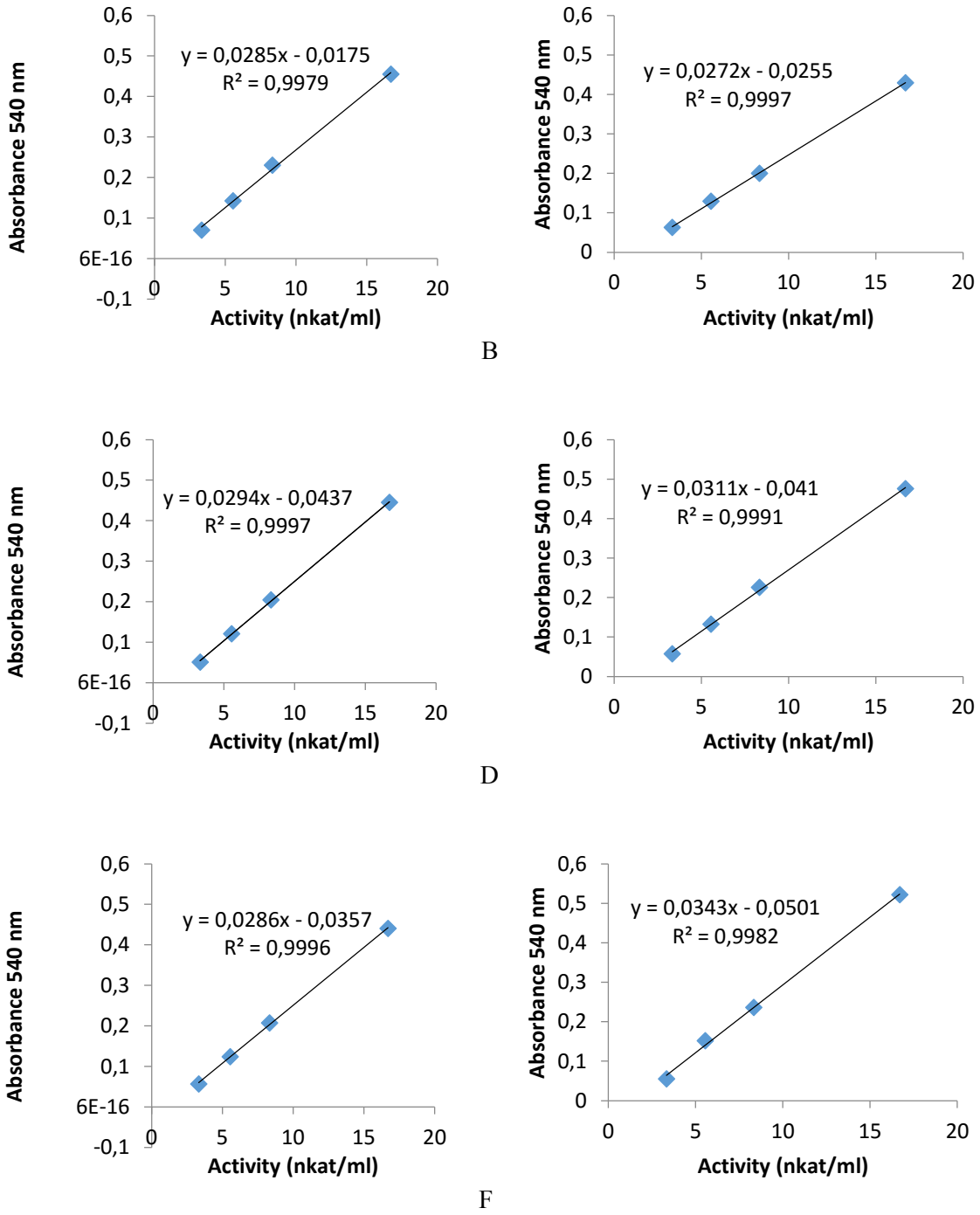
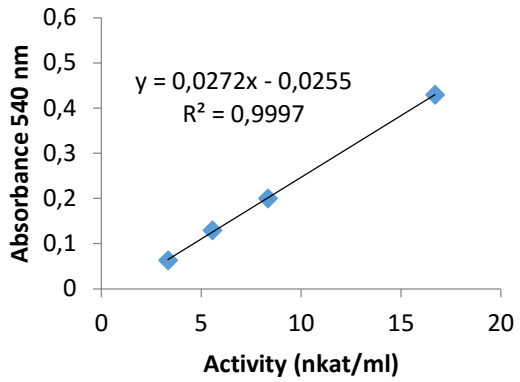
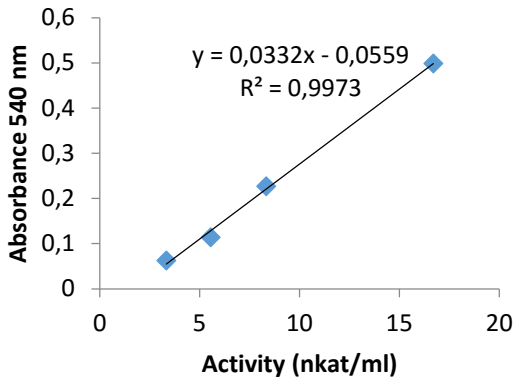


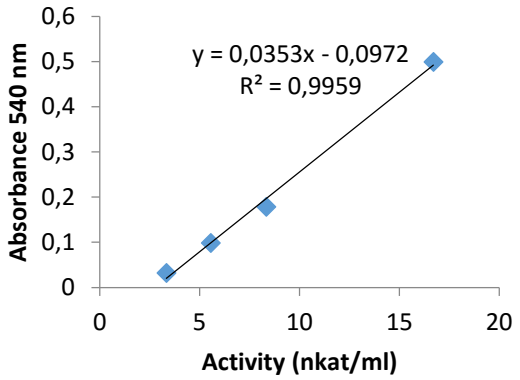
Figure A3: Glucose standard curve for the determination of endoglucanase activity at: A: pH 4; B: pH 5; C: pH 6; D: pH 7; E: pH 8; F: pH 9



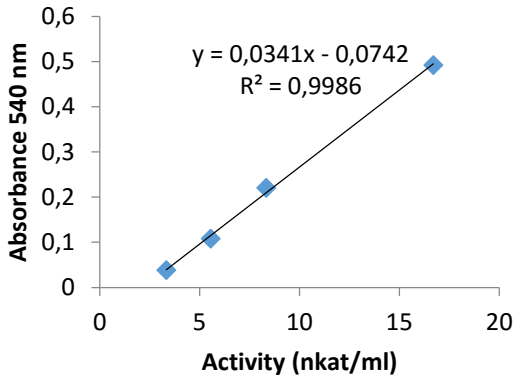
A



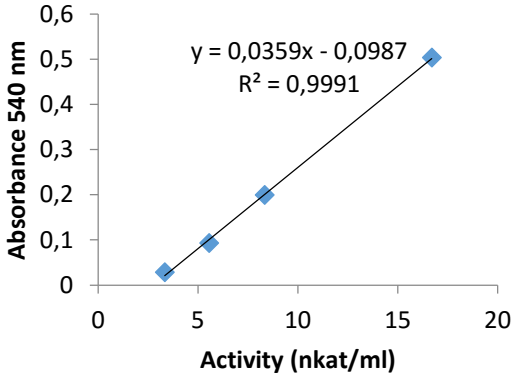
B



C



D



E

Figure A4: Glucose standard curve for the determination of endoglucanase activity at: A: 50°C; B: 60°C; C: 70°C; D: 80°C; E: 90

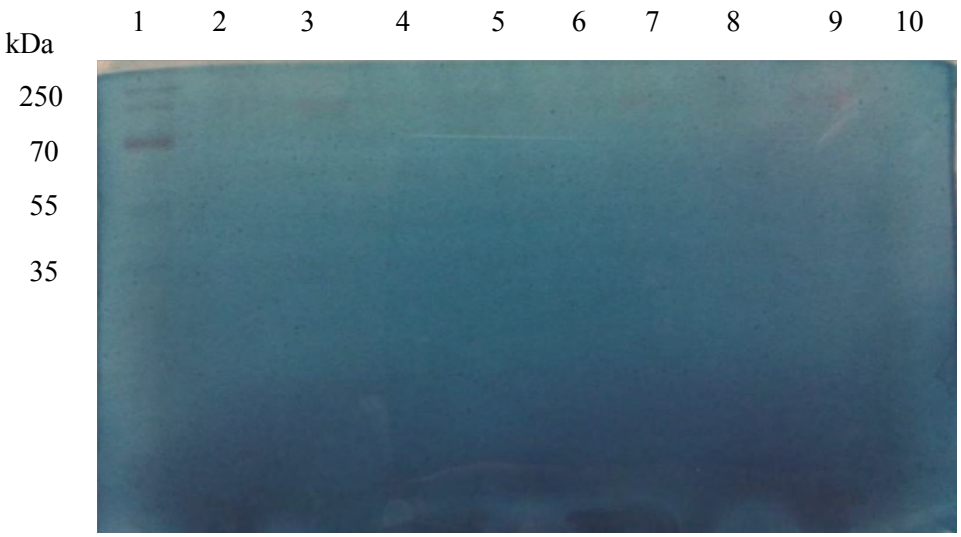


Figure A5: Zymogram analysis of *Pichia* clones extracellular proteins (Lane 1: Marker: PageRuler Plus Prestained Protein Ladder, Lanes: 2-4: *Pichia* pPIC9-no insert Day 0, 2, 4; Lanes: 5-7: *Pichia* pPIC9-S1 Day 0, 2, 4; Lanes: 8-10: *Pichia* pPIC9-B2 Day 0, 2, 4)

Appendix B

Recipes

1. 1 M Tris-HCl, pH 8 (500 ml)
   60.57 g Tris
   500 ml distilled water
   pH adjusted to 8.0 with concentrated HCl
   Autoclaved and stored at room temperature

2. 0.5 M EDTA, pH 8 (1000 ml)
   181.6 g EDTA
   1000 ml distilled
   pH adjusted to pH 8.0 with NaOH pellets
   Autoclaved and stored at room temperature

3. 5 M NaCl (500 ml)
   146.1 g NaCl
   500 ml distilled water
   Autoclaved and stored at room temperature

4. Fungal extraction buffer (10 ml)
   1 ml Tris-HCl
   0.5 ml EDTA
   3 ml NaCl
   5.5 ml distilled water

Appendix C

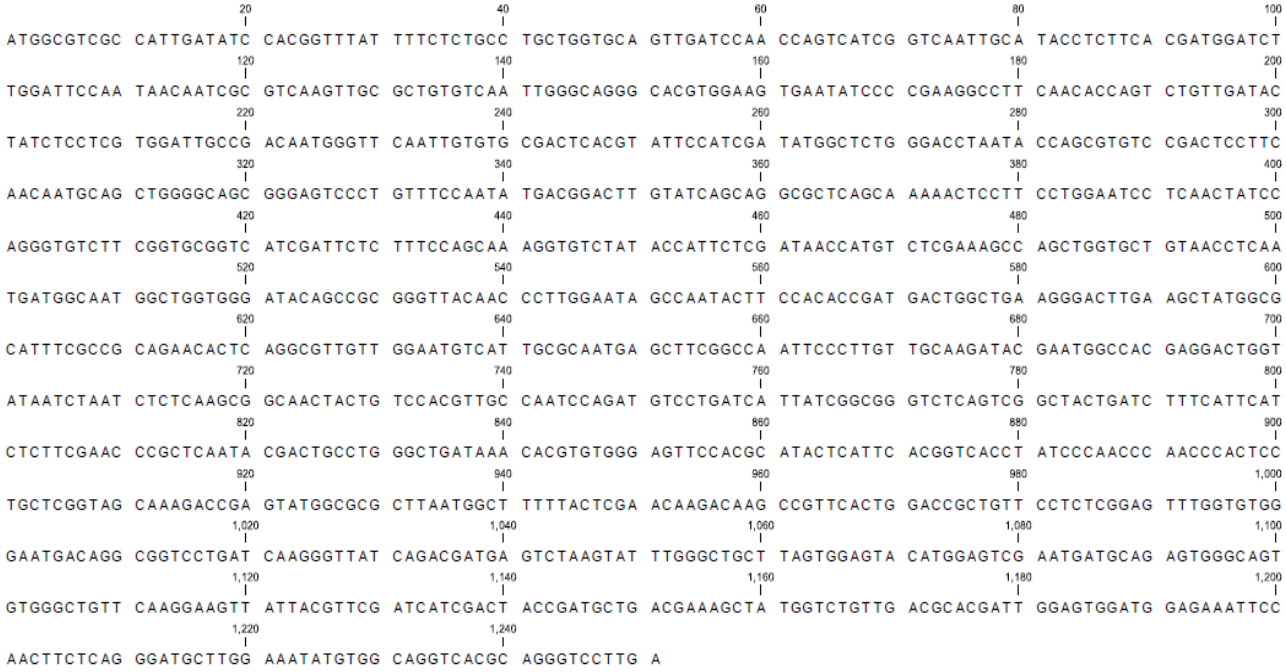


Figure C1: Nucleotide sequence of GH5 endoglucanase gene from *P. alba*