Cloning, Expression and Characterization of Novel Fungal Endoglucanases

HUA NAN REN

A Thesis in the Department of

Biology

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

November 2009

Your file Votre référence
ISBN: 978-0-494-80149-9
Our file Notre référence
ISBN: 978-0-494-80149-9

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: HUA NAN REN

Entitled: **Cloning, Expression and Characterization of Novel Fungal Endoglucanases**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair

_____Examiner

_____Examiner

_____Examiner

_____Supervisor

_____Supervisor

Approved by

_____
Chair of Department or Graduate Program Director

_____ 2009

_____
Dean of Faculty

ABSTRACT

Cloning, Expression and Characterization of Novel Fungal Endoglucanases

Hua Nan Ren

Global energy consumption is projected to double between 1980 and 2020. Furthermore, about 65% of present energy consumption is derived by burning nonrenewable fossil fuels that increase atmospheric levels of CO_2, the major greenhouse gas implicated in global warming. It is therefore important that carbon neutral alternative energy sources be developed. For transportation fuels bioethanol derived from renewable energy sources is considered among the most attractive alternatives. Presently, most bioethanol is produced from corn kernels and sugar cane sucrose, with the United State and Brazil being the world's major producers. Cellulose is another possible feedstock that can be used to produce bioethanol. Compared with the currently used feedstocks, cellulose is very abundant in nature and its use for biofuels production would not compete with the food or animal feed industries. In order to use cellulose for bioethanol production, the cellulose must be converted into the fermentable sugar glucose. Currently, the high cost of converting the cellulose into glucose is a major impediment to using cellulose as a feedstock for the production of bioethanol. The goal of my research was to address this issue by discovering new fungal enzymes that could improve the cellulose hydrolyzing efficiency of existing commercial cellulase systems. In this study I identified, cloned, functionally expressed and characterized six endoglucanases, AfumEgl2010,

AnidEgl2020, FgraEgl2010, FgraEgl1020, FgraEgl2020 and NcraEgl2010. Three of these endoglucanases, AfumEgl2010, AnidEgl2020, FgraEgl2010, were characterized in detail in this study. Only AfumEgl2010 harbours a carbohydrate-binding module (CBM). All three of the endoglucanases that were subject to detailed characterization show highest activity at pH 5.0. The temperature optima of these three endoglucanases were determined to be 40°C, 60°C and 70°C. All of them were stable during 30 minutes pre-incubation at 60°C. The kinetic parameters of these three endoglucanases and four other endoglucanases, the *Trichoderma reesei* (*T. reesei)* Eg2/Cel5A and ApulSEQ15654, StheSEQ13822, GtraSEQ630 (previously identified by undergraduate student Christopher St-Francois) were determined at 37°C and the pH optima of each endoglucanase. The Km values ranged from 2.0 to 29 mg/ml. The Vmax values ranged from 5.7 to 41 μmole/mg/min. The degree of synergism when these endoglucanases were combined with the *T. reesei* Cbh1/Cel7A was also determined.

ACKNOWLEGMENTS

I would like to express my gratitude to my two supervisors Dr. Reg Storms and Dr. Justin Powlowski for their support and patience. I also want to thank the other members of my MSc supervisory committee, Dr. Luc Varin and Dr. Muriel Herrington for giving suggestion and advice throughout my MSc research. I also wish to thank my lab mates, Dr. Humberto Tambor, Yun Zheng, Rebecca Sydenham, Minghui Yang, Dr. Sophia Ushinsky, Dr. Edith M. Munro and Yan Li for their help during my research. I would also like to thank Tricia John and Marc Champagne for their knowledge and technical assistance.

Finally, I wish to give special thanks to my parents and my girlfriend for always giving support and encouragement in all my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa	amino acid
Amp	ampicillin
bp	base pair
BCA	bicinchoninic acid
CBM	carbohydrate-binding module
Cbh	cellobiohydrolase
cDNA	complementary DNA
CM	complete medium
CMC	carboxymethylcellulose
Da	dalton
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
Eg	endoglucanase
GH	glycoside hydrolase
JGI	Joint Genome Institute
Kan	kanamycin
M	molar
MM	minimal media
MMJ	minimal media J
PASC	phosphoric acid swollen cellulose
PCR	polymerase chain reaction
PEG	polyethylene glycol
sec	second

SDS-PAGE	sodium dodecylsufate-polyacrylamide gel electrophoresis
VTO	vector transformants only (*A. niger* transformant obtained with the expression vector without an insert)

1. Introduction

1.1 Traditional and alternative resources used for energy

Energy demand has increased dramatically in the last one hundred years (Sun and Cheng, 2002). To meet the demand of energy consumption, worldwide energy production has increased by 118 percent from 1970 to 2006 (DOE/EIA-0384(2008)). The traditional resources used for energy are mainly fossil fuels including oil, coal and natural gas (http://www.energy.gov/energysources/fossilfuels.htm). In 2006, the energy produced by burning fossil fuels accounted for 86 percent of total energy production in the world (DOE/EIA-0384(2008)). Fossil fuels are finite and take a very long time to generate (DOE/GO-102001-1102). We therefore need to find alternative energy resources. In 2006, renewable energy and nuclear electric power accounted for 8 percent and 6 percent respectively of total energy production in the whole world (DOE/EIA-0384(2008)). Renewable energy is generated from renewable sources such as hydroelectric power, biofuels, wood, wind, geothermal, waste, and solar (DOE/EIA-0384(2008)). Compared with fossil fuels, renewable energy sources can be renewed in a short period of time and their use reduces greenhouse gas emissions (DOE/GO-102001-1102). For example, the use of bioethanol, produced from renewable cellulosic biomass as a transportation fuel can reduce greenhouse gas emission relative to using gasoline produced from fossil fuels (Wang et al., 2007).

1.2 Bioethanol

Bioethanol has the potential to replace gasoline in the transportation sector. In Brazil, the United States, some European countries and Canada, bioethanol is already in use. Essentially all the bioethanol that is presently produced is made by fermentable sugars generated from different types of raw materials (Galbe and Zacchi, 2007).

1.2.1 Bioethanol produced from starch- and sugar- based crops

Currently, bioethanol is generated mainly from starch- and sugar- based crops such as corn and sugar cane (Galbe and Zacchi, 2007). The advantage of using these feedstocks to produce bioethanol is that the cost of production stays competitively low (Galbe and Zacchi, 2007). The disadvantage of using these feedstocks is that starch- and sugar- based crops are agricultural crops, therefore using these feedstocks to produce bioethanol competes with food and feed production (Sun and Cheng, 2002). In the United States, bioethanol is produced mainly from corn and in Brazil the bioethanol is produced mainly from sugar cane (DOE/EIA-0383(2007)). In 2008, the United States produced 9.2 billion gallons of bioethanol (http://www.eia.doe.gov/conference/2009/session5/session5.pdf). The cost of production of bioethanol from corn ranged from 60 cents to 80 cents per liter of gasoline equivalent (http://www.iea.org/techno/essentials2.pdf). Brazil produced more than 6.4 billion gallons of bioethanol in 2008 according to União da Agroindústria Canavieira (Unica) and Renewable Fuels Association (http://www.ethanolrfa.org/industry/statistics/#E). The cost of production of bioethanol using sugar cane ranged from 25 cents to 35 cents per liter of gasoline equivalent (http://www.iea.org/techno/essentials2.pdf).

1.2.2 Bioethanol produced from cellulosic biomass

Another way to produce bioethanol is to use cellulosic biomass from plants, agricultural residues, forestry wastes, and paper waste (DOE/EIA-0383(2007)). The advantages of using cellulosic biomass are that it is very abundant in nature, it does not compete with food and feed production, and it has potential to reduce greenhouse gas emissions (Galbe and Zacchi, 2007; Sun and Cheng, 2002). The review paper by Sun and Cheng (2002) reported that in the past 20 years there has been significant research focusing on the conversion of cellulosic feedstocks to bioethanol. There are two main steps that are required to convert cellulosic feedstocks to bioethanol: 1) hydrolysis of hemicellulose and cellulose of feedstocks to fermentable sugars 2) conversion of the sugar into ethanol through a fermentation step (Sun and Cheng, 2002). Currently, the main challenge of using cellulosic biomass to produce bioethanol as a transportation fuel is the relatively high cost of production of bioethanol (Galbe and Zacchi, 2007). For instance, currently the cost of production of bioethanol using cellulosic biomass is over 1 dollar per liter of gasoline equivalent (http://www.iea.org/techno/essentials2.pdf). One reason for the high cost of bioethanol production is the relatively low efficiency of feedstock conversion to fermentable sugars (Sun and Cheng, 2002).

1.3 Cellulose and hemicellulose

Cellulose is the major component in plant cell walls and it is very abundant in nature (Bayer et al., 1998). Cellulose is a linear polymer of many D-glucose units that are linked by β-(1, 4) - glucosidic bonds (Crawford, 1981; Updegraff, 1969). The individual cellulose chains can hydrogen bond with each other to form a cellulose microfibril (Albersheim, 1975). Cellulose microfibrils can assume both a crystalline structure and an amorphous structure (Purves et al., 1998). Figure 1 shows the structure of a cellulose microfibril.

Hemicellulose is another major component in plant cell walls and accounts for 20 to 35 percent of lignocellulosic biomass (Saha, 2003). Hemicellulose is a branched heteropolymer with a β-(1, 4) linked backbone (Rose, 2003). In plant cell walls, the hemicellulose can be linked to cellulose via hydrogen bonds (Rose, 2003). The linkage between hemicellulose and cellulose makes less cellulose available for enzymatic degradation thereby reducing both the efficiency of the enzymatic hydrolysis of cellulose and the yield of fermentable sugars (Sun and Cheng, 2002).

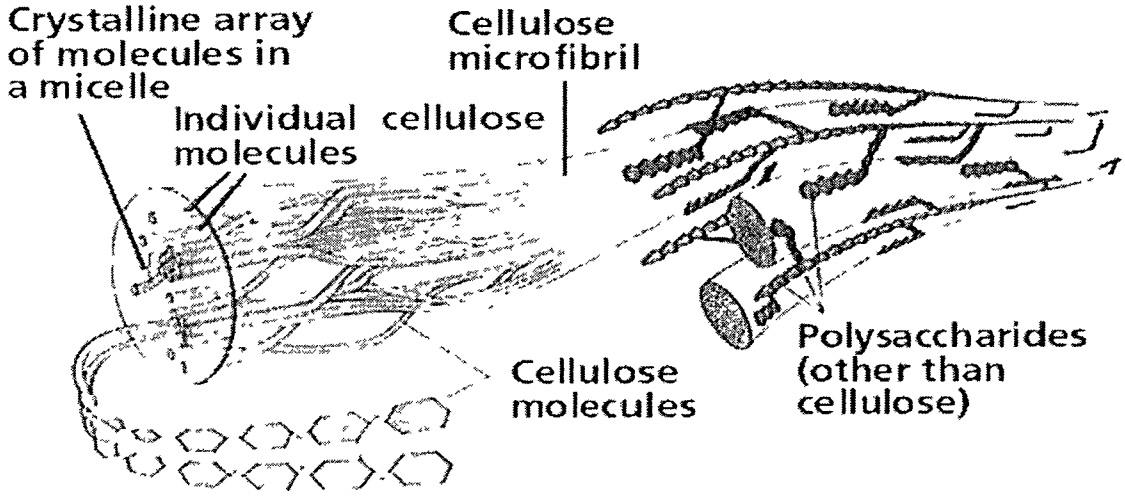


Figure 1: The structure of a cellulose microfibril (Purves et al., 1998). Reprinted from "Life: The Science of Biology, Fifth Edition" Copyright (2009), with permission from SINAUER ASSOCIATES, Inc.

1.4 Cellulases

1.4.1 Background of cellulases and cellulase systems

Enzymes that hydrolyze the β-1, 4-glucosidic bonds in the cellulose chain are called cellulases (Bayer et al., 1998). Cellulases are glycoside hydrolases (EC 3.2.1.-), which are enzymes responsible for the breakdown of glycosidic bonds between carbohydrates (http://www.cazy.org/fam/acc_GH.html). Cellulases are produced by a wide variety of species, such as fungi, bacteria, plants and animals such as termites and crayfish (Lynd et al., 2002). The major role of cellulases in bacterial and fungal species is to break down cellulose into an energy form that they can readily use (Lynd et al., 2002). The function of cellulases in plants involves construction and decomposition of their cell walls (Urbanowicz et al., 2007). Many cellulases have a catalytic domain linked via a shorter linker to a carbohydrate-binding module (CBM) (Lynd et al., 2002). The major function of the CBM of cellulases is to bind to insoluble cellulose, which brings the catalytic domain closer to the insoluble cellulose and therefore increases the efficiency of cellulose hydrolysis (Lynd et al., 2002). Another possible function of the CBM is to partially separate cellulose strands from the surface of cellulose microfibrils, thereby making the cellulose strands more accessible for hydrolysis (Din et al., 1994).

The collection of cellulase enzymes that are produced by a strain of bacteria or fungus are called cellulase enzyme systems (Lynd et al., 2002). In general, there are two different types of cellulase enzyme system. One type is produced by anaerobic bacteria, while the other type is produced by aerobic bacteria and fungi (Lynd et al., 2002). In anaerobic bacteria, the enzymes that are produced form stable high-molecular weight

complexes (Lynd et al., 2002). These complex cellulase systems, also known as cellulosomes, are responsible for the degradation of cellulose by these organisms (Lynd et al., 2002). Aerobic bacterial and fungal organisms produce multiple cellulase enzymes that do not associate to form stable high-molecular weight complexes (Lynd et al., 2002).

1.4.2 Classification of cellulases

There are three major types of cellulases: endoglucanases (EC 3.2.1.4), exoglucanases which are either cellodextrinases (EC 3.2.1.74) or cellobiohydrolases (EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21) (Lynd et al., 2002). Endoglucanases hydrolyze cellulose at internal amorphous sites and reduce the length of the cellulose chain, thus creating more reducing and non-reducing chain ends (Lynd et al., 2002). Exoglucanases hydrolyze the cellulose chain from either the reducing or non-reducing end and the major final products are cellobiose or glucose (Lynd et al., 2002). Another function of exoglucanases is the release of single cellulose molecules from the cellulose microfibril structure (Teeri, 1997). β-glucosidases break down cellobiose and cellodextrins into glucose (Lynd et al., 2002). Figure 2 summarizes the role of three major cellulases in the hydrolysis of cellulose.

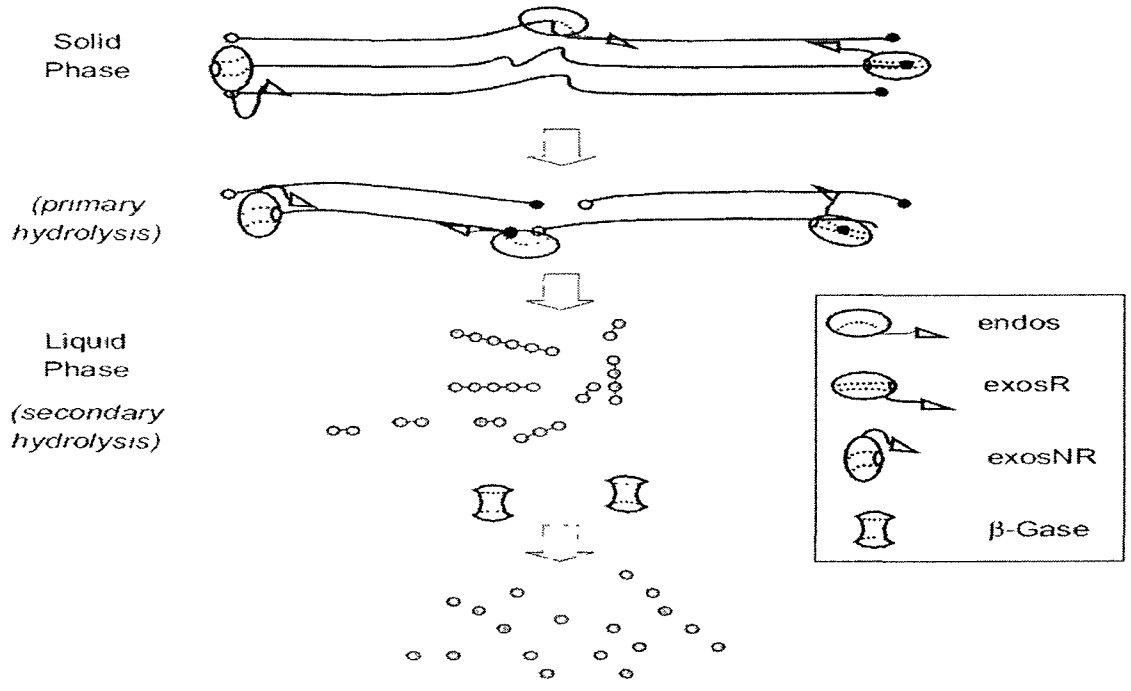


Figure 2: The role of three major cellulases in the hydrolysis of cellulose (Percival Zhang et al., 2006). The endos represents endoglucanases, the exosR and exosNR represents exoglucanases that act at the reducing and non reducing ends of cellulose polymer, and the β-Gase represents β-glucosidases (Percival Zhang et al., 2006). Reprinted from "Outlook for cellulase improvement: Screening and selection strategies" Copyright (2009), with permission from Elsevier.

1.4.3 Classification of endoglucanases (EC 3.2.1.4) of *Trichoderma reesei*

There are five endoglucanases present in the *T. reesei* cellulase system (Lynd et al., 2002). They are Cel7B (EgI), Cel5A (EgII), Cel12A (EgIII), Cel61A (EgIV), and Cel45A (EgV). The most abundant endoglucanases present in the *T. reesei* cellulase system are Cel7B and Cel5A (Lynd et al., 2002). Except for Cel12A, all the major *T. reesei* endoglucanases harbour a CBM (Karlsson et al., 2002). The 3D structures of the catalytic domains of Cel7B and Cel12A of *T. reesei* are already solved and both have similar 3D structures consisting of an open glove structure that can bind to cellulose strands (Kleywegt et al., 1997; Sandgren et al., 2001). The 3D structures of the other 3 *T. reesei* endoglucanases have not been elucidated.

1.4.4 Commercial cellulases in industrial applications

According to the review paper of Lynd et al. (2002), the cellulases that are the most widely used in the cellulosic bioethanol industry are produced from *T. reesei* and *Aspergillus* species. One commercial cellulase system includes minimally two exoglucanases (CBHI and CBHII), five endoglucanases (EgI to EgV), and two β-glucosidases (BglI and BglII) all produced from *T. reesei* (Lynd et al., 2002). In addition, β-glucosidase produced from *Aspergillus* species has been used to supplement the *T. reesei* cellulase in some commercial cellulase preparations (Lynd et al., 2002). The β-glucosidase from *Aspergillus* species is added because it is more tolerant to glucose inhibition than are the two β-glucosidases produced by *T. reesei* (Lynd et al., 2002).

1.5 Synergism

Cellulase synergism occurs when cellulases working together show higher activity levels than the level of activity obtained by adding the activities of the individual cellulases working under the same conditions (Lynd et al., 2002). Four types of synergism have been described (Lynd et al., 2002): 1) synergy between endoglucanases and exoglucanases (Henrissat et al., 1985; Nidetzky et al., 1993) 2) synergy between exoglucanases (Henrissat et al., 1985; Tomme et al., 1988) 3) synergy between exoglucanases (endoglucanases) and β-glucosidases (Lamed et al., 1991; Woodward, 1991) 4) synergy between the CBM and catalytic domain of individual cellulases (Din et al., 1994). It has been suggested that the synergy between endoglucanases and exoglucanases is due to two types of cooperative activity. Firstly, endoglucanases working on the amorphous sites of cellulose chains create more chain ends that can be hydrolyzed by exoglucanases (Henrissat et al., 1985). Secondly, exoglucanases release cellulose molecules from the cellulose microfibril making the cellulose more accessible to endoglucanases (Teeri, 1997). Synergy between exoglucanases occurs because CBHI acts on the reducing ends of the cellulose chain and CBHII acts on the non-reducing ends of cellulose chains (Teeri, 1997). Both endoglucanase and exoglucanase activity can be inhibited by the accumulation of cellobiose (Gruno et al., 2004). Since β–glucosidase can break down cellobiose to glucose, there is synergism when β–glucosidase is added to reactions with endoglucanases or exoglucanases (Lamed et al., 1991). Lastly the intramolecular synergy between the CBM and the catalytic domain of cellulases occurs because CBM binding to the cellulose substrate brings the catalytic domain to the substrate surface and the CBM loosens the crystalline structure by partially separating the

cellulose strands from the surface of cellulose microfibrils making the substrate easier to hydrolyze (Din et al., 1994).

The degree of synergism is the ratio between the measured activity of a cellulase mixture and the sum of individual activity of each cellulase in the mixture (Zhang and Lynd, 2004). There are several factors that can affect the degree of synergism such as the degree of polymerization of the substrate, the amount of enzyme loading, and the reaction time (Zhang and Lynd, 2004).

1.6 Expression vector and expression system

In our laboratory, integrating plasmids are used to express proteins in *Aspergillus niger* (*A niger*) (Storms et al., 2005). The integrating plasmids were constructed using several cassettes: 1) a pUC18 derivative for selection and maintenance in *E. coli*; 2) a pyrG cassette for selection of *A. niger* transformants; 3) an expression cassette that includes the glucoamylase promoter region (GlaPr) and glucoamylase transcription termination region (GlaTt) for directing gene expression; 4) *lacA* reporter gene cassette flanked by unique *NheI* and *FseI* restriction cut sites; and 5) the attR1 and attR2 recombinant cassette (Invitrogen) in the Gateway destination vector (Storms et al., 2005). The *A. niger* glucoamylase promoter is a strong promoter that can direct high levels of expression (Carrez et al., 1990; Fowler et al., 1990; Hata et al., 1992; Jeenes et al., 1993; Santerre Henriksen et al., 1999; Verdoes et al., 1993; Withers et al., 1998). In this study, two integrating plasmids, which were previously constructed in our laboratory using the

above cassettes (Storms et al., 2005), were used. These plasmids are shown in Figure 3 and Figure 4.

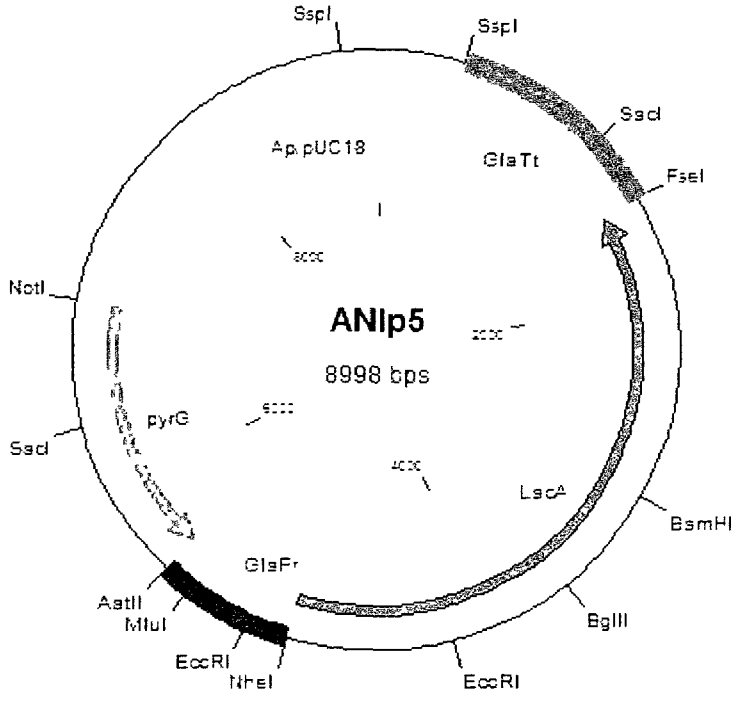


Figure 3: The integrating vector ANIp5. ANIp5 contains the Ap/pUC18 cassette which includes the ampicillin resistance gene and a bacterial origin of replication. The selectable marker cassette contains the *Aspergillus nidulans pyrG* gene that is used as a selectable marker for *A. niger* transformations. The glucoamylase promoter region (GlaPr) and glucoamylase termination region (GlaTt) regulate the transcription and transcription termination of the cloned gene. The reporter gene cassette includes the coding region of the *A. niger lacA* gene. *lacA* is replaced by the target genes during directional cloning using the *NheI* and *FseI* restriction enzyme sites.

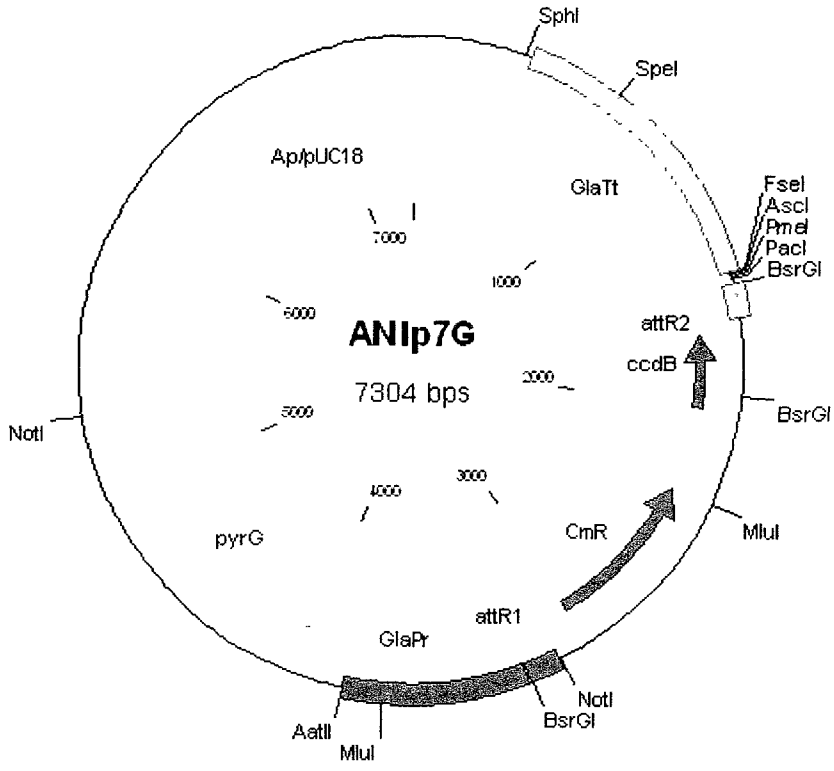


Figure 4: The integrating vector ANIp7G. ANIp7G is an integrating vector. It is used for "Gateway recombination" cloning. ANIp7G was constructed by ligating the attR1 and attR2 Gateway recombination cassette (Invitrogen) into the *NheI* restriction site of ANIp7. Cm^R and ccdB indicate the chloramphenicol resistance gene and ccdB gene.

1.7 Objective

Currently the cost of production of bioethanol from cellulose is much higher than the cost of producing gasoline from fossil fuels. One major problem is the low efficiency of the cellulases used to hydrolyze cellulose to fermentable sugars. Thus my objective was to look for new highly efficient cellulases that could replace the existing cellulases in the commercial cellulase systems or create an altogether new cellulase system. To achieve this objective I: cloned target genes that encoded putative hemicellulases, cellulases and accessory proteins from several fungal species; heterologously expressed them in *A. niger*; screened the expressed enzyme for high endoglucanase activity; and subjected several functionally expressed endoglucanases to biochemical characterization.

2. Material and Methods

2.1 Bioinformatics analysis

I cloned two sets of fungal genes into our expression vectors. The cDNA-derived set included 113 cDNA genes that were part of a collection of about full length 2000 cDNAs identified previously (Semova et al., 2006) and cloned into a Gateway Donor vector, pDONR 201. This set of genes coded for cellulases, hemicellulases and cellulose accessory proteins. Since these genes were already partially characterized, I used the following criteria to select a subset for further analysis: 1) *A. niger* transformed with the gene expressed an active enzyme at significant levels as assessed by enzyme assays of the culture filtrate; 2) and/or *A. niger* transformed with the gene expressed the target protein at a significant level as assessed by SDS-PAGE analysis of the culture filtrate. The genomic DNA-derived set of genes were identified using the protein sequences of 11 previously characterized enzymes, 9 cellulases, one cellobiose dehydrogenase, and one swollenin (Table 2) to search the genomes of 13 fungal species (Table 3) using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) "tbalstn". In this study, I have functionally expressed and biochemically characterized three endoglucanases that were encoded by three genes from the genomic DNA-derived set. These three endoglucanases AfumEgl2010, AnidEgl2020, and FgraEgl2010 were identified by a tblastn search of the *Aspergillus fumigatus*, *Aspergillus nidulans*, and *Fusarium graminearum* genomes using the protein sequence of the *T. reesei* Eg2/Cel5A (Genebank accession #: M19373) as query. I also performed enzyme kinetic and synergy assays on three endoglucanases (ApulSEQ15654, GtraSEQ630 and StheSEQ13822) that were

encoded by three genes from cDNA-derived set. Potential signal peptides were predicted by using the SignalP 3.0 server tool (http://www.cbs.dtu.dk/services/SignalP-3.0/) (Bendtsen et al., 2004). Conserved domains and glycoside hydrolase families were predicted by using the A Conserved Domain Database and Search Service, v2.17 (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al., 2009) at National Center for Biotechnology Information (NCBI). Protein sequences were aligned using ClustalW2 (matrix setup "blosum") from EMBL-EBI (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (Larkin et al., 2007). The molecular mass of the proteins was estimated using the EXPASY Compute pI/Mw tool (http://www.expasy.ch/tools/pi_tool.html) (Bjellqvist et al., 1993). Swiss-model software was used to predict the 3-D structures of proteins (http://swissmodel.expasy.org//SWISS-MODEL.html) (Arnold et al., 2006; Kiefer et al., 2009 ; Peitsch, 1995).

Table 1: The fungal species used as the source of mRNA for cloning the cDNA-derived set

Fungal species name	Source and stock number
Aureoblasidium pullulans	American Type Culture Collection Strain # 62921
Amorphotheca resinae	Canadian Collection of Fungal Cultures Strain # DAOM194228
Coprinus cinereus	Okayama 7
Gloeophyllum trabeum	American Type Culture Collection Strain # 11539
Cryptococcus laurentii	University of Alberta Microfungus collection and Herbarium (UAMH) # 6337
Geomyces pannorum	University of Alberta Devonian Botanic Garden Microfungus Collection & Herbarium Strain # UAMH 1568.
Lentinula edodes	American Type Culture Collection Strain # 48564
Leucosporidium scottii	American Type Culture Collection Strain # 90774
Ophiostoma piliferum	Forintek Culture Collection Strain # FCC 55A
Phanerochaete chrysosporium	USDA Forests Products Laboratory Strain # RP78
Sporotrichum thermophile	American Type Culture Collection Strain # 42464
Thermomyces lanuginosa	American Type Culture Collection Strain # 200065
Aspergillus niger	Wild type strain A733 (Fungal Genomic Stock Group, Kansas City, Kansas
Cunnighamella elegans	American Type Culture Collection Strain #36112
Trametes versicolor	Paprican strain 52P (ATCC 20869)

Table 2: The query genes that were used to identify the 207 target genes in the genomic DNA-derived set

Name	Fungal species	Genebank accession number
Eg1/Cel7B	*Trichoderma reesei*	M15665
Eg2/Cel5A	*Trichoderma reesei*	M19373
Eg3/Cel12A	*Trichoderma reesei*	AB003694
Eg4/Cel61A	*Trichoderma reesei*	Y11113
Eg5/Cel45A	*Trichoderma reesei*	Z33381
Cbh1/Cel7A	*Trichoderma reesei*	P62694
Cbh2/Cel6A	*Trichoderma reesei*	M16190
Bgl1/Cel3	*Aspergillus niger*	AJ132386
Bgl5/ HGT-BG	*Aspergillus oryzae*	Q7Z9L3
CDH/CDH[a]	*Phanerochaete chrysosporium*	CAA61359
Swo1/Swollenin[a]	*Trichoderma reesei*	AJ245918

[a] Cellobiose dehydrogenase and Swollenin act as accessory proteins during cellulose degradation.

Table 3: The fungal species used as the source of genomic DNA templates that were used to clone the genomic DNA-derived set

Fungal species	Source and stock number
Aspergillus fumigatus	Fungal Genetics Stock Center # A1100
Aspergillus nidulans	Fungal Genetics Stock Center # A4
Aspergillus niger	Fungal Genetics Stock Center # A1143
Chaetomium globosum	ATCC 6205
Coprinus cinereus	Fungal Genetics Stock Center # 9003
Fusarium graminearum	Fungal Genetics Stock Center # 9075
Nectria haematococca (Fusarium solani)	Fungal Genetics Stock Center #9596
Neurospora crassa	Fungal Genetics Stock Center # 2489
Phanerochaete chrysosporium	Fungal Genetics Stock Center # 9002
Phytophthora sojae	P. sojae strain P6497 obtained from Dr. Gijzen from University of Western Ontario
Rhizopus oryzae	Fungal Genetics Stock Center # 9543
Ustilago maydis	Fungal Genetics Stock Center # 9021
Phycomyces blaskesleeeanus	Fungal Genetics Stock Center # 10004

2.2 Strains and plasmid

cDNA clones from the 15 fungal species presented in Table 1 were used as the source of the full length ORF templates that were cloned to generate the cDNA-derived set. The 13 fungal species presented in Table 3 were used as the source of genomic DNA to amplify putative genes that encoded cellulases and accessory proteins to generate the genomic DNA-derived set. Heterologous proteins were expressed in *A. niger* strain N593 *Δ glaA::hisG* (Storms et al., 2005) and strain CBS 513.88 (personal communication, Anja Riemens). *E. coli* strain *DH5α* was used as the host strain for recombinant plasmid construction and routine plasmid propagation. *E. coli* strain *Top10* was used as host strain for the propagation of recombinant entry clones and expression clones when using the Gateway cloning system (Invitrogen). *E. coli* strain *DB3.1 (F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm^r) xyl5 Δleu mtl1)* (Bernard and Couturier, 1992; Miki et al., 1992), purchased from Invitrogen, was used as host strain for propagating the Donor vector, pDONR 201, (Invitrogen) and the Gateway Destination vector ANIp7G.

The integrating plasmid ANIp5 (Storms et al., 2005) was used to construct the expression clones by directional cloning. The integrating plasmid ANIp7G, constructed by ligating the attR1 and attR2 Gateway recombination cassette (Invitrogen) into the *NheI* restriction site of ANIp7 plasmid, was used to construct the Gateway expression clones (Storms et al., 2005).

2.3 Media and culture conditions

For directional cloning, *E. coli* cultures of the expression clones were incubated in LB medium containing 100 μg/ml ampicillin for 14-16 hours at 37°C in a shaking incubator at 200 rpm. For cloning using the Gateway in vitro recombination system, *E. coli* cultures of entry clones were incubated in LB medium containing 50 μg/ml kanamycin for 14-16 hours at 37°C incubator shaking at 200 rpm. *E. coli* cultures of expression clones were incubated in LB medium containing 100 μg/ml ampicillin for 14-16 hours in a 37°C shaking at 200 rpm. The LB medium was made of 1% (w/v) bio-tryptone, 0.5% (w/v) yeast extract, and 1.0% (w/v) NaCl. Complete medium (CM) was prepared as described previously (Käfer, 1977). Minimal medium J (MMJ), is the same as minimal medium (MM) (Käfer, 1977) except the amount of nitrate salts, $MgSO_4$, and trace elements were 4 times higher than MM and 15 % (w/v) maltose was used as the carbon source. MMJ was used to grow liquid cultures of *A. niger* harbouring the target endoglucanases. Saline-Tween solution was 0.5% (w/v) NaCl and 0.002% (v/v) Tween-80.

2.4 Cloning

2.4.1 Directional cloning

2.4.1.1 Gene amplification

Full-length cDNA clones isolated from the species listed in Table 1 (Semova et al., 2006) were used as the source of template DNAs for amplifying the cDNA-derived target genes. Genomic DNA isolated from 13 different fungal species (Table 3) was used as the

template for amplifying the genomic DNA-derived target genes. The genomic DNA was isolated using the Qiagen DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol with one minor modification. Harvested mycelia were ground to a fine powder using liquid nitrogen.

The 3' ends of the forward and reverse primer pairs that were used to amplify each target gene harboured 18 to 25 nucleotides complementary to the template strand and the nontemplate strand respectively. The region of the oligo that hybridized to the template strand of each gene began at the A residue of the start codon and extended into the coding region. The region of the oligo that hybridized to the non coding strand began at the third base of the stop codon and extended towards the start codon. The 5' end of the forward primers started with the filler nucleotides AGAAT to increase the digestion efficiency of restriction enzyme *Nhe1*, followed by GCTAGC, the sequence for *NheI* recognition, and then ATA to insert Kozak consensus sequence nucleotides for *A. niger* immediately before the start codon. The 5' end of each reverse primer began with filler nucleotides AGACTA to increase the digestion efficiency of restriction enzyme *Fse1*, followed by the *FseI* recognition sequence, GGCCGGCC immediately adjacent the 18-25 nucleotides that are complementary to the target gene. For target genes that harboured a *NheI* restriction site, *SpeI* or *XbaI* were used to replace the *NheI* restriction site in the forward primers.

PCR amplification was carried out in 50 µl reactions using Pfu buffer with $MgSO_4$. Each gene amplification contained 0.375 units of Pfu polymerase, 0.75 units of Taq polymerase and 100 ng of fungal genomic DNA. The final concentration of dNTPs and primers were 0.2 mM and 0.5 µM. The PCR program used to perform the gene

amplifications was the Touchdown PCR program (Don et al., 1991). The initial cycles in the touchdown PCR program start at an annealing temperature higher than the Tm of the primers and then decrease 0.5°C per cycle until the annealing temperature reaches or is even below the Tm of the primers (Don et al., 1991). To check for the amplification of fragments of the correct size, 5 µl of each PCR product was fractionated on a 1% agarose gel. PCR products of the correct size were gel-purified using the GE Healthcare "GFX^{tm} PCR DNA and Gel Band Purification Kit" following the method provided by the manufacturer. The purified PCR products were double digested with either restriction enzyme *NheI*, *SpeI* or *XbaI*, and *FseI* and then ligated into the prepared backbone of expression vector ANIp5.

2.4.1.2 Backbone preparation

To prepare the backbone for directional cloning, ANIp5 was first digested with *NheI* and *FseI*. Next, the digested product was fractionated on a 1% agarose gel and the vector backbone band of 5562 bp was cut from the agarose gel and extracted using a GE Healthcare "GFX^{tm} PCR DNA and Gel Band Purification Kit".

2.4.1.3 Plasmid construction and *E. coli* transformation

ANIp5 backbone and inserts were mixed and ligated for 2 hours at room temperature. Five µl of each ligation product was mixed with 40 µl of *E. coli DH5α* competent cells, followed by incubation on ice for 30 minutes. The mixture was then subjected to a heat shock at 37°C for 5 minutes. The mixture was added to 100 µl of LB medium and incubated for 45 minutes at 37°C and 250 rpm. Each transformation mix was spread on

an LB agar plate containing 100 μg/ml ampicillin and the plate was incubated overnight at 37°C.

Three putative expression colonies obtained with each insert were inoculated in LB + Amp medium and incubated for 14-16 hours at 37°C with shaking. The plasmid DNAs were extracted using Qiagen Plasmid Miniprep Kits (Qiagen) following the manufacturer's instructions. To determine whether clones with the correct insert were isolated, the expression clones were digested with *SacI* and the digestion products were fractionated on 1% agarose gels. Plasmids harbouring inserts of the correct size were subjected to DNA sequencing to verify that they contained a faithful copy of the desired ORF.

2.4.2 Gateway cloning

2.4.2.1 Gene amplification

Forward and reverse primers that were used to generate attB-PCR products for Gateway cloning were designed following the guidelines in the Gateway Technology manual (Invitrogen). The 5' end of forward primers start with GGGG, then the 25 nucleotide attB1 site ACAAGTTTGTACAAAAAAGCAGGCT, followed by 18 to 25 nucleotides of gene specific sequence. The 5' end of each reverse primer also starts with GGGG, then the 25 nucleotides attB2 site ACCACTTTGTACAAGAAAGCTGGGT, followed by 18 to 25 nucleotides of gene specific sequence. The attB-PCR products were amplified and purified as described in section 2.4.1.1.

2.4.2.2 BP recombination reaction and *E. coli* transformation

The BP recombination reaction was performed with Gateway BP Clonase™ Enzyme Mix (Invitrogen) and the provided protocol with the following modification. In each BP recombination reaction, 2 µl of pDONR 201 vector (30 ng/µl) was mixed with 1 µl 5x BP Clonase™ Reaction buffer, 0.5 µl of the Gateway BP Clonase™ enzyme mix, and 0.5 µl of TE buffer (pH 8.0). One µl of the attB-PCR produced from the gene amplification step was added into the BP master mix. The BP recombination reaction was incubated at 25°C overnight. The BP recombination reaction was terminated by adding 2 µg of Proteinase K. After the Proteinase K was added, the mixture was vortexed briefly and incubated at 37°C for 10 minutes. To transform the desired BP recombination products into competent cells, firstly, 2.5 µl of each BP recombination reaction mixture was mixed with 40 µl *E. coli Top10* chemically prepared competent cells. Secondly, the transformation procedure was performed as described in section 2.4.1.3. Finally, the transformed cells were spread on an LB agar plate containing 50 µg/ml kanamycin. The plate was incubated overnight in a 37°C incubator.

A single colony of each entry clone was inoculated in LB + Kan medium at 37°C and incubated with shaking for 14-16 hours. Plasmid DNA extractions of putative entry clones were done as described in section 2.4.1.3. To confirm the entry clone had the correct insert, the entry clone was digested with *BsrGI*, the digestion mixture was run on 1% agrose gel and the resulting banding pattern for each clone was examined to determine whether the desired fragment had been recombined into the backbone of the Gateway Donor vector.

2.4.2.3 LR recombination reaction and *E. coli* transformation

The LR recombination reaction was performed with Gateway LR Clonase™ Enzyme Mix (Invitrogen) and the manufacturer's protocol with the following modifications. In each LR recombination reaction, 2 µl of ANIp7G vector (30 ng/µl) was mixed with 1 µl 5x LR Clonase™ Reaction buffer, 0.5 µl Gateway LR Clonase™ Enzyme Mix, and 0.5 µl TE buffer (pH 8.0). One µl of the desired entry clone generated in the BP recombination reaction step was added into the LR recombination reaction. The LR recombination reaction was incubated at 25°C overnight. The same procedures were used to terminate and transform the BP and LR recombination reactions except that the transformed *E. coli Top10* chemical treated competent cells were spread on an LB agar plate containing 100 µg/ml ampicillin. A single colony of each expression clone was inoculated in LB + Amp medium and incubated at 37°C with shaking for 14-16 hours. Plasmid DNA extraction and clone confirmation were performed as described in section 2.4.2.2.

2.4.3 DNA Sequencing

The inserts in each expression clone were sequenced at the Genome Quebec Innovation Centre (McGill University). The same universal forward and reverse primers were used for sequencing the ANIp7G and ANIp5 clones (forward primer: 5'-TGCTTCTTCCTTCAGCTTCC-3', reverse primer: 5'-ATGACTTGGCTTCCATTTCAC-3').

2.5 *A. niger* protoplast formation and transformation

A. niger protoplasts used for transformation were prepared as described previously (Debets and Bos, 1986) with some modifications. Recently harvested conidia (less than two weeks old) were used to inoculate a flask of liquid CM medium at a concentration of $2x10^6$ conidia/ml. The mycelia after 16-18 hours growth were harvested by filtration through Miracloth. The mycelial mass was washed with 100 ml 0.6 M $MgSO_4$, then dried and weighed. For every 2 grams of mycelium 10 ml OM solution (1.6 mM NaH_2PO_4, 8.4 mM Na_2HPO_4, and 1.2 M $MgSO_4 \cdot 7H_2O$) and 1 ml of β–glucannase solution (250mg/ml) were added to the mycelium to digest the cell wall. After 2.5 to 3 hours incubation in a 30°C shaking incubator at 100 rpm, the protoplast suspension was transferred into a 50 ml centrifuge tube, and an equal volume of ice-cold TB-layer (0.6 M sorbitol and 0.1 M Tris-HCl pH 7.5) was added carefully and slowly onto the protoplast suspension to avoid mixing of the two layers. After centrifugation, the protoplasts were harvested from the interface between the TB layer and OM layer. The harvested protoplast solution was centrifuged and washed three times with ice cold S/C solution (1 M sorbitol and 50 mM $CaCl_2$). Following the last wash the protoplasts were resuspended in ice cold S/C solution at a final concentration of $1x10^7$ protoplasts/ml.

Transformation was done as described previously (Wernars et al., 1987) with several modifications. Two hundred µl of protoplast suspension was mixed with 20 µl 0.4 M ATA (aurintricarboxylic acid), 1 to 5 µg of plasmid DNA, and a 100 µl solution of 20% (w/v) polyethyleneglycol (PEG) 4000, 16.7 mM $CaCl_2$, 3.3 mM Tris-HCl pH 7.5, and incubated at room temperature for 10 minutes. After the 10 minute incubation, 1.5 ml PEG 4000 solution (60% (w/v) PEG 4000 , 50 mM $CaCl_2$, 10 mM Tris-HCl pH 7.5) was

added, mixed, and incubated at room temperature for another 20 minutes. Five ml of 1.2 M sorbitol was added and the solution was mixed carefully. The mixture was centrifuged at room temperature for 10 minutes at 4000 rpm, and the pellet was then resuspended in 1 ml 1.2 M sorbitol. The transformed protoplasts were finally plated on MM agar plates with 4.47% (w/v) KCl using MM with 4.47% (w/v) KCl and 1.5% agar overlayer for selection (Storms et al., 2005). Plates were incubated for 4 to 6 days in a 30°C incubator until individual transformants appeared.

2.6 Microtiter plate culture growth and initial screening

Recombinant versions of expression vectors ANIp5 and ANIp7G with the cloned cellulase and accessory protein genes were introduced into *A. niger* strain N593 *Δ glaA::hisG* using the protoplast generation and transformation method described above.

Microtiter plate cultures were inoculated as described previously (Master et al., 2008). A toothpick soaked in saline-tween was used to harvest conidia from the surface of a single colony on transformation plates (MM + 4.47% (w/v) KCl). Conidia on the tip of the toothpick were used to inoculate individual 200 µl MMJ medium cultures in the individual wells of a microtiter plate. The microtiter plate cultures were grown for 4 to 6 days inside a styrofoam or plastic box lined on the bottom with paper towels soaked in water and incubated at 30°C without shaking. The medium from each culture was harvested and transferred into 96 well PCR plates using a multi-channel pipettor.

Initial screening for endoglucanase activity was performed using agar plates containing carboxylmethylcellulose (CMC) sodium salt and activity staining with Congo red essentially as described previously (Wood et al., 1988). To prepare the CMC agar

plate, 0.5% (w/v) CMC (degree of substitution: 0.65-0.90) sodium salt (Sigma) and 1.5% (w/v) agar (Bioshop) were dissolved in 100 mM citrate buffer pH 5.0 and poured into NUNC OmniTray single-well plates (VWR). Three µl of culture medium suitably diluted with 10 mM citrate buffer pH 5.0 was then spotted in duplicate on two CMC agar plates. The plates were incubated upside down at 37°C overnight. Both CMC agar plates were stained with 0.05% (w/v) Congo red solution for 20 to 30 minutes and then destained with 0.5% (w/v) NaCl for 20 to 30 minutes. The destaining step was repeated until clear spots were visible.

2.7 Petri dish culture growth and purification of enzymes

Petri dish cultures were inoculated as described previously (Master et al., 2008) with the following modification. 15 ml of MMJ medium was inoculated using conidia resuspended in saline-Tween to a final concentration of $1x10^6$ conidia/ml. The growth conditions were the same as those used for the microtiter plate cultures described in section 2.6.

Culture supernatant was centrifuged for 15 minutes at 3000 rpm and 4°C to pellet the mycelia. The supernatant was then filtered through Ultrafree-0.5 centrifugal filter devices (Millipore) to remove small molecules including sugars, following the manufacturer's instructions. A maximum of 500 µl of culture supernatant was loaded into the Ultrafree-0.5 centrifugal filter device. The supernatant was centrifuged for 30 minutes at 10,000 rpm and 4°C to bring the volume of supernatant from 500 µl down to 20 µl. The 20 µl of concentrated culture supernatant was washed five times by adding 500 µl of citrate buffer (10 mM, pH 5.0) to the filter device and centrifuging for 30 minutes at 10,000 rpm and

4°C. During each washing step, if the remaining volume of solution after centrifugation was greater than 20 µl, an extra centrifugation step was done to bring the final volume of solution down to 20 µl. After five washes, the solution that remained on the filter was transferred into a microfuge tube. The Ultrafree-0.5 centrifugal filter device was then washed with 480 µl of citrate buffer (10 mM, pH 5.0) and the retained liquid was combined with the previously transferred solution. Purified enzymes were stored at -80°C.

2.8 Three cDNA-derived target genes

The ApulSEQ15654, StheSEQ13822, and GtraSEQ630 genes were amplified from cDNA and cloned into the pGBFIN-GTW vector (Tsang et al., 2008). The recombinant versions of expression vector pGBFIN-GTW were introduced into *A. niger* strain CBS513.88 and the culture filtrate purified to remove sugar using an Ultrafree-0.5 centrifugal filter device, as described above.

2.9 SDS-PAGE and enzyme concentration determination

SDS-PAGE analysis was performed using a 12%, 10 well precast polyacrylamide gel (DGel Science). Proteins were visualized by staining with Coomassie blue R250 (Biorad).

In order to estimate the protein concentration, different amounts (0.25 µg, 0.5 µg, 0.75 µg, and 1 µg) of BSA (Albumin, bovine serum fraction V, ~99% purity, protease free, essentially r-globulin free, Sigma) were loaded adjacent to the suitably diluted endoglucanases on the same SDS-PAGE gel. After the proteins were stained with Coomassie blue R250, an image of the gel was taken using the Gene Genius Bio Imaging

System (Syngene). To estimate the concentration of each endoglucanase, Gene Tools (Version 3.08.03) was used.

2.10 Enzyme assays

2.10.1 Enzyme activity assay

Endoglucanase activity was determined by measuring the production of reducing sugar ends. The assay was performed in 133 mM citrate buffer (pH 5.0) at 37°C using filter paper disks as substrate (Xiao et al., 2004). The amount of reducing sugar ends produced were quantitated using the bicinchoninic acid (BCA) assay adapted for 96-well microtiter plates as previously described (Doner and Irwin, 1992 ; Grishutin et al., 2004; Zorov et al., 1997). Ten µl of enzyme reaction mixture was added to the wells of a 96-well PCR plate containing 90 µl of H_2O and 100 µl of BCA mixture (Doner and Irwin, 1992 ; Grishutin et al., 2004; Zorov et al., 1997). The BCA mixture contains equal volumes of reagent A [5.43% (w/v) of Na_2CO_3, 2.42% (w/v) of $NaHCO_3$, and 0.1942% (w/v) 2'2 bicinchoninic disodium BCA] and reagent B [0.125% (w/v) $CuSO_4 \cdot 5H_2O$, 0.126% (w/v) L-serine]. The assay mixtures were incubated for 40 minutes at 80°C in a heating block. Once the incubation was finished, the plate was transferred into an ice box and then centrifuged for 3 minutes at 3000 rpm. Eighty µl of the BCA reaction mixture was removed and the absorbance was determined at 562 nm using PowerWave™ HT microplate scanning spectrophotometer (Bio-Tek). In order to convert the absorbance into the amount of reducing sugar ends, a standard curve was generated using D-glucose (5 µM to 25 µM, final concentration of D-glucose in BCA assay). One unit of enzyme

activity is defined as 1 μmole of reducing sugar ends released per minute at 37°C. All assays were done in triplicate.

2.10.2 Enzyme pH profiles

To determine the pH profile of each endoglucanase, the microplate filter paper assay was performed as previously described (Xiao et al., 2004) with the following modifications. Glycine-HCl buffer (pH 2.0 to 3.0), citrate buffer (pH 3.0 to 6.0), and sodium phosphate buffer (pH 6.0 to 8.0) were used to generate the pH range. Twenty μl of suitably diluted enzyme (diluted in 10 mM citrate buffer, pH 5.0) was mixed with 40 μl of 200 mM buffer at the desired pH and a filter paper disk in the wells of a 96-well PCR plate. The reactions were carried out for 60 minutes at the optimal temperature of each endoglucanase (in the case of AnidEgl2020 it was at 60°C, AfumEgl2010 at 70 °C, and FgraEgl2010 at 40 °C). To terminate the reaction the 96-well PCR plate was transferred on ice, and 10 μl of each reaction mixture was removed and a BCA assay was performed as described in section 2.10.1.

2.10.3 Enzyme temperature profiles

The microplate filter paper assay (Xiao et al., 2004) was also used to determine the temperature profile of each endoglucanase with the following modifications. Twenty μl of suitably diluted enzyme was mixed with 40 μl of 200 mM citrate buffer (pH 5.0) and a filter paper disk in the wells of a 96-well PCR plate. The enzyme assay was incubated for 60 minutes at the following temperatures: 25°C, 30 °C, 37 °C, 50 °C, 60 °C, 70 °C, and

80°C. The reactions were terminated by placing the PCR plate on ice and the BCA assay was performed as described in section 2.10.1.

2.10.4 Enzyme thermostability profiles

To determine each enzyme's thermostability, a microplate filter paper assay was performed as previously described (Xiao et al., 2004) with the following modifications. Twenty µl of suitably diluted enzyme was incubated with 40 µl of 200 mM citrate buffer (pH 5.0) for 30 minutes without the addition of a filter paper disk at the following temperatures: 25°C, 30 °C, 37 °C, 50 °C, 60 °C, 70 °C, and 80°C. The enzyme buffer mixture was cooled down on ice and transferred into another 96-well PCR plate containing the filter paper disk. The plate was incubated for 60 minutes at 37°C. Finally, the amount of reducing sugar ends produced were determined using the BCA assay performed as described in section 2.10.1.

2.11 Enzyme kinetics

Enzyme kinetic parameters were estimated at 37°C and the optimal pH of each endoglucanase. Parallel assays were also performed with a commercial endoglucanase, the *T. reesei* Eg2/Cel5A (Iogen Corp.). Phosphoric acid swollen cellulose (PASC) was used as the substrate. PASC was prepared as described previously (Wood, 1988) with the following modifications. 5 grams of minicrystalline cellulose PH-105 Avicel (Sigma) was resuspended in 150 ml of 85% (w/v) phosphoric acid. The suspension was slowly stirred for 3 hours at 4°C. After adding two liters of ice cold distilled H_2O, the suspension was stirred vigorously for 15 minutes. The suspension was centrifuged for 10 minutes at

14,000 g and 4°C. The precipitated cellulose was washed several times with ice cold distilled H_2O until the pH was between 5.0 and 7.0. The pellet formed during the final centrifugation was resuspended in 10 mM citrate buffer pH 5.0. Finally, a blender was used to generate a homogeneous mixture. The final concentration of PASC used in the enzyme assays ranged from 0.05% (w/v) to 1% (w/v). Each enzyme assay consists of 40 µl of each endoglucanase at a known concentration (Table 4), 60 µl of 3.3 X buffer at each endoglucanase's optimal pH (Table 5) and 100 µl of different concentrations of PASC in a 96-well flat bottom microplate. Enzyme kinetic parameters were determined using enzyme and substrate concentrations that resulted in a linear increase of reducing sugar ends with time. Enzyme hydrolysis was carried for 60 minutes at 37°C with shaking at 250 rpm. The reaction mixtures were mixed using a multi-channel pipettor until the solution was homogeneous and then 10 µl aliquots were transferred at intervals of 5 minutes over a 60 minute time course into a new 96-well PCR plate. The amount of reducing sugar ends produced was determined using the BCA assay as described in section 2.10.1. The values of µmoles of reducing sugar ends produced per minute per mg of enzyme with variable concentration of PASC were used to determine the kinetic parameters of each endoglucanase using the GraFit v6.0 software program (Erithacus Software Limited), by fitting to the Michaelis-Menten equation.

Table 4: The final concentration of target enzymes used in the kinetic assays

Target name	Amount of enzyme (pmole)
GtraSEQ630	5.3
ApulSEQ15654	3.0
StheSEQ13822	4.5
FgraEgl2010	2.9
AfumEgl2010	4.5
AnidEgl2020	31.5
Eg2/Cel5A (Iogen)[a]	7.2

[a] *T. reesei* Eg2/Cel5A obtained from Iogen Corporation.

Table 5: The final concentration and pH of the buffers used for determining enzyme kinetic parameters and synergism

Name of enzyme	Buffer
Eg2/Cel5A (Iogen)[a]	60 mM Citrate buffer pH 5.0
AnidEgl2020	60 mM Citrate buffer pH 5.0
AfumEgl2010	60 mM Citrate buffer pH 5.0
FgraEgl2010	60 mM Citrate buffer pH 5.0
ApulSEQ15654	44 mM McIlvaine buffer pH 4.5
StheSEQ13822	49 mM McIlvaine buffer pH 6.0
GtraSEQ630	44 mM McIlvaine buffer pH 4.5

[a] *T. reesei* Eg2/Cel5A obtained from Iogen Corporation.

2.12 Synergy assay

The substrates used for the synergy assays were 1% (w/v) Avicel (Sigma) and 1% (w/v) PASC. PASC was prepared as described in section 2.11. 1% (w/v) Avicel was prepared by suspending 1 gram of Avicel in 100 ml of 10 mM citrate buffer (pH 5.0). In order to assess synergism of enzyme mixtures, five different combinations (Table 6) of each endoglucanase and the *T. reesei* exoglucanase (Cbh1/Cel7A) (Iogen Corp.) were mixed in a final volume of 20 µl. The 20 µl enzyme mixtures were added to 30 µl of 3.3 X optimal pH buffer (Table 5) and 50 µl of 1% (w/v) PASC or 1% (w/v) Avicel in the well of a 96-well flat bottom microplate. The activity of each enzyme was determined as follows. Four different concentrations of each endoglucanase and the *T. reesei* exoglucanase (Cbh1/Cel7A) (Iogen Corp.) (Table 7) were assayed by combining 20 µl of appropriately diluted enzyme, 30 µl of 3.3 X optimal pH buffer (Table 5) and 50 µl of 1% (w/v) PASC or 1% (w/v) Avicel (personal communication, Rasha Cheikh-Ibrahim). The enzyme assays were carried out for 120 minutes in a shaking incubator at 250 rpm and 37°C. The BCA assay (as described in section 2.10.1) was performed to follow the production of reducing sugar ends.

Table 6: The final concentrations of the *T. reesei* exoglucanase (Cbh1/Cel7A) and endoglucanases that were present in the 20 µl of enzyme mixtures that were added to buffer and Avicel or PASC for the synergy assays

Number of combination	Amount of enzyme (pmole)			
	Avicel as substrate		PASC as substrate	
	Exoglucanase (Cbh1/Cel7A)[a]	Endoglucanase	Exoglucanase (Cbh1/Cel7A)	Endoglucanase
1	0	1.81	0	0.91
2	10.2	1.36	3.4	0.68
3	20.4	0.91	6.8	0.45
4	30.6	0.45	10.2	0.23
5	40.8	0	13.6	0

[a] *T. reesei* Cbh1/Cel7A obtained from Iogen Corporation.

Table 7: The final concentration of exoglucanase and endoglucanase enzymes that were present in the 20 µl samples used to perform assays with various concentrations of the individual enzyme

Amount of enzyme (pmole)			
Avicel as substrate		PASC as substrate	
Exoglucanase (Cbh1/Cel7A)[a]	Endoglucanases	Exoglucanase (Cbh1/Cel7A)	Endoglucanases
10.2	0.45	3.4	0.23
20.4	0.91	6.8	0.45
30.6	1.36	10.2	0.68
40.8	1.81	13.6	0.91

[a] *T. reesei* Cbh1/Cel7A obtained from Iogen Corporation.

3. Results

3.1 Cloning and plasmid construction

The 113 target genes in the cDNA-derived set that were selected from over 2,000 cDNA-derived fungal Gateway entry clones (Semova et al., 2006) were transferred into the integrating plasmid ANIp7G in order to test whether they could be expressed in *A. niger* strain N593 *ΔglaA::hisG*. The Gateway LR recombination reaction (described in section 2.4.2.3) was used to move the target gene from the entry clone into the expression vector ANIp7G. To verify the correct target gene was cloned into the expression vector, the restriction enzyme *BsrGI* was used to digest each putative expression clone. The digestion results showed that 70 out of the113 target genes were successfully cloned into the integrating plasmid. Table 8 lists the 113 target genes in the cDNA-derived set.

Table 8: The 113 target genes of the cDNA-derived set

Target gene	Cloned into ANIp7G	Target gene	Cloned into ANIp7G
AresSEQ4824glaA0	Yes	StheSEQ1892	No
GtraSEQ1243chi	Yes	CcinSEQ19341fun	No
GtraSEQ3958man	No	ApulSEQ15671egl	No
GtraSEQ4830xyn	No	GtraSEQ1990prt	No
StheSEQ2292cbhE0	No	TverSEQ8985axe	No
StheSEQ4494eglC0	Yes	GtraSEQ10347osp (tlp)	Yes
StheSEQ457cbhC0	Yes	StheSEQ4870fun	Yes
StheSEQ648cbhD0	No	OpilSEQ9021	Yes
GtraSEQ9704abn	No	StheSEQ15225osp	Yes
CcinSEQ4423eglF0	Yes	LscoSEQ14813	Yes
CcinSEQ11398CbhB0	Yes	AresSEQ3057lcc	Yes
StheSEQ437cbhA0	Yes	StheSEQ12575eng	Yes
StheSEQ643amyA0	No	StheSEQ13822egl	Yes
GtraSEQ4064RNase	Yes	TlanSEQ9064eng	Yes
AresSEQ8410osp	No	LedoSEQ3170fun	No
OpilSEQ11503xyn	Yes	GtraSEQ8518egl	Yes
ApulSEQ18636lam	Yes	GtraSEQ3053fae	Yes
GpanSEQ12534	No	StheSEQ15153fun	No
StheSEQ15248	Yes	StheSEQ64fun	Yes
CcinSEQ16583xyn	Yes	ApulOPL1293	Yes
TverCcn3344	Yes	ApulSEQ12207	Yes

AresSEQ12121bgl	Yes	ApulSEQ17044	Yes
TverSEQ9157	No	ApulSEQ17217	No
PchrSEQ2380fun	Yes	ApulSEQ2789	Yes
TverCcn8594pga	No	ApulSEQ3139	Yes
StheSEQ3268	Yes	ApulSEQ60	Yes
StheSEQ11380	Yes	ApulSEQ6709	Yes
ApulSEQ15654egl	No	ApulSEQ8695	No
GpanSEQ16040pga	Yes	GtraSEQ5285xynA0	No
LedoSEQ12272egl	Yes	CcinSEQ14118	Yes
ClauSEQ1118osp	Yes	CcinSEQ15107	No
GpanSEQ12029pgx	Yes	CcinSEQ15948	Yes
CcinSEQ14390egl	Yes	CcinSEQ18776	Yes
TverSEQ4796egl	Yes	CcinSEQ18940	No
GtraSEQ7011xyn	Yes	CcinSEQ5116EglB0	Yes
StheSEQ3090xyn	No	CcinSEQ6020	Yes
ApulSEQ17085exg	Yes	CcinSEQ9568	No
CcinSEQ14540fun	Yes	CcinSEQ9885	Yes
GtraSEQ13688lam	Yes	GpanSEQ11398	No
CcinSEQ19498fae	No	GpanSEQ6135	Yes
TverSEQ8543chi	Yes	GpanSEQ6729	No
LscoSEQ12576pga	No	GtraSEQ11412	No
LedoSEQ3581bgl	Yes	GtraSEQ5873OspH0	Yes
CcinSEQ14918osp	Yes	LedoSEQ11599	Yes

ApulSEQ5780fae	Yes	LedoSEQ12465	No
CcinSEQ14835egl	Yes	LedoSEQ3005	No
AresSEQ4925osp	Yes	LedoSEQ7334	No
AresSEQ10811bla	Yes	LscoSEQ1311	Yes
StheSEQ14955bgl	Yes	TlanSEQ10520	Yes
GtraSEQ8352eglB0	Yes	TverSEQ7178	Yes
GpanSEQ1806bgl	Yes	GpanSEQ2075lip	No
PchrTLN5267amyA0	No	CcinSEQ4635prt	No
StheSEQ3724XynB0	Yes	AresSEQ11525lip	No
StheSEQ2730	No	Pchr4630ospGo	No
GtraSEQ2907funF0	Yes	TverSEQ436mnp	No
PchrSEQ4420chi	No	CcinSEQ9181apr	No
TverCcn5797	No		

I also cloned a set of 207 genomic DNA-derived genes. The genomic DNA-derived set was PCR amplified from genomic DNA prepared as described in section 2.4.1.1. Directional cloning was used to clone 142 of the 207 genomic DNA-derived genes, whereas Gateway recombination was used to clone 65 of the 207 genomic DNA-derived genes. The amplified target genes were cloned into either integrating plasmid ANIp5 or ANIp7G using directional cloning and Gateway recombination, respectively (as described in sections 2.4.1 and 2.4.2). For expression clones generated using the directional cloning method, *SacI* restriction enzyme digests were used to confirm the target gene had been cloned. Restriction enzyme *BsrGI* was used to verify that the Gateway cloning was successful. Based on the restriction enzyme digestion results, 149 out of 207 target genes were successfully cloned into the integrating plasmids. Table 9 shows the list of 149 cloned genes, and Figure 5 shows the results of 1% agarose gel elecrophoretic analysis of the restriction digests.

Table 9A: List of successfully cloned genomic DNA-derived genes

Name [a]	Position on gel [b]	Digestion (*SacI*) [c]
AnigEgl4020	Row#1, lane1	3773, 2155, 846
AfumEgl4040	Row#1, lane2	3773, 2845
NcraEgl4060	Row#1, lane4	3773, 2310, 746
NcraEgl4100	Row#1, lane5	3773, 2585
AfumEgl4020	Row#1, lane6	3773, 2214, 759
AfumEgl4010	Row#1, lane7	3773, 2218,713
AfumEgl4030	Row#1, lane8	3773, 2657
FgraEgl4060	Row#1, lane9	3773,1892, 632
NhaeEgl4040	Row#1, lane10	3773, 2492, 437
PchrEgl4010	Row#1, lane11	3773, 2595, 711
FgraEgl3020	Row#2, lane1	3773, 2598
FgraEgl3010	Row#2, lane2	3773, 1990, 619
AnidEgl3010	Row#2, lane3	3773, 1747, 690, 133
AnigBgl5010	Row#2, lane4	3773, 1613, 1370, 180
AnigBgl5020	Row#2, lane5	3773, 2309, 857
FgraBgl5010	Row#2, lane6	3773, 3959
UmayBgl5060	Row#2, lane7	3773, 1974, 1685, 749, 229, 75
UmayBgl5010	Row#2, lane8	3773, 3292
FgraEgl4030	Row#2, lane9	3773, 1604, 1196
AfumBgl5010	Row#2, lane10	3773, 2136, 1021
AfumBgl5030	Row#2, lane11	3773, 3042
RoryBgl5010	Row#2, lane12	3773, 3209

[a] Gene that was successfully cloned into ANIp5 based on restriction enzyme analysis.
[b] Location of the lanes in Figure 5A harbouring the positive clones.
[c] *SacI* fragments obtained and expected for positive clones.

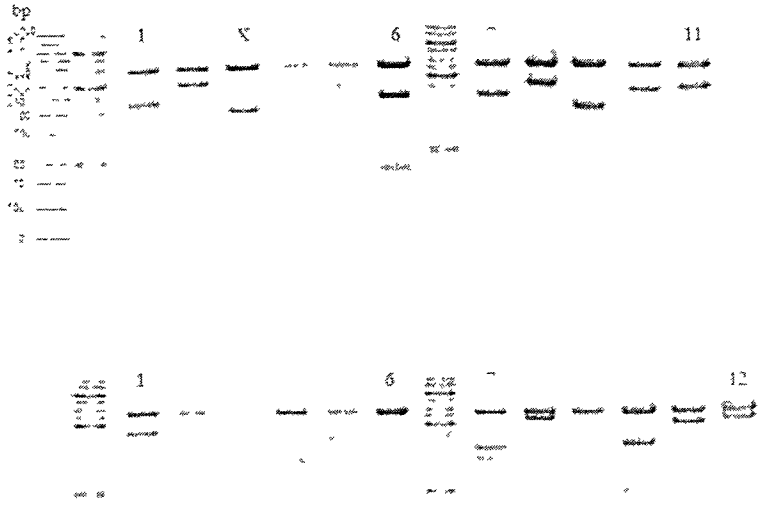


Figure 5A: *SacI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9A. The lane marked X is a clone that did not harbour the expected.

Table 9B: List of successfully cloned genomic DNA-derived genes [a]

Name	Position on gel	Digestion (*SacI*)
FgraEgl1020	Row#1, lane1	3773, 3146
AfumCbh1020	Row#1, lane2	3773, 1741, 1536
AfumCbh1010	Row#1, lane3	3773, 3394
PchrCbh1060	Row#1, lane5	3773, 1926, 1317
AnigEgl1010	Row#1, lane7	3773, 4596
CgloCbh1040	Row#1, lane8	3773, 2674, 857
FgraEgl2020	Row#1, lane9	3773, 2989
AfumEgl2030	Row#1,lane10	3773, 2574, 454
FgraEgl2010	Row#1,lane11	3773, 2896
NhaeCbh2010	Row#1,lane12	3773, 3428
NcraEgl4050	Row#2, lane1	3773, 3018
RoryBgl5040	Row#2, lane2	3773, 2920
NcraEgl4030	Row#2, lane3	3773, 2722
AfumCbh2010	Row#2, lane4	3773, 3508
AnidCbh2020	Row#2, lane 5	3773, 3236
AnigCbh2020	Row#2, lane 6	3773, 1735, 1546
FgraCbh2010	Row#2, lane 7	3773, 2369, 985, 24
NcraEgl5010	Row#2, lane 9	3773, 2149, 581
AfumEgl5010	Row#2, lane 10	3773, 2577
AfumSwo1010	Row#2, lane 11	3773, 3526

[a] Columns designated as indicated for Table 9A.

Figure 5B: *SacI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9B. The lane marked X is a clone that did not harbour the expected.

Table 9C: List of successfully cloned genomic DNA-derived genes

Name [a]	Position on gel [b]	Digestion (*BsrGI*) [c]
AnidEgl5010	Row#1, lane1	5619, 1095
AnidCbh1010	Row#1, lane2	5619, 1574, 37
AnidEgl4060	Row#1, lane3	5619, 1104
AnidEgl4040	Row#1, lane4	5619, 960, 670
AnidEgl4010	Row#1, lane5	5619, 796, 499
CcinEgl4010	Row#1, lane6	5619, 902, 451
CcinEgl4060	Row#1, lane7	5619, 1790
CcinEgl3010	Row#1, lane8	5619, 1250, 1160
CcinEgl4080	Row#1, lane9	5619, 1583
CcinEgl2010	Row#1, lane10	5619, 1887
CcinCbh2010	Row#1, lane11	5619, 1501
NcraCbh1020	Row#1, lane12	5619, 1594
NcraCbh2010	Row#2, lane1	5619, 1546
CcinCbh1020	Row#2, lane2	5619, 1470
CcinCbh1040	Row#2, lane3	5619, 705
NcraEgl2010	Row#2, lane4	5619, 1271
AfumBgl5020	Row#2, lane5	5619, 2648
AnidBgl5030	Row#2, lane6	5619, 1422
CcinBgl5130	Row#2, lane7	5619, 1997
CcinBgl5010	Row#2, lane8	5619, 838, 822
CcinBgl5100	Row#2, lane9	5619, 1728, 932
CcinBgl5090	Row#2, lane10	5619, 1965

[a] Genes that was successfully cloned into ANIp7G based on restriction enzyme analysis.
[b] Location of the lanes in Figure 5C harbouring the positive clones.
[c] *BsrGI* fragments obtained and expected for positive clones.

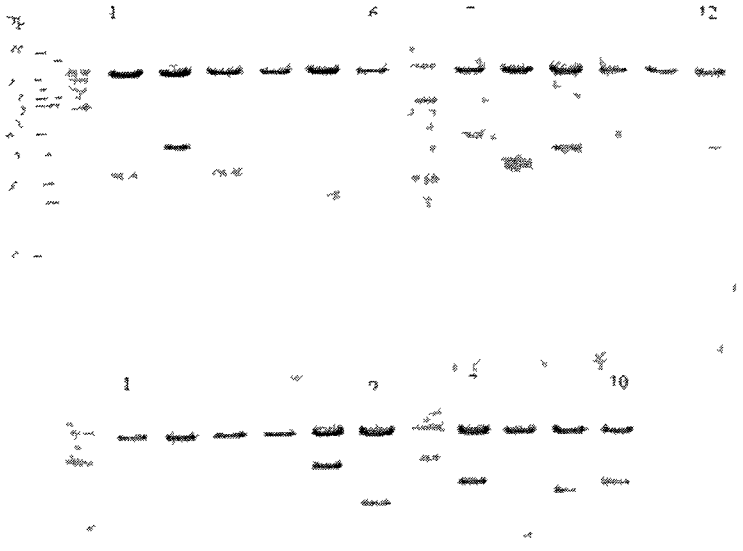


Figure 5C: *BsrGI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9C.

Table 9D: List of successfully cloned genomic DNA-derived genes [a]

Name	Position on gel	Digestion (*SacI*)
AfumCdh020	Row#1, lane2	3773, 1492, 1402, 881, 558
FgraCdh030	Row#1, lane3	3773, 4207
NhaeCdh010	Row#1, lane4	3773, 4244
NhaeCdh030	Row#1, lane5	3773, 3335, 866
AnigCbh2010	Row#1, lane6	3773, 1990, 1687
CcinEgl4050	Row#1, lane7	3773, 2198, 782
FgraEgl1010	Row#1, lane8	3773, 3441
CgloEgl2040	Row#1, lane9	3773, 1994, 996
CgloEgl2030	Row#1, lane10	3773, 3122
CgloCbh1030	Row#1, lane12	3773, 2609
FgraBgl5020	Row#2, lane2	3773, 3190
NhaeCbh1020	Row#2, lane3	3773, 3423
FgraCdh050	Row#2, lane4	3773, 3451
AnigCdh020	Row#2, lane5	3773, 3515
PchrEgl3010	Row#2, lane6	3773, 2678
NhaeEgl3010	Row#2, lane7	3773, 1940, 587
AfumEgl1020	Row#2, lane8	3773, 1544, 779, 696

[a] Columns designated as indicated for Table 9A.

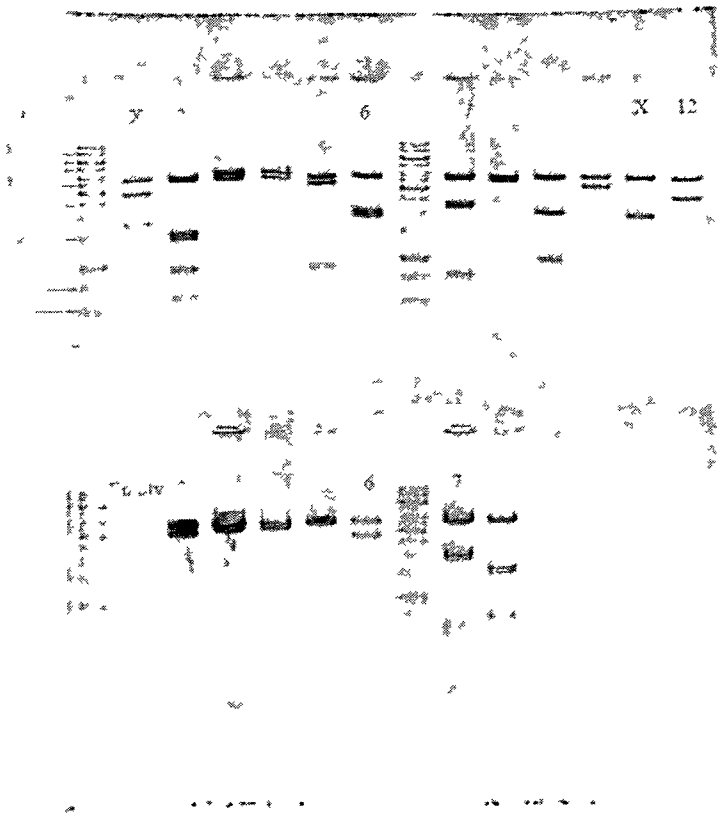


Figure 5D: *SacI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9D. The lane marked X is a clone that did not harbour the expected

Table 9E: List of successfully cloned genomic DNA-derived genes [a]

Name	Position on gel	Digestion (*SacI*)
FgraEgl4050	Row#1, lane1	3773, 1660, 1504
FgraEgl4040	Row#1, lane2	3773, 2875
FgraEgl4010	Row#1, lane3	3773, 2620
FgraEgl4020	Row#1, lane4	3773, 3033
NhaeEgl1010	Row#1, lane5	3773, 2140, 1046
NhaeEgl2020	Row#1, lane6	3773, 2390, 497
NhaeEgl2010	Row#1, lane7	3773, 3342
NhaeEgl2030	Row#1, lane8	3773, 3002
NhaeEgl4020	Row#1, lane9	3773, 3084
NhaeEgl4060	Row#1, lane10	3773, 2611
NhaeEgl4030	Row#1, lane11	3773, 2013, 517
NhaeEgl4010	Row#1, lane12	3773, 2623
NhaeBgl5010	Row#2, lane1	3773, 2760, 1138
PchrCbh1020	Row#2, lane2	3773, 2898, 551
PsojCbh1050	Row#2, lane3	3773, 3008
PsojEgl3090	Row#2, lane4	3773, 1828, 617
PsojEgl3060	Row#2, lane5	3773, 2524
PsojEgl3070	Row#2, lane6	3773, 2057, 476
PsojEgl3011	Row#2, lane7	3773, 1733, 821
PsojEgl3040	Row#2, lane8	3773, 1828, 690
UmayBgl5030	Row#2, lane9	3773, 3385
UmayBgl5020	Row#2, lane10	3773, 3313

[a] Columns designated as indicated for Table 9A.

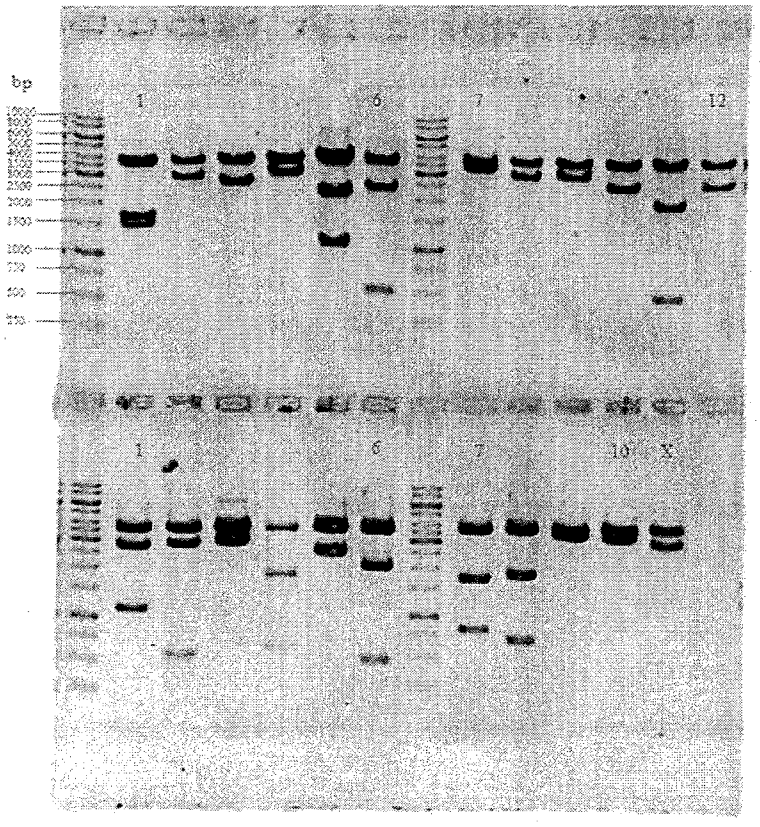


Figure 5E: *SacI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9E. The lane marked X is a clone that did not harbour the expected.

Table 9F: List of successfully cloned genomic DNA-derived genes

Name [a]	Position on gel [b]	Digestion (*SacI*) [c]
PsojEgl3010	Row#1, lane1	3773, 2246, 537, 164
AfumEgl4050	Row#1, lane3	3773, 1703, 1085
AfumEgl2010	Row#1, lane4	3773, 2364, 828
AnidEgl1010	Row#1, lane5	3773, 3088
AnidEgl2020	Row#1, lane6	3773, 3074
NcraEgl4020	Row#1, lane7	3773, 2821
NcraEgl4090	Row#1, lane8	3773, 2962
Name [d]	**Position on gel**	**Digestion (*BsrGI*) [e]**
AfumBgl1010	Row#1, lane9	5619, 3090
AnidBgl1010	Row#1, lane10	5619, 1852, 1338
AnigBgl1010	Row#1, lane11	5619, 2163, 815
CcinBgl1010	Row#1, lane12	5619, 1983, 1254
FgraBgl1010	Row#2, lane1	5619, 1789, 1073, 302
NcraBgl1010	Row#2, lane2	5619, 2731, 302
PblaBgl1020	Row#2, lane3	5619, 2947
PblaBgl1010	Row#2, lane4	5619, 1159, 746, 736, 383, 54
PchrBgl1010	Row#2, lane5	5619, 2840, 245
RoryBgl1010	Row#2, lane6	5619, 1314, 1103
RoryBgl1030	Row#2, lane7	5619, 1651, 782
RoryBgl1040	Row#2, lane8	5619, 1644, 932
UmayBgl1020	Row#2, lane9	5619, 2490

AnidEgl4050	Row#2, lane10	5619, 800, 25
CcinEgl4030	Row#2, lane11	5619, 1288

[a] Gene that was successfully cloned into ANIp5 based on restriction enzyme analysis.
[b] Location of the lane in Figure 5F harbouring the positive clones.
[c] *SacI* fragments obtained and expected for positive clones.
[d] Gene that was successfully cloned into ANIp7G based on restriction enzyme analysis.
[e] *BsrGI* fragments obtained and expected for positive clones.

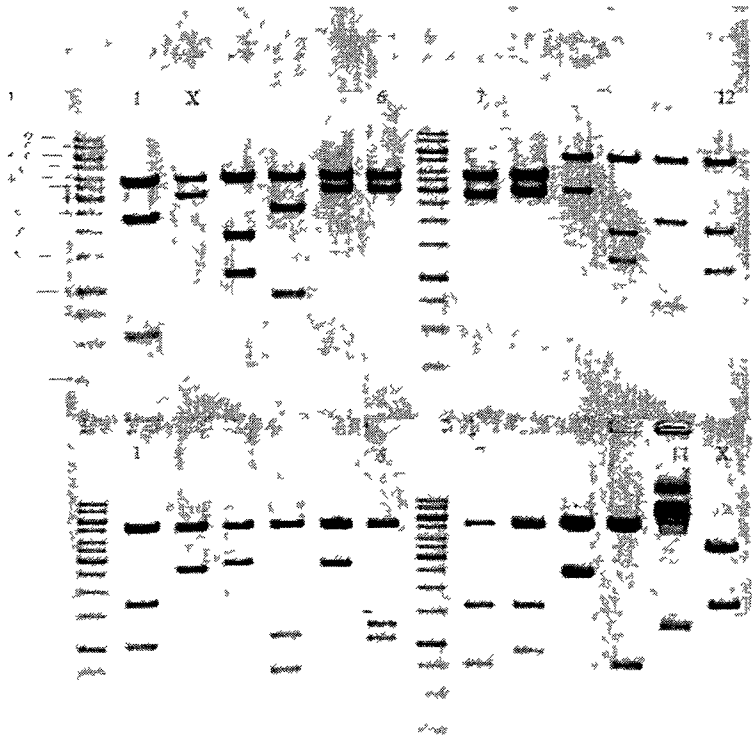


Figure 5F: *SacI* or *BsrGI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9F. The lane marked X is a clone that did not harbour the expected

Table 9G: List of successfully cloned genomic DNA-derived genes [a]

Name	Position on gel	Digestion (*SacI*)
FgraCdh020	Lane 1	3773, 3329, 896
NhaeCbh1010	Lane 2	3773, 3823
NcraEgl4070	Lane 3	5619, 1315 (***BsrGI***)
AnigEgl3010	Lane 4	3773, 1867, 761
AfumCbh1030	Lane 5	3773, 1892, 1286
UmayBgl5050	Lane 6	3773, 3652
NcraEgl1010	Lane 7	3773, 1822, 1362

[a] Columns designated as indicated for Table 9A.

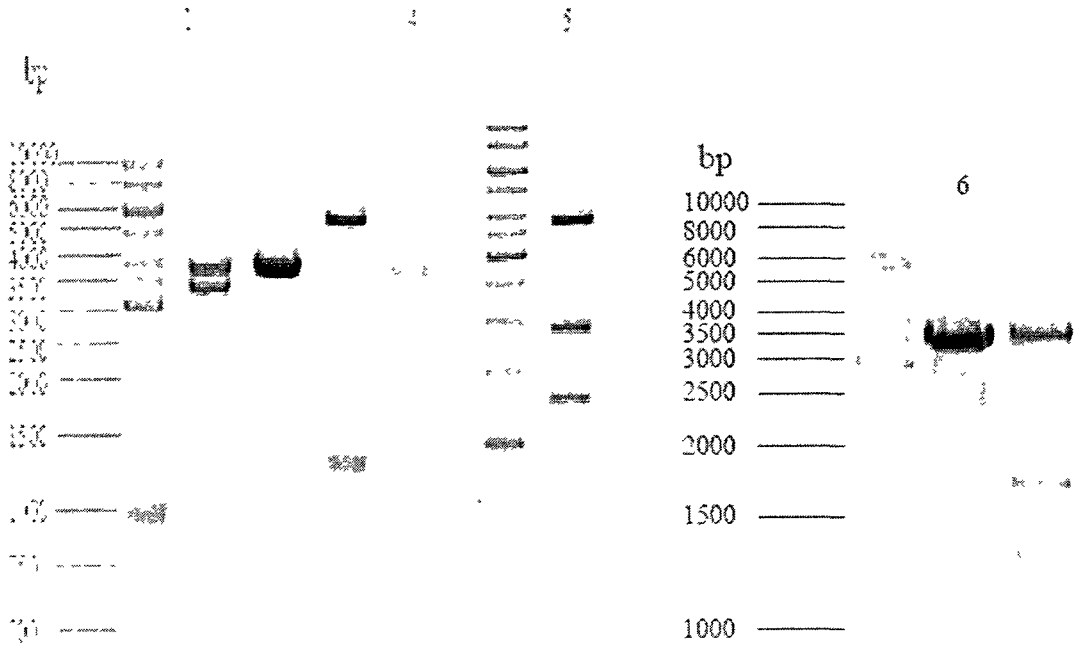


Figure 5G: *SacI* or *BsrGI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9G.

Table 9H: List of successfully cloned genomic DNA-derived genes [a]

Name [b]	Position [c]	Digestion (*SacI*) [d]
PsojCbh1040	1	3773, 2159, 1028
CcinBgl5110	2	3773, 3524, 534
AnidCbh1020	3	3773, 3245
PsojEgl3050	4	3773, 2510
NcraCbh1010	5	3773, 1925, 1030, 522
NcraEgl4010	6	3773, 2087, 815
NcraEgl1020	7	3773, 1929, 1279
NhaeCdh020	8	3773, 3490
Name [e]	**Position**	**Digestion (*BsrGI*) [f]**
CcinEgl4070	9	5619, 1066
CcinBgl5120	10	5619, 2892
CcinBgl5060	11	5619, 1914
AnidBgl5020	12	5619, 2645
CcinBgl5050	13	5619, 2195, 504
CcinBgl5080	14	5619, 3420
CcinEgl4090	15	5619, 1710
CcinCbh1030	16	5619, 1325
CcinSEQ4423	17	5619, 1143

[a] Genes were successfully cloned by Yan Li.
[b] Gene that was successfully cloned into ANIp5 based on restriction enzyme analysis.
[c] Location of the lane in Figure 5H harbouring the positive clones.
[d] *SacI* fragments obtained and expected for positive clones.
[e] Gene that was successfully cloned into ANIp7G based on restriction enzyme analysis.
[f] *BsrGI* fragments obtained and expected for positive clones.

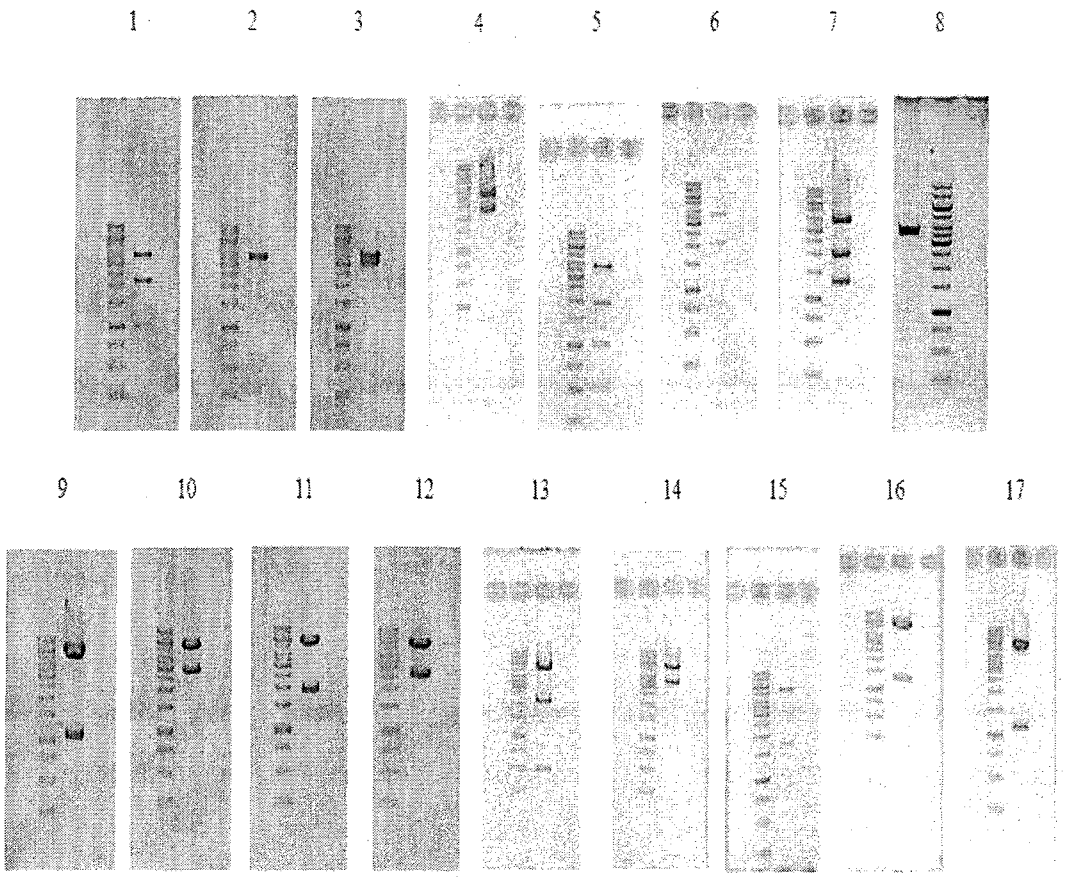


Figure 5H: *SacI* or *BsrGI* restriction enzyme screening results to identify the positive genomic DNA-derived clones presented in Table 9H.

3.2 Sequencing

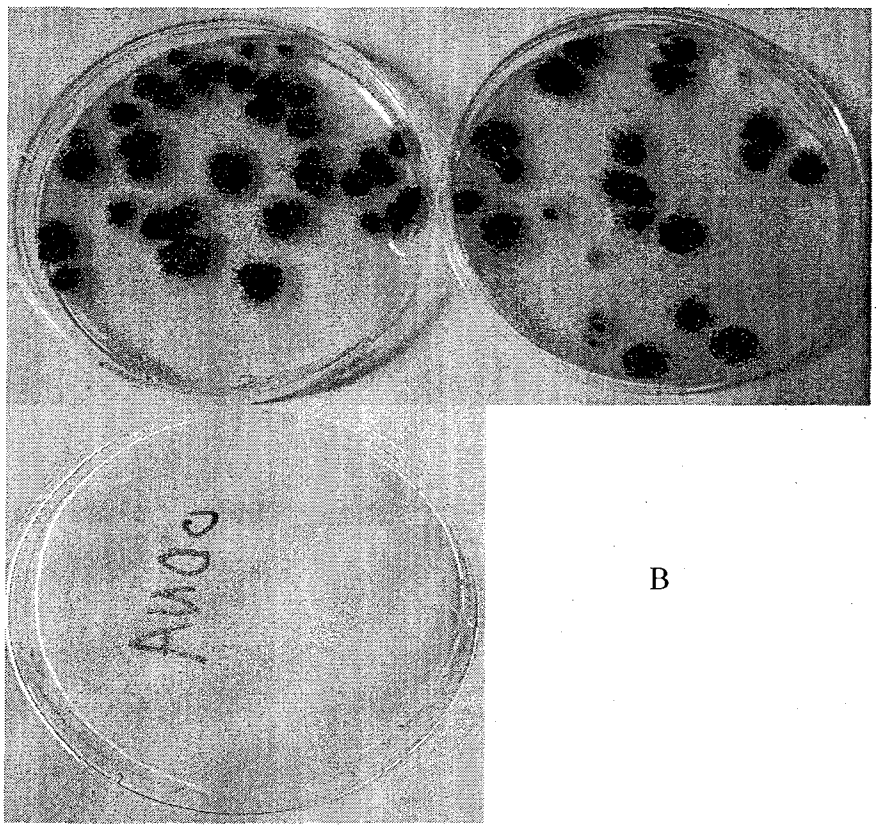
I successfully cloned a total of 149 genomic DNA-derived genes and 70 cDNA-derived genes. Since the cDNA-derived genes were previously sequenced as entry clones, DNA sequencing was only performed on the genomic DNA-derived genes. The sequencing results were compared with the target gene DNA sequence downloaded from two fungal genome databases (JGI and BROAD INSTITUTE). Summarizing the sequencing results, 41 of the completely sequenced target genes were 100% identical with the fungal genome database sequence. Two completely sequenced target genes had a 1 base pair difference that resulted in a silent mutation. Two of the completely sequenced target genes had 1 or 2 base pair difference within an intron. 39 partially sequenced target genes (with about 500 bp reads from each end) were found to be 100% identical with the gene in the respective fungal genome sequence. The remaining 65 sequenced target genes were found to differ from the genome sequences by at least one base pair that changed the encoded protein.

Three target genes were biochemically characterized. The DNA sequence of two of these genes, AnidEgl2020 and FgraEgl2010 were 100% identical with the publicly available gene sequences, whereas AfumEgl2010 was identical to its gene sequences in the sequenced genomes over the 788 nucleotides that were sequenced; however about 600 nucleotides of the cloned gene still need to be sequenced.

3.3 *A. niger* transformations and microtiter plate-scale culture growth

The *A. niger* host strain N593 *ΔglaA::hisG* was transformed (as described in section 2.5) with 74 out of the 149 plasmid constructions (genomic DNA-derived set) that were

verified by restriction enzyme digestion. Figure 6 shows an example of *A. niger* transformation results. Sixty out of the 74 clones harbouring the expected gene inserts were screened using microplate culture filtrates (Figure 7) (as described in section 2.6) to identify transformants that efficiently expressed the target gene.



A

B

Figure 6: *A. niger* transformation. A. *A. niger* strain N593 *Δ glaA::hisG* transformants obtained with ANIp7G-NcraEgl2010. Each colony represents an individual transformant. **B.** Results obtained when the transformation procedure was performed without the inclusion of plasmid DNA. Since *A. niger* strain N593 *Δ glaA::hisG* is *pyrG*⁻, the untransformed protoplasts can not survive on MM plate without uracil and uridine.

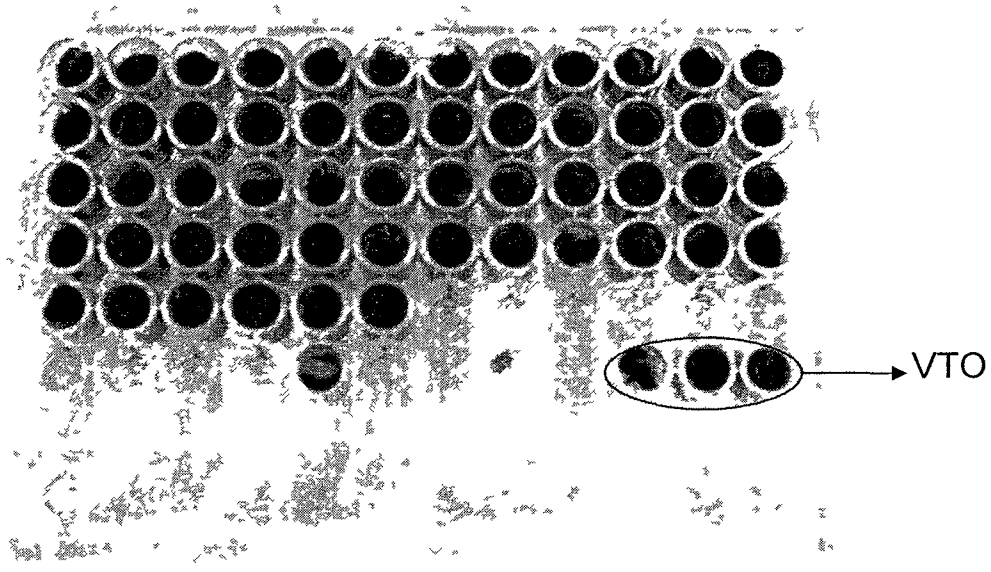


Figure 7: Microtiter plate cultures of individual transformants for ANIp7G-NcraEgl2010. The VTO (vector transformants only) is *A niger* strain N593 *Δ glaA hisG* transformed with integrating plasmid ANIp7G without a target gene

3.4 Screening for functionally-expressed endoglucanases

The initial screening to identify functionally-expressed target endoglucanases was performed as described in section 2.6. The screening of 51 endoglucanase genes (genomic DNA-derived set) identified six that were functionally expressed. They were AfumEgl2010, AnidEgl2020, FgraEgl2010, FgraEgl1020, FgraEgl2020 and NcraEgl2010. This study completed the detailed characterization of AfumEgl2010, AnidEgl2020, and FgraEgl2010. An example of the initial screening results of one functionally expressed target endoglucanase, FgraEgl2010, and a target endoglucanase, FgraEgl4020, that was not expressed at levels significantly above background is shown in Figure 8.

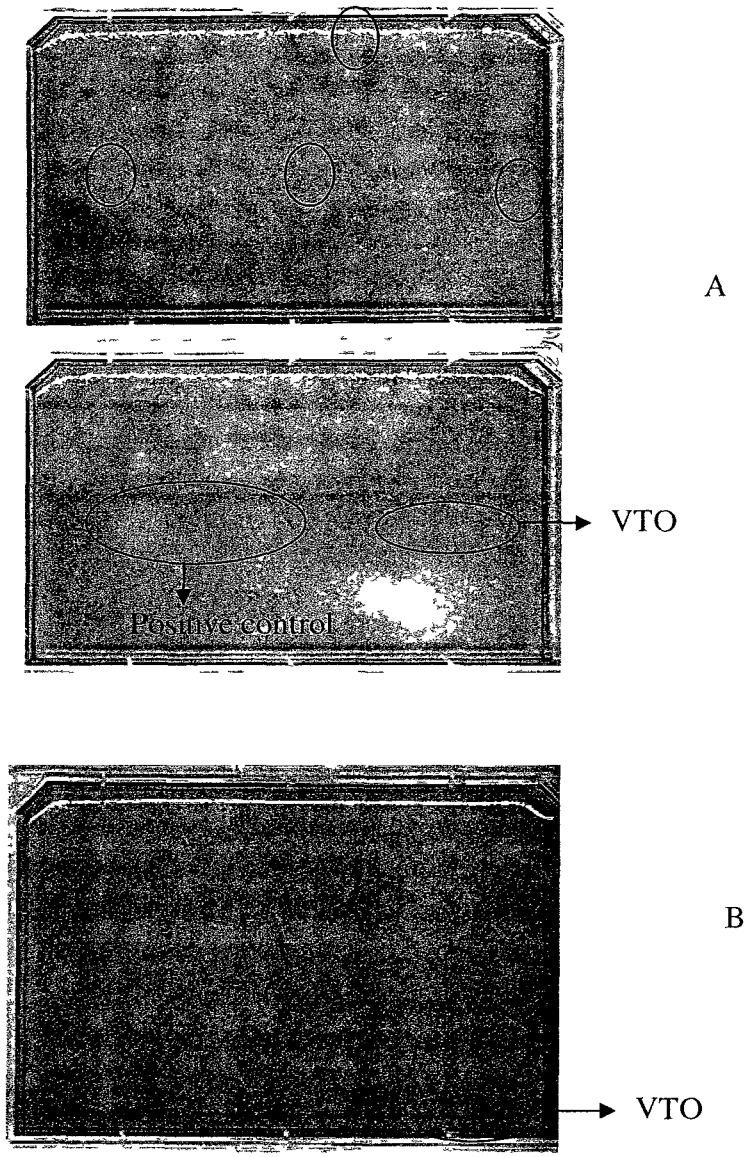
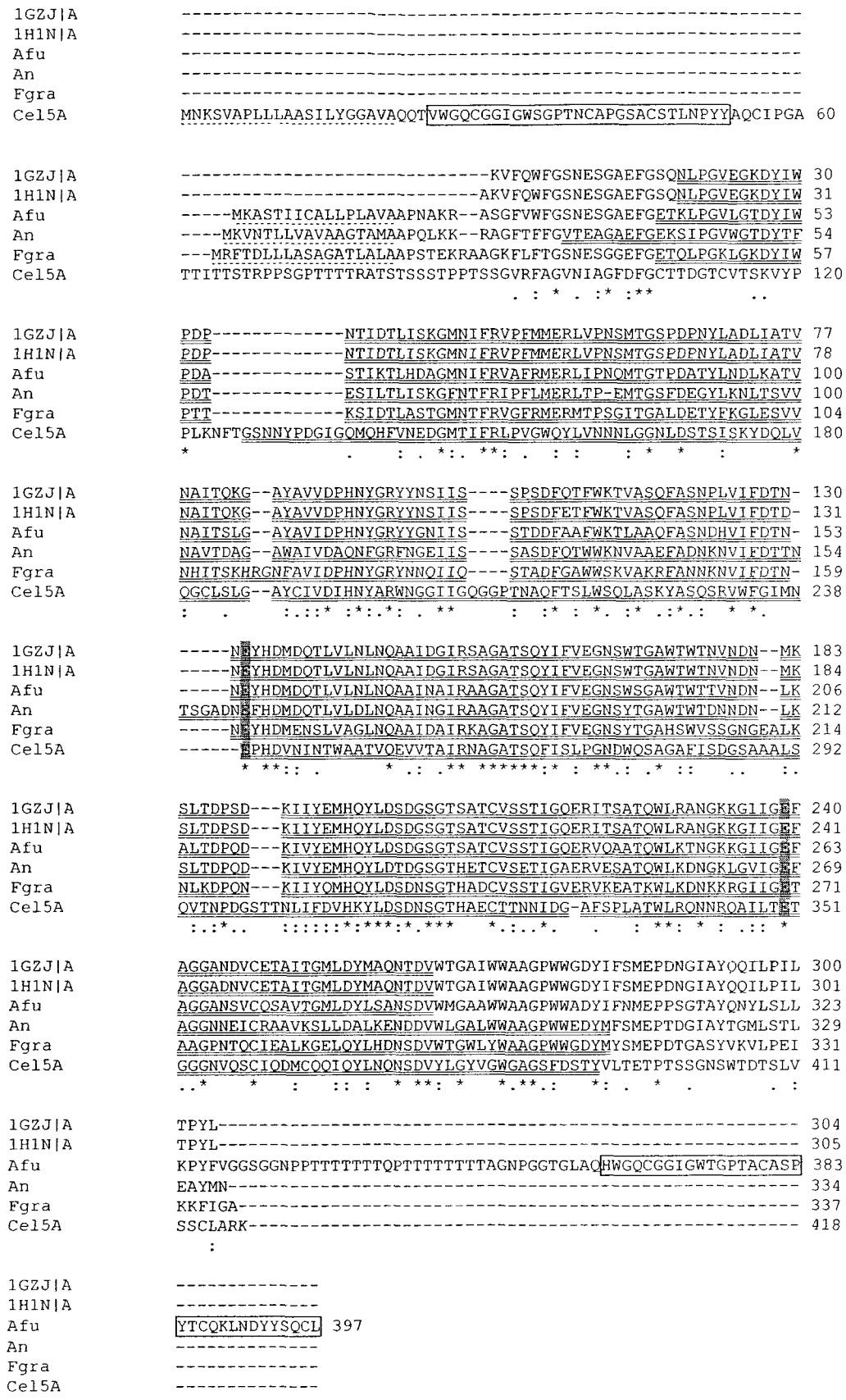


Figure 8: Screening for endoglucanase activity. A. Screening results for independent transformants obtained with functionally expressed target endoglucanase FgraEgl2010. FgraEgl2010 was expressed at much higher levels than the VTO control. The halos indicated by ovals in the lower image in Panel A were generated by a commercial endoglucanase II (Megazyme). Ovals in the top image of Panel A identify the independent transformants that were chosen for further characterization. **B.** Initial screening results obtained with FgraEgl4020 that did not express endoglucanase at levels that were significantly above background expression by the VTO control.

3.5 Protein sequence analysis of three functionally expressed endoglucanases

The A Conserved Domain Database and Search Service (NCBI) was used to predict the glycoside hydrolase family and conserved domains for each target endoglucanase. The results showed all three endoglucanases (AfumEgl2010, AnidEgl2020, and FgraEgl2010) belong to glycoside hydrolase family 5 (data not shown). In addition to the GH family domain, one of the endoglucanases (AfumEgl2010) has a carbohydrate-binding module (CBM) (Figure 9). The SignalP tool was used to predict whether the three endoglucanases had signal peptide that could direct secretion. SignalP predicted that all three target endoglucanases had signal peptides (Figure 9). The number of amino acids in the predicted signal peptides for AfumEgl2010, AnidEgl2020, and FgraEgl2010 were 16 aa, 17 aa and 18 aa respectively. The three target endoglucanases were aligned with the *T. reesei* Eg2/Cel5A (Genebank accession #: M19373) and the two proteins (1GZJ|A and 1H1N|A) that were used to predict the 3D structure of the catalytic domains of the three target enzymes. The ClustalW2 alignment (Figure 9) identified the two conserved catalytic glutamic acid residues that were predicted to be the catalytic site amino acid. The predicted molecular masses of AfumEgl2010, AnidEgl2020, and FgraEgl2010 were 41.0, 35.0 and 35.3 kDa, without signal peptide.

Figure 9: Protein sequence alignments using ClustalW2. Alignment of Afu (AfumEgl2010), An (AnidEgl2020), and Fgra (FgraEgl2010), 1GZJ/A, and 1H1N/A with the *T. reesei* Cel5A (Eg2/Cel5A) (Genebank accession #: M19373). 1GZJ/A (gi|23200107) was used as template to predict the 3D structure of the catalytic domains of AfumEgl2010 and AnidEgl2020. 1H1N/A (gi|23200122) was used as template to predict the 3D structure of the catalytic domain of the FgraEgl2010. The residues highlighted in grey are the predicted catalytic residues. Regions labelled by a dotted line represent secretory N terminal pre-sequences (signal peptide sequence); regions boxed represent carbohydrate-binding module (CBM); regions with a double line below them represent the glycoside hydrolase family 5 domains.


1GZJ|A
1H1N|A
Afu
An
Fgra
Cel5A

The 3D structure of the catalytic domains of the three target endoglucanases were predicted using Swiss-Model software. The structure of *Thermoascus aurantiacus* family 5 endoglucanase (accession: 1GZJ/A; gi/23200107) was used as template to predict the 3D structures of the catalytic domains of AfumEgl2010 and AnidEgl2020 (Figure 10 and 11 respectively). *Thermoascus aurantiacus* family 5 endoglucanase has 77% amino acid sequence identity with AfumEgl2010 and 65% identity with AnidEgl2020. The atomic resolution structure of the major endoglucanase from *Thermoascus aurantiacus* (accession: 1H1N/A; gi/23200122) was used as the template structure to model the 3D structure of the catalytic domain of FgraEgl2010 (Figure 12). FgraEgl2010 and 1H1N/A share 60% amino acid sequence identity. All three modelled endoglucanases showed an eight fold (α/β) barrel structure. The predicted catalytic residues were glutamic acid residues 155 and 262 for AfumEgl2010, residues 161 and 268 for AnidEgl2020 and residues 161 and 270 for FgraEgl2010. The predicted catalytic glutamates of all three modelled endoglucanases were located on opposite interior walls of the (α/β) barrel.

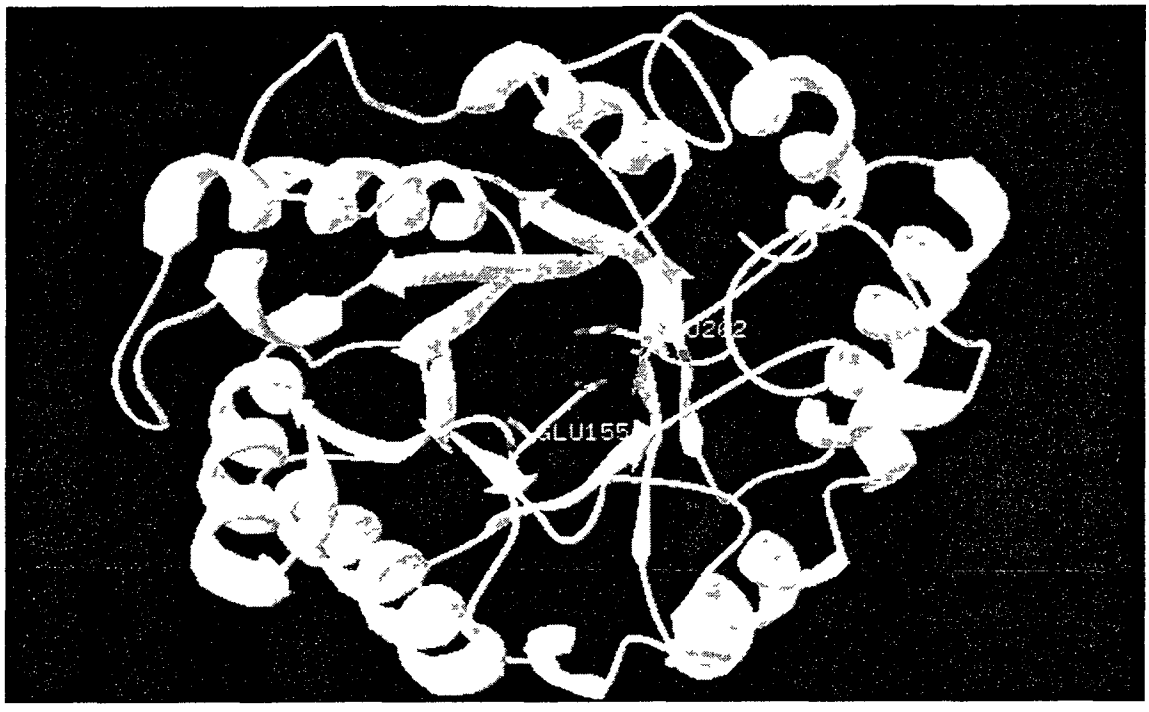


Figure 10: Predicted 3D structure of the catalytic domain of AfumEgl2010. The two predicted catalytic residues, glutamic acid residues 155 and 262, are shown in the ball and stick format, and rest of the protein in ribbon format.

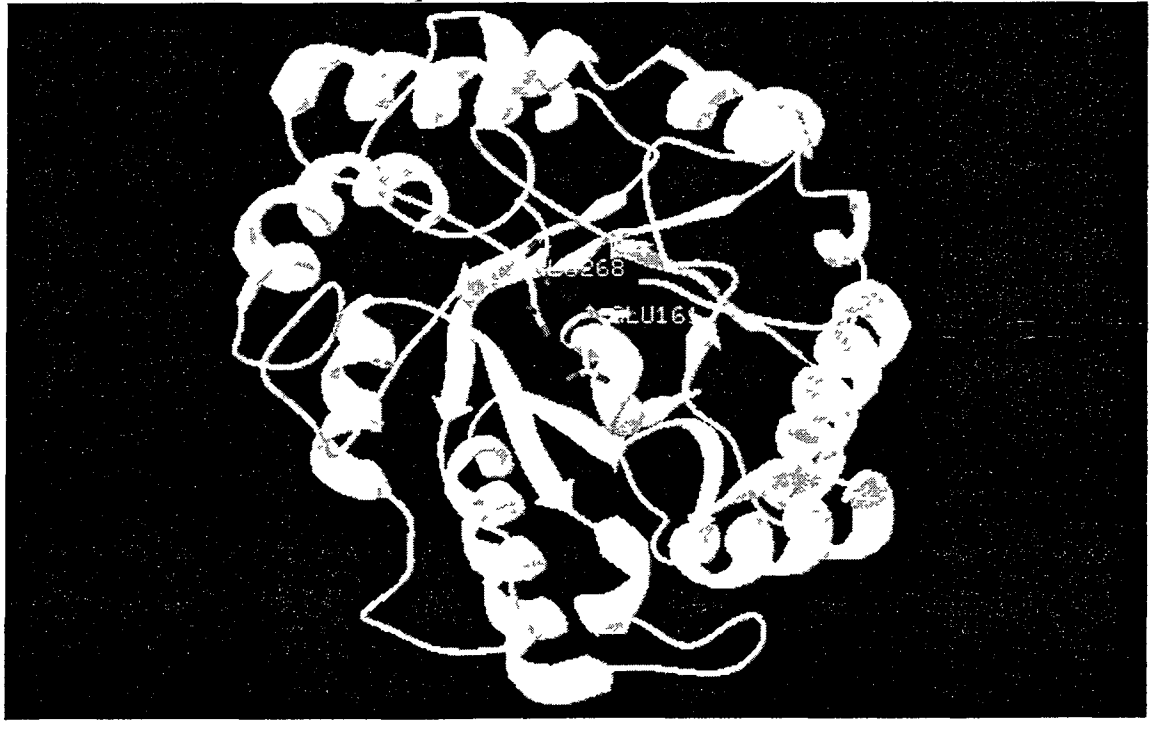


Figure 11: Predicted 3D structure of the catalytic domain of AnidEgl2020. The two predicted catalytic residues, glutamic acid 161 and 268, are shown in the ball and stick format, and rest of the protein in ribbon format.

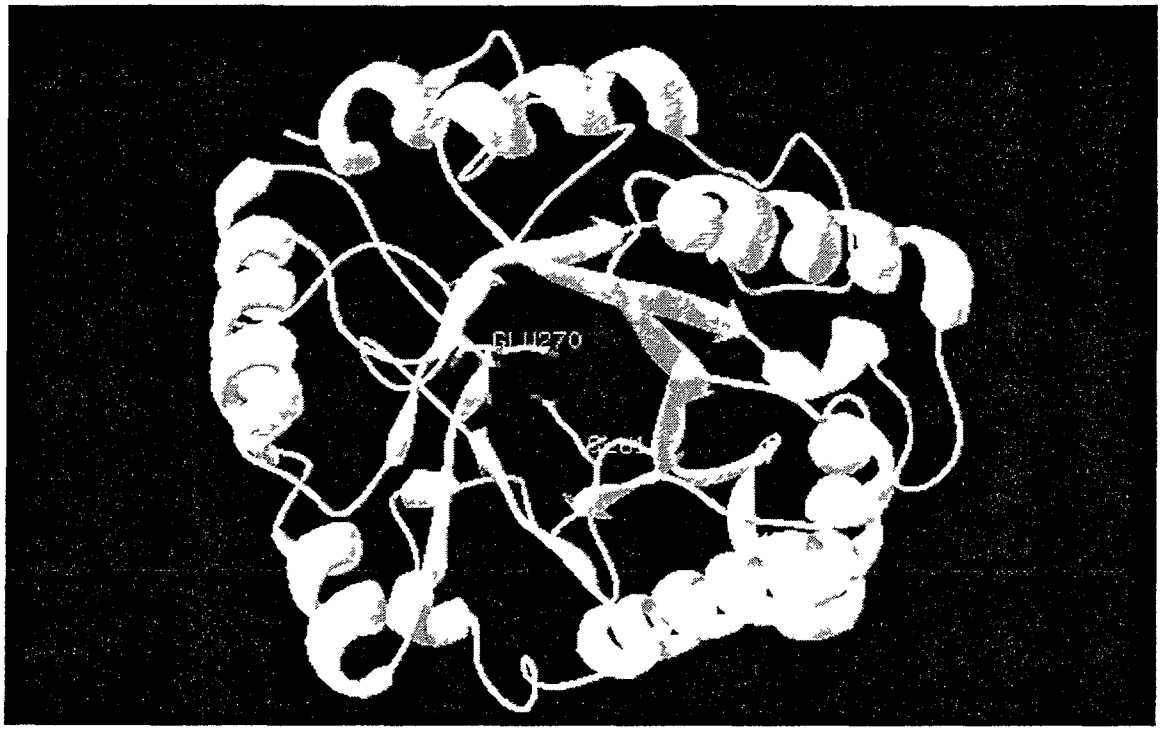


Figure 12: Predicted 3D structure of the catalytic domain of FgraEgl2010. The two predicted catalytic residues, glutamic acid 161 and 270, are shown in the ball and stick format, and rest of the protein in ribbon format

3.6 Target endoglucanase expression and desalting

A total of three functionally expressed endoglucanases (AfumEgl2010, AnidEgl2020 and FgraEgl2010) were chosen from the initial screening experiments for further biochemical characterization. For each target endoglucanase, the four individual transformants that showed the highest activity in the initial screening were used to inoculate MMJ micro plate well cultures in Petri dishes as described in section 2.7. The culture filtrate containing the target endoglucanase was partially purified through Ultrafree-0.5 centrifugal filter devices (Millipore) as described in section 2.7. The culture filtrate was assayed for residual sugar essentially as described in section 2.10.1 except that substrate was not added. The results showed the purification step successfully removed any remaining sugar in the culture filtrate (data not shown). The presence and expression level in the four individual transformants of each target secreted endoglucanase was estimated on SDS-PAGE gel as described in section 2.9. The results (Figure 13) showed that the estimated molecular mass of AfumEgl2010 was 46.0 kDa and AnidEgl2020 was 35.0 kDa. In Figure 14, the estimated molecular mass of FgraEgl2010 was 33.0 kDa. Enzyme activity assay as described in section 2.10.1 was also performed to check the activity of endoglucanase expressed by each transformant. Based on the SDS-PAGE gel and enzyme activity assay results (data not shown), transformants #36 and #43 (Figure 13), and #60 (Figure 14) of AnidEgl2020, AfumEgl2010 and FgraEgl2010, respectively were chosen for further detailed characterization. Three cDNA-derived endoglucanases (ApulSEQ15654, StheSEQ13822 and GtraSEQ630) were also chosen for detail characterization.

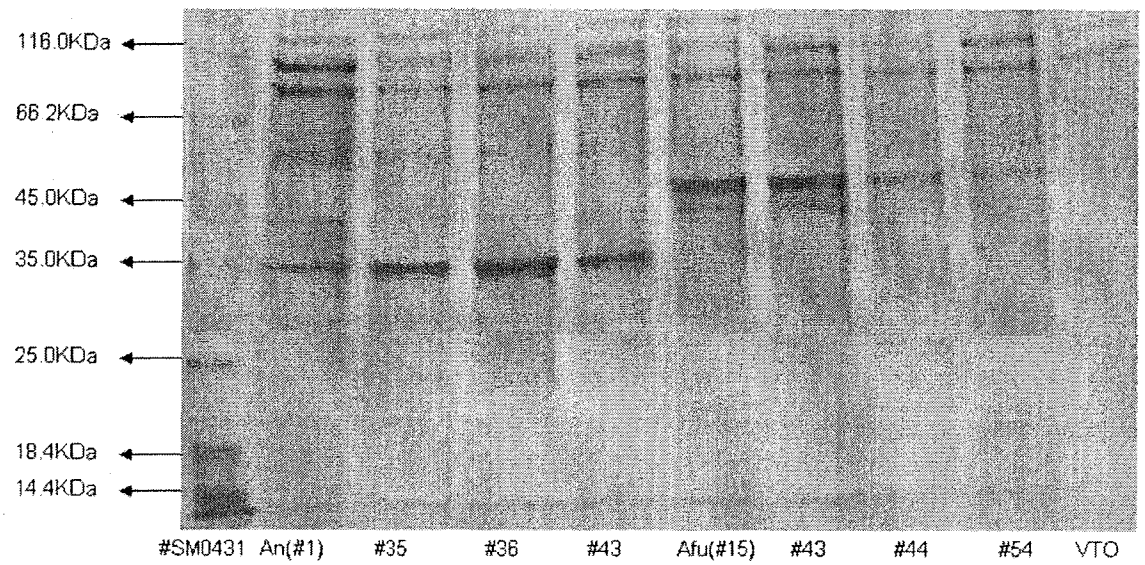


Figure 13: SDS-PAGE analysis of four independent *A. niger* transformants obtained with An (AnidEgl2020) and Afu (AfumEgl2010).

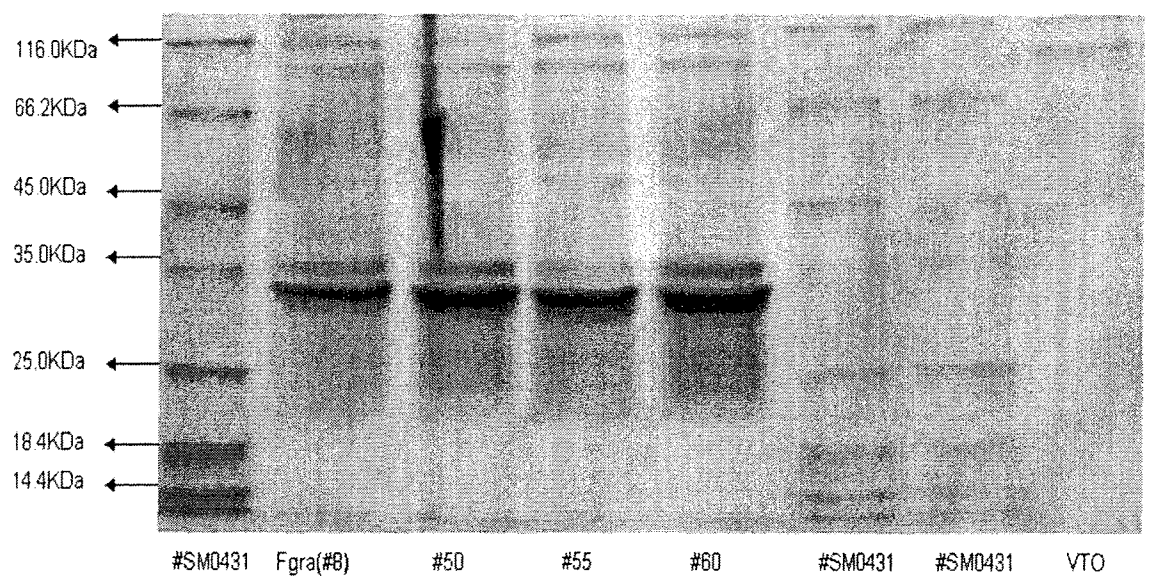


Figure 14: SDS-PAGE analysis of four independent *A. niger* transformants obtained with Fgra (FgraEgl2010).

3.7 Protein concentration determination

Protein concentrations of 3 genomic and 3 cDNA derived, as well as *T. reesei* Cbh1/Cel7A (Iogen Corp.) and *T. reesei* Eg2/Cel5A (Iogen Corp.) were determined as described in section 2.9. The SDS-PAGE gel images that used to determine the protein concentrations are shown in Figures 15 and 16. The predicted and estimated molecular masses of the proteins and protein concentration of target endoglucanases are shown in Table 10.

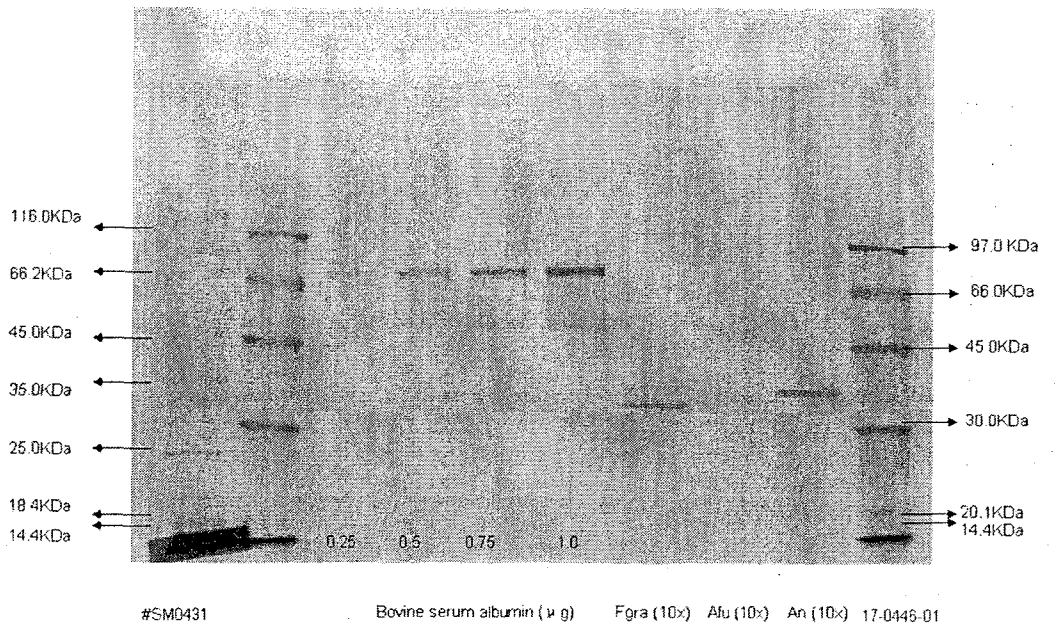


Figure 15: The SDS-PAGE gel of Fgra, Afu and An. Each diluted 10 fold (10x). Fgra (FgraEgl2010), Afu (AfumEgl2010) and An (AnidEgl2020).

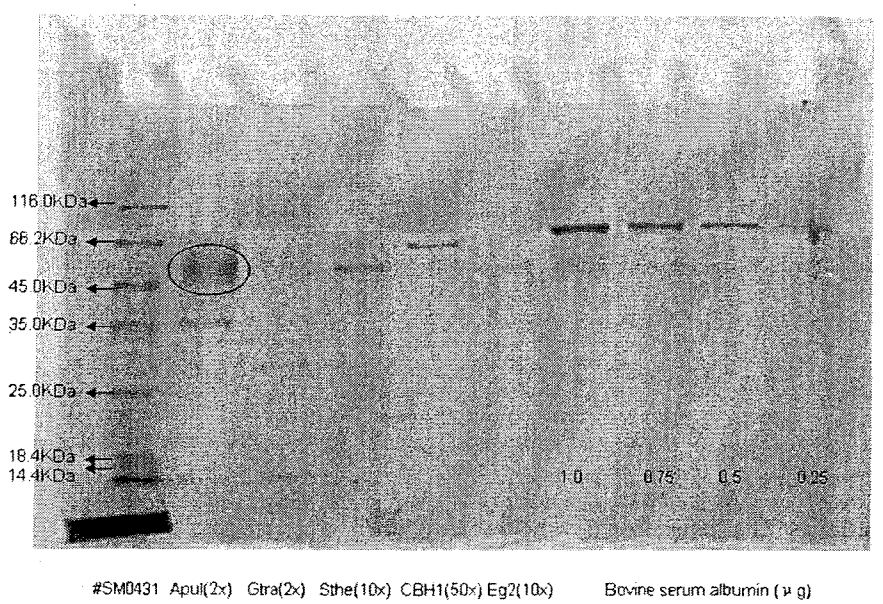


Figure 16: The SDS-PAGE gel of Apul, Gtra, Sthe, CBH1, and Eg2. Apul (ApulSEQ15654) (2x), Gtra (GtraSEQ630) (2x), Sthe (StheSEQ13822) (10x), CBH1 (Cbh1/Cel7A) (50x) (Iogen Corp.) and Eg2 (Eg2/Cel5A) (10x) (Iogen Corp.).

Table 10: The predicted molecular mass (kDa), estimated molecular mass (kDa) and estimated protein concentration (mg/ml and μM) of target endoglucanases and Cbh1/Cel7A and Eg2/Cel5A

Target protein	Predicted molecular mass (kDa) without signal peptide	Estimated molecular mass (kDa)	Experimental concentration (mg/ml) [a]	Experimental concentration (μM)
FgraEgl2010	35.3	33.0	1.28	36.3
AfumEgl2010	41.0	46.0	0.72	17.6
AnidEgl2020	35.0	35.0	1.26	36.0
ApulSEQ15654	42.8	47.0	0.32	7.5
GtraSEQ630	24.0	29.0	0.16	6.7
StheSEQ13822	40.9	46.0	0.92	22.5
Cbhl/Cel7A (Iogen) [b]	52.2	60.0	5.30	101.5
Eg2/Cel5A(Iogen) [c]	42.1	46.0	0.76	18.1

[a] Final concentration of target endoglucanases after sugar was removed.
[b] *T. reesei* Cbh1/Cel7A obtained from Iogen Corporation.
[c] *T. reesei* Eg2/Cel5A obtained from Iogen Corporation.

3.8 Basic characterization of target endoglucanases

The optimal pH of three endoglucanases (AfumEgl2010, AnidEgl2020 and FgraEgl2010) was determined as described in section 2.10.2. All three target endoglucanases showed the highest activity in citrate buffer pH 5.0 (Figure 17). To determine the optimal temperature, the method described in section 2.10.3 was used. AfumEgl2010 and AnidEgl2020 showed maximal activity at temperatures of 70°C and 60°C respectively. FgraEgl2010 showed maximal activity at a relatively low optimal temperature of 40°C (Figure 18). The thermostability assay was performed as described in section 2.10.4. All three endoglucanases were stable during 30 minutes pre-incubation at 60°C (Figure 19). Compared to AnidEgl2020 and FgraEgl2010, AfumEgl2010 is the most stable endoglucanase. AfumEgl2010, AnidEgl2020 and FgraEgl2010 retained half of their activity during a 30 minute pre-incubation at 75°C, 65°C and 70°C respectively (Figure 19).

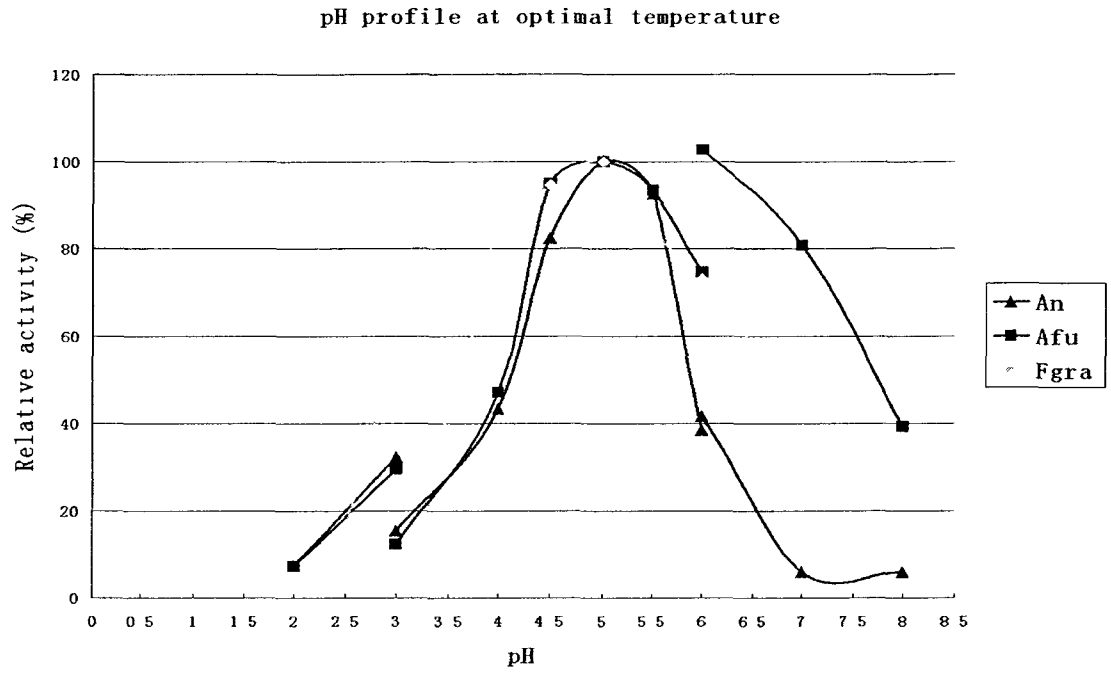


Figure 17: The pH profiles of An, Afu and Fgra at their optimal temperature. An (AnidEgl2020), Afu (AfumEgl2010), Fgra (FgraEgl2010)

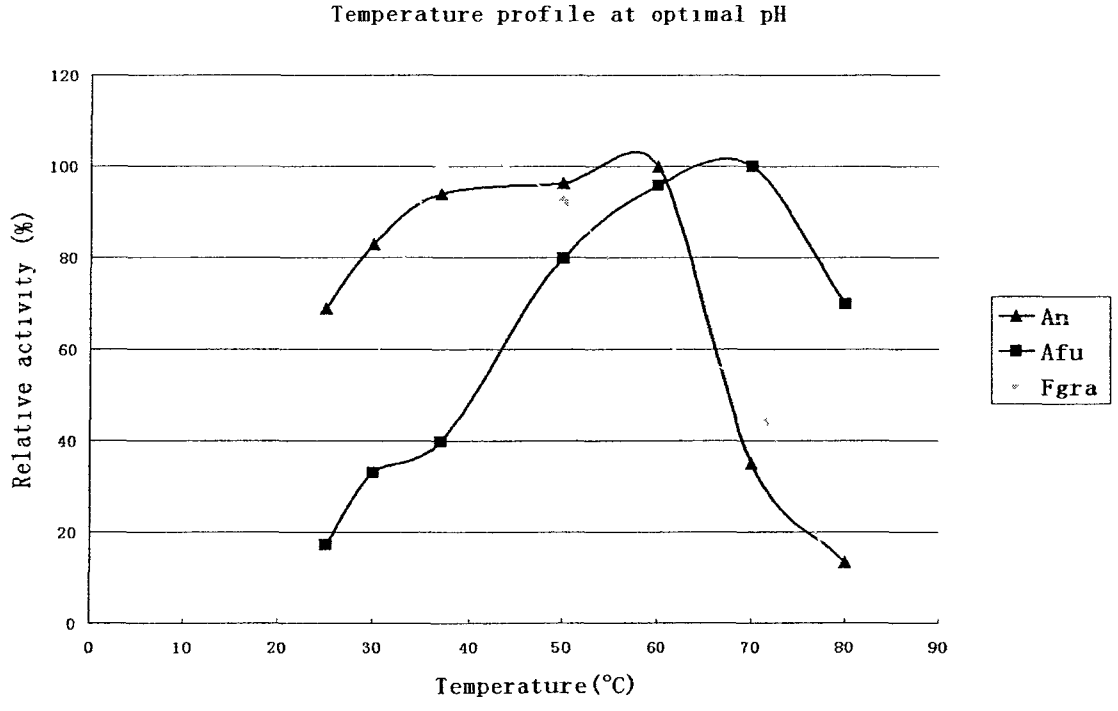


Figure 18: The temperature profiles of An, Afu and Fgra at their optimal pH. An (AnidEgl2020), Afu (AfumEgl2010), Fgra (FgraEgl2010)

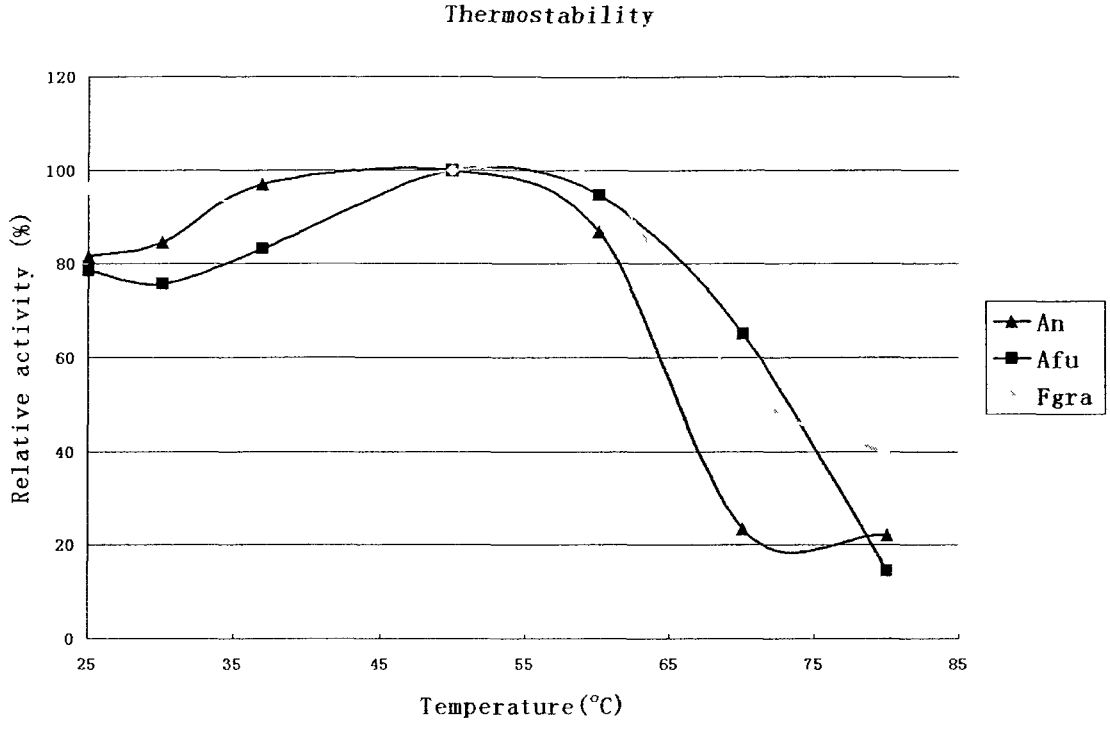


Figure 19: The thermostability profiles of An, Afu and Fgra after a 30 minute pre-incubation at temperature from 25°C to 80°C. An (AnidEgl2020), Afu (AfumEgl2010), Fgra (FgraEgl2010)

3.9 Enzyme kinetics of target endoglucanases

The kinetic parameters of the six endoglucanases and the *T. reesei* Eg2/Cel5A (Iogen Corp.) were determined as described in section 2.11. The results are shown in Table 11. The Km values range from 2.0 to 29 mg PASC/ml. The Vmax values range from 5.7 to 41 μmole of reducing sugar ends per mg of enzyme per minute. The kcat values range from 4/sec to 29/sec. The kcat/Km ratio is also shown in Table 11. AfumEgl2010 has the highest kcat/Km ratio of 3.0 which is 2 times higher than the kcat/Km ratio of the *T. reesei* Eg2/Cel5A (Iogen Corp.). AnidEgl2020 has a significantly lower kcat/Km ratio of 0.1 compared to that of *T. reesei* Eg2/Cel5A (Iogen Corp.).

Table 11: Enzyme kinetic parameters of 6 endoglucanases and the Eg2/Cel5A at 37°C and the optimal pH of each endoglucanase

Target name	Vmax (µmole/mg/min)		Km (mg/ml)		kcat (sec^{-1})	kcat/Km (ml/sec*mg)
	Value	Std.error	Value	Std.error	Value	Value
GtraSEQ630	16	3.1	13	3.9	6	0.5
ApulSEQ15654	41	6.9	10	2.9	29	2.9
StheSEQ13822	8.5	0.62	3.3	0.58	6	1.8
FgraEgl2010	24	5.6	13	4.8	14	1.1
AfumEgl2010	8.8	0.85	2.0	0.55	6	3.0
AnidEgl2020	7.2	1.7	29	8.8	4	0.1
Eg2/Cel5A(Iogen)[a]	5.7	0.59	2.6	0.70	4	1.5

[a] *T. reesei* Eg2/Cel5A obtained from Iogen Corporation.

3.10 Synergy assays of target endoglucanases

Synergy assays were performed using either Avicel or PASC as described in section 2.12. The degree of synergism is determined by the ratio between the experimental enzyme activity of the combined enzyme mixture and the added enzyme activities of the individual enzymes (Zhang and Lynd, 2004). The synergy assay results of the target endoglucanases combined with the *T. reesei* Cbh1/Cel7A are shown in Figure 20. Based on these results, all the target enzymes showed a higher degree of synergism on Avicel than on PASC (Table 12). On Avicel, *T. reesei* Eg2/Cel5A (Iogen Corp.) showed the highest degree of synergism of 3.3 when combined with the *T. reesei* Cbh1/Cel7A (Iogen Corp.). AfumEgl2010 combined with the *T. reesei* Cbh1/Cel7A (Iogen Corp.) showed a maximal degree of synergism of 2.9. AnidEgl2020 showed the lowest degree of synergism of 1.3. On PASC, the highest degree of synergism was observed when AfumEgl2010 was combined with the *T. reesei* Cbh1/Cel7A (Iogen Corp.). The lowest degree of synergism on PASC was observed when AnidEgl2020 was combined with the *T. reesei* Cbh1/Cel7A (Iogen Corp.).

Figure 20: The degree of synergism when the target endoglucanases were combined with the *T. reesei* Cbh1/Cel7A (Iogen Corp.). The top x axis is the amount of endoglucanase that was added into each enzyme assay. The bottom x axis is the amount of exoglucanase that was added into each enzyme assay. The data points in blue indicate the enzyme activity of the *T. reesei* Cbh1/Cel7A (Iogen Corp.) alone at different concentrations. The data points in plum indicate the enzyme activity of endoglucanase alone at different concentrations. The data points in lime indicate the added activity of individual enzyme activity of the *T. reesei* Cbh1/Cel7A (Iogen Corp.) and an endoglucanase at the concentration indicated on the lower and upper x-axis. The data points in blue-gray indicate the experimental enzyme activity of combining the *T. reesei* Cbh1/Cel7A (Iogen Corp.) and the endoglucanase at the concentration indicated on the lower and upper x-axis. The arrows on the figure indicate the combination of the *T. reesei* Cbh1/Cel7A (Iogen Corp.) and the endoglucanase that gives the highest degree of synergism. In the figure, 0.1 *OD* equals to 0.0056 μmole/min/ml, which is μmole of reducing sugar ends produced per minute per ml of enzyme. In the figure, cbh1 represents the *T. reesei* Cbh1/Cel7A obtained from Iogen Corporation; eg2 represents the *T. reesei* Eg2/Cel5A obtained from Iogen Corporation.

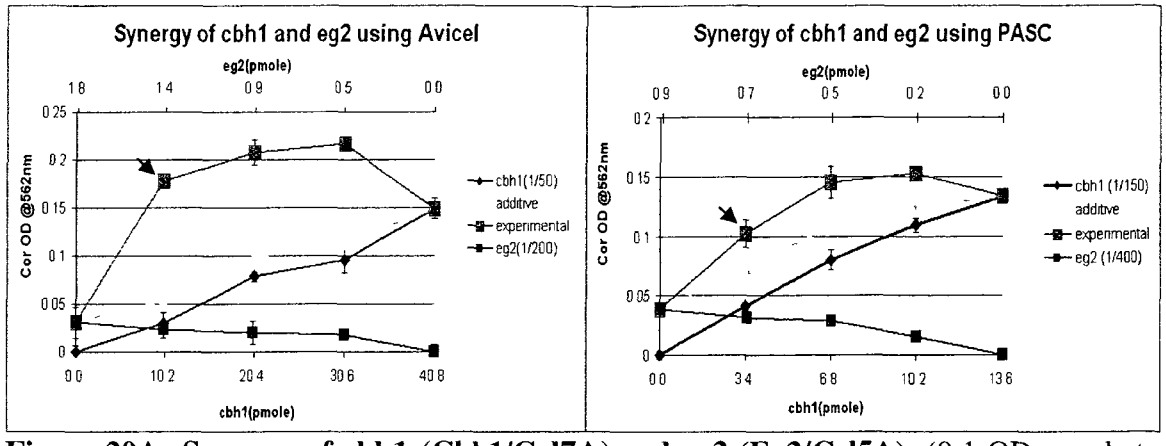


Figure 20A: Synergy of cbh1 (Cbh1/Cel7A) and eg2 (Eg2/Cel5A). (0.1 OD equals to 0.0056 μmole/min/ml.)

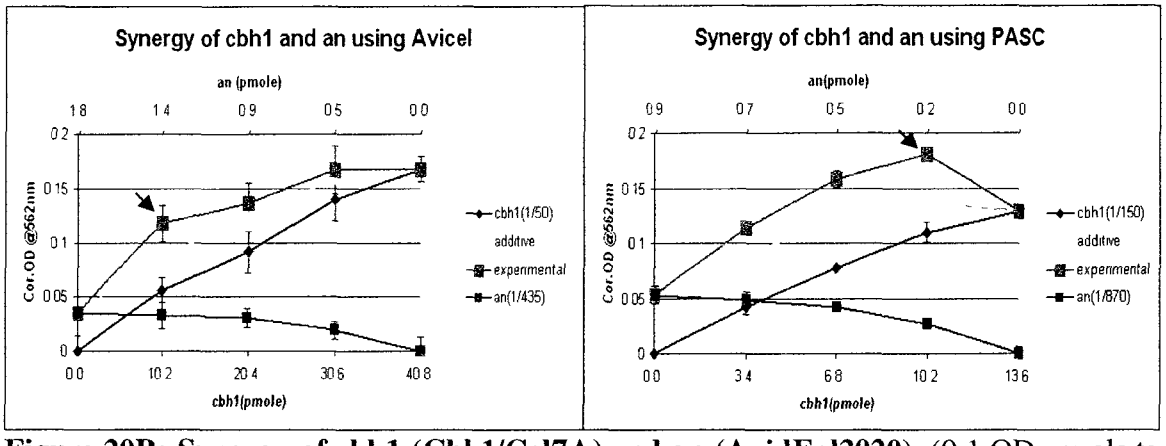


Figure 20B: Synergy of cbh1 (Cbh1/Cel7A) and an (AnidEgl2020). (0.1 OD equals to 0.0056 μmole/min/ml)

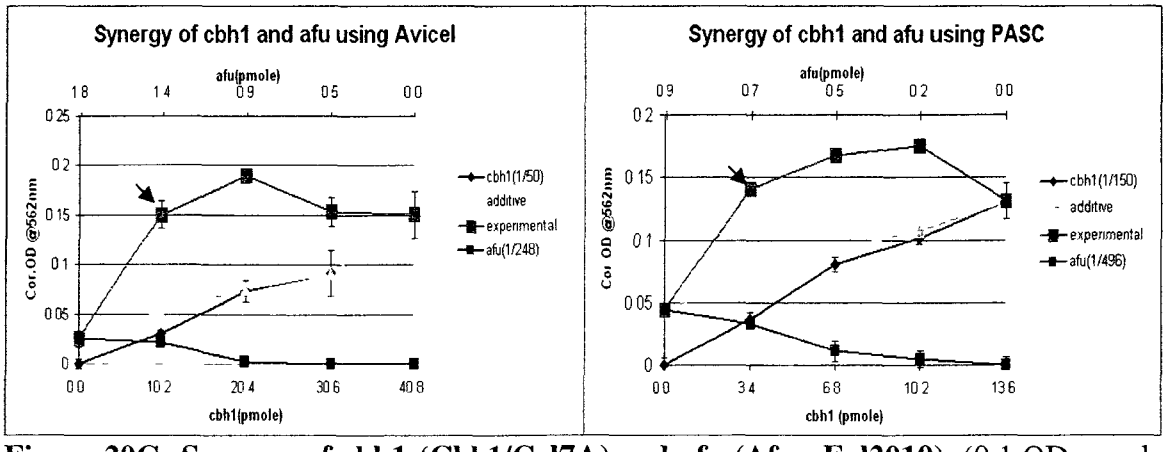


Figure 20C: Synergy of cbh1 (Cbh1/Cel7A) and afu (AfumEgl2010). (0.1 OD equals to 0.0056 μmole/min/ml)

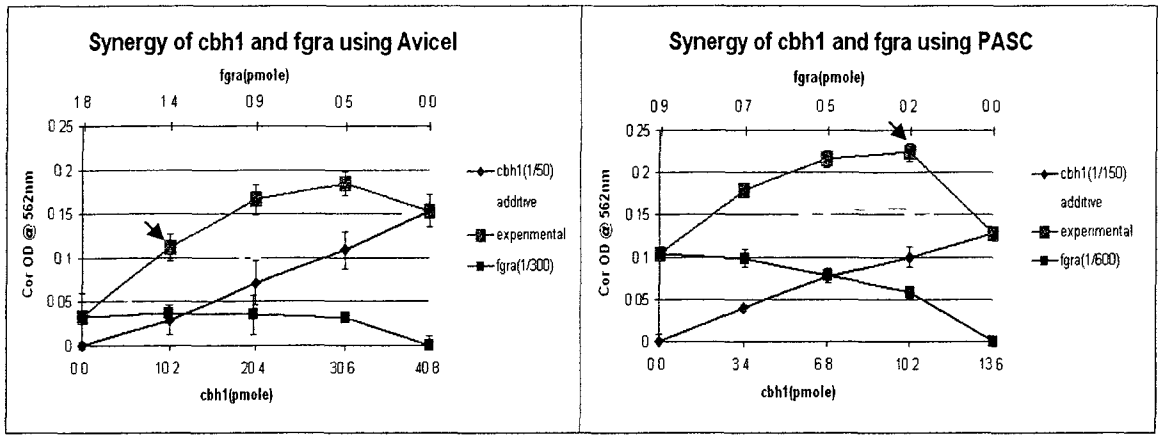


Figure 20D: Synergy of cbh1 (Cbh1/Cel7A) and fgra (FgraEgl2010). (0.1 OD equals to 0.0056 μmole/min/ml)

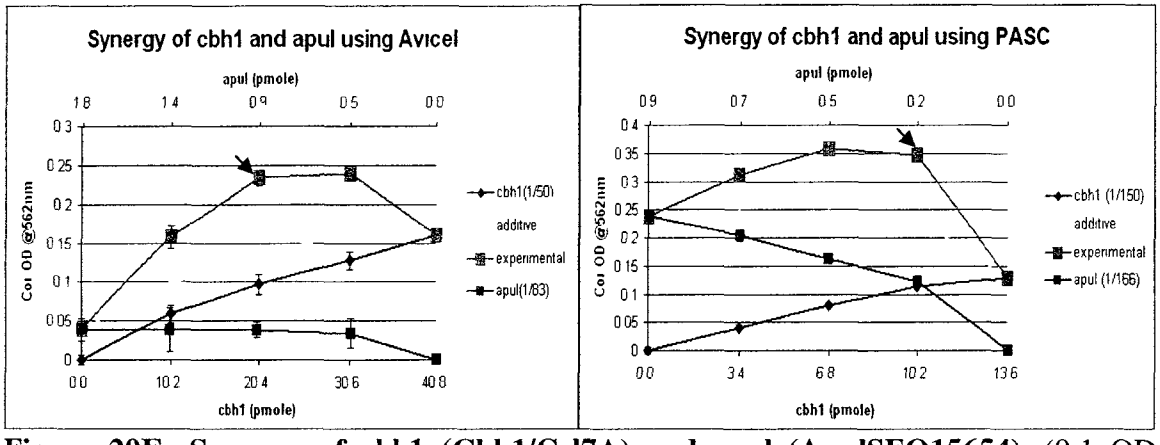


Figure 20E: Synergy of cbh1 (Cbh1/Cel7A) and apul (ApulSEQ15654). (0.1 OD equals to 0.0056 μmole/min/ml)

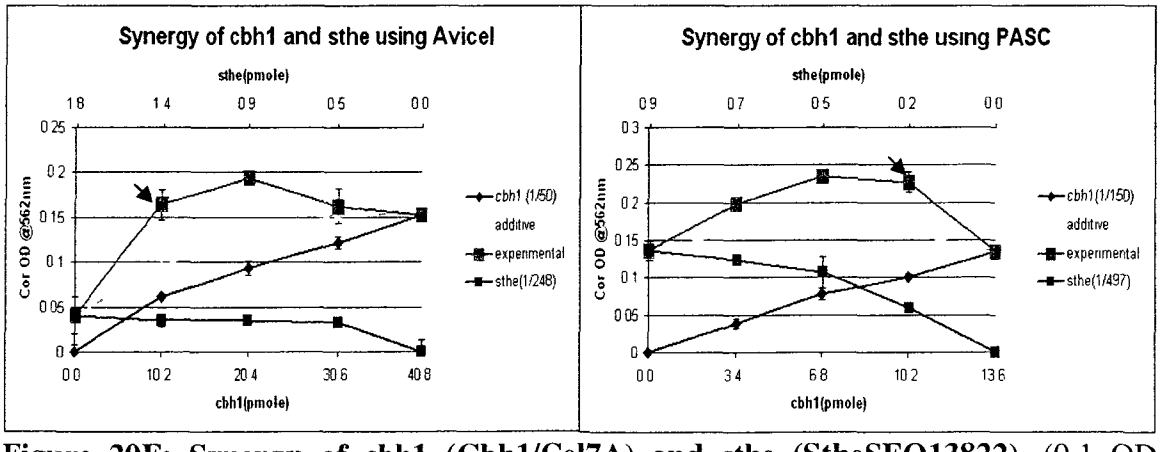


Figure 20F: Synergy of cbh1 (Cbh1/Cel7A) and sthe (StheSEQ13822). (0.1 OD equals to 0.0056 μmole/min/ml)

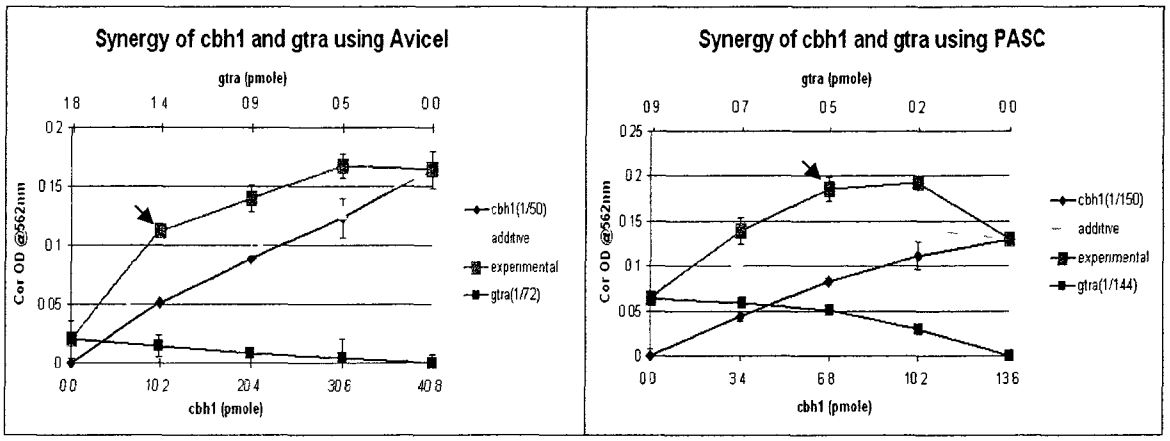


Figure 20G: Synergy of cbh1 (Cbh1/Cel7A) and gtra (GtraSEQ630). (0.1 OD equals to 0.0056 μmole/min/ml)

Table 12: The degree of synergism observed when combining each endoglucanase with the *T. reesei* Cbh1/Cel7A at 37°C

Name	Degree of synergism[a]			
	Avicel		PASC	
Eg2 (Iogen)[b]	10.2 pmole Cbh1[1] and 1.36 pmole Eg2	3.3	3.4 pmole Cbh1 and 0.68 pmole Eg2	1.4
An[c]	10.2 pmole Cbh1 and 1.36 pmole An	1.3	10.2 pmole Cbh1 and 0.23 pmole An	1.3
Afu[d]	10.2 pmole Cbh1 and 1.36 pmole Afu	2.9	3.4 pmole Cbh1 and 0.68 pmole Afu	2.0
Fgra[e]	10.2 pmole Cbh1 and 1.36 pmole Fgra	1.7	10.2 pmole Cbh1 and 0.23 pmole Fgra	1.4
Apul[f]	20.4 pmole Cbh1 and 0.91 pmole Apul	1.7	10.2 pmole Cbh1 and 0.23 pmole Apul	1.5
Sthe[g]	10.2 pmole Cbh1 and 1.36 pmole Sthe	1.7	10.2 pmole Cbh1 and 0.23 pmole Sthe	1.4
Gtra[h]	10.2 pmole Cbh1 and 1.36 pmole Gtra	1.7	6.8 pmole Cbh1 and 0.45 pmole Gtra	1.4

[a] Degree of synergism presented for each enzyme combination is equal to highest ratio of the observed combined activity divided by the predicted activity obtained by adding the activities of the individual enzymes.
[b] Eg2 represents T. reesei Eg2/Cel5A obtained from Iogen Corporation.
[c] An represents AnidEgl2020.
[d] Afu represents AfumEgl2010.
[e] Fgra represents FgraEgl2010.
[f] Apul represents ApulSEQ15654.
[g] Sthe represents StheSEQ13822.
[h] Gtra represents GtraSEQ630.
[1] Cbh1 represents T. reesei Cbh1/Cel7A obtained from Iogen Corporation.

4. Discussion

4.1 Isolation of putative cellulase genes

I cloned a set of cDNA-derived and a set of genomic DNA-derived genes that code for proteins that are predicted to be secreted enzymes or accessory proteins that are active on cellulose or hemicellulose. The genomic DNA-derived set included 207 genes from 13 different fungal species. We chose these 13 different fungal species because their complete genome sequences were available, they are known to degrade plant biomass and their cellulase systems are not well characterized.

4.2 Identification of functionally expressed endoglucanases

In this study, I performed an initial screening of 51 of the endoglucanase coding genes from the genomic DNA-derived set. From this set, I identified six functionally expressed endoglucanases. From previous experiments the success rate in identifying functionally expressed fungal enzymes is around 30% (Personal communication, Dr. Storms). Compared to that, my success rate (12%) is relatively low. In order to explain that, we sent these 51 cloned endoglucanase genes for sequencing. The sequencing results showed that 25 of the completely sequenced target genes were 100% identical with the fungal genome database sequence. Eight partially sequenced genes were found to be 100% identical with the fungal genome database sequence. One of the sequenced genes has one silent mutation. The rest of the 17 endoglucanase genes were found to have at least one missense mutation. The mutations may be the result of PCR amplification using a Taq and Pfu polymerase mix. These mutations in the target protein can be one

explanation for the low success rate of identifying functionally expressed endoglucanases. Another reason could be due to the use of genomic DNA as the template. When genomic DNA is used as template, the introns of target genes are also cloned and thus have to be removed before translation. Since the target proteins are encoded by heterologous genes, there is the possibility that the introns are not correctly spliced by *A. niger*. However, previous studies in our laboratory found that the expression of 3 heterologous xylanases cloned from *Phanerochaete chrysosporium* was similar whether the cloned genes were derived using cDNA or genomic DNA (Decelle, 2006). Other possible explanations for the low success rate are; background activity may have masked the expression of genes that were expressed at low levels and, the wrong substrates were used to test for activity.

4.3 Bioinformatic analysis of functionally expressed endoglucanases

From the six functionally expressed endoglucanases, I chose three for detailed analysis. The three target endoglucanases that were chosen all belong to glycoside hydrolase family 5 (GH5) as determined by blastp (NCBI). Previous studies of Cel5A from *T. reesei*, also a member of the GH5 family, showed that it has a CBM located at the N terminus (Saloheimo et al., 1988). However, my Blast analysis showed that AnidEgl2020 and FgraEgl2010 do not have a CBM. The third endoglucanase, AfumEgl2010, has a CBM at its C terminus.

Two 3D structures of Cel5A proteins (accession: 1GZJ/A; gi/23200107, accession: 1H1N/A; gi/23200122) were used to model 3D structures of the catalytic domains of the three GH5 endoglucanases. All three modelled endoglucanases showed an eight fold (α/β) barrel structure, which is similar to the structure found in other GH5 family

endoglucanases (Davies et al., 1998). From the protein sequence alignment results, the predicted catalytic residues of my three target endoglucanases were determined using the catalytic residues as previously determined (Lo Leggio and Larsen, 2002; Van Petegem et al., 2002). The predicted catalytic residues were glutamic acid residues 155 and 262 for AfumEgl2010, residues 161 and 268 for AnidEgl2020 and residues 161 and 270 for FgraEgl2010.

4.4 Characterization of the functionally expressed endoglucanases

Basic biochemical characterization of the selected endoglucanases was performed. The pH profile, temperature profile and thermostability were determined for each of the three GH5 endoglucanases. The optimal pH of the three GH5 endoglucanases was pH 5.0. The optimal temperature of the three GH5 endoglucanases ranged from 40°C to 70°C. All three GH5 endoglucanases retained activity after a 30 minute pre-incubation at 60°C. The optimal temperature of *T. reesei* Cel5A purified from *T. reesei* culture filtrates is 70°C (Karlsson et al., 2002). Subsequent studies reported that the optimal pH of *T. reesei* Cel5A expressed in *T. reesei* ranged from pH 4.6 to 5.0, and that Cel5A lost 50% of its activity when incubated at 70°C for 30 minutes (Qin et al., 2008). The superior thermostability of the AfumEgl2010 cellulase that I identified and characterized in this study may make it an attractive replacement for *T. reesei* Cel5A in future commercial preparations of the *T. reesei* cellulase system used for industrial scale cellulose hydrolysis.

I determined the kinetic parameters of 7 endoglucanases. These endoglucanases are the three I identified, the *T. reesei* Eg2/Cel5A obtained from Iogen Corporation, and three endoglucanases that were identified by undergraduate student Christopher St-Francois.

Three endoglucanases (GtraSEQ630, FgraEgl2010, AnidEgl2020) all without a CBM showed higher Km values than the other four endoglucanases (ApulSEQ15654, StheSEQ13822, AfumEgl2010, Eg2/Cel5A) containing a CBM. The Km represents how efficiently the enzyme binds to the substrate. The lower Km associated with the four enzymes harbour CBMs is consistent with the CBM contributing to higher substrate binding efficiency (Champe et al., 2007). When comparing the Km values of the six endoglucanases and the *T. reesei* Eg2/Cel5A (which is used in industry to convert the cellulosic biomass to fermentable sugars), AfumEgl2010 has the lowest Km of 2.0 mg/ml, *T. reesei* Eg2/Cel5A has a Km of 2.6 mg/ml and StheSEQ13822 has a Km of 3.3 mg/ml. When comparing the kcat/Km values, which indicate how efficiently the enzyme degrades its substrate (Eun, 1996), three of the endoglucanases (ApulSEQ15654, StheSEQ13822, AfumEgl2010) show higher kcat/Km values when compared with the *T. reesei* Eg2/Cel5A. For industrial applications, using less enzyme through increased enzyme efficiency, will lead to a reduction in the cost of cellulose hydrolysis for bioethanol production. ApulSEQ15654, StheSEQ13822 or AfumEgl2010 endoglucanases may be good candidates to replace the *T. reesei* Eg2/Cel5A in a commercial cellulase system because these three endoglucanases have higher kcat/Km values. However, further work needs to be done in order to confirm the commercial potential of these endoglucanases to supplement or replace Eg2/Cel5A in the *T. reesei* cellulases system used industrially. To begin, these three endoglucanases could be added to a commercial cellulase system lacking Eg2/Cel5A to see how they perform relative to the *T. reesei* Eg2/Cel5A.

4.5 Synergy analysis

In assessing synergy between each of the seven endoglucanases and the *T. reesei* Cbh1/Cel7A (Iogen Corp.), two substrates were used: Avicel, a microcrystalline cellulose and PASC, an amorphous cellulose. According to my results, the degree of synergism using Avicel is higher than the degree of synergism using PASC for all the endoglucanases tested in combination with the *T. reesei* Cbh1/Cel7A (Iogen Corp.). My results agree with results reported previously (Nidetzky et al., 1993). These authors found that the degree of synergism was higher on microcrystalline cellulose than amorphous cellulose. This is probably due to microcrystalline cellulose having relatively few accessible sites for endoglucanases and therefore exoglucanases can generate new sites for endoglucanase access (Teeri, 1997). On the other hand, endoglucanases and exoglucanases showed less cooperativity on amorphous cellulose (PASC) because endoglucanases have access to more sites on amorphous cellulose. Comparing the degree of synergism of different target endoglucanases on Avicel or PASC, generally, the results show that the endoglucanases with a CBM have a higher degree of synergism compared with the endoglucanases without a CBM. For example, AfumEgl2010 which has a CBM shows a high degree of synergism with the *T. reesei* Cbh1/Cel7A (Iogen Corp.) when using both PASC and Avicel. On the other hand, AnidEgl2020 which does not have a CBM shows the lowest degree of synergism under both conditions. The function of the CBM may explain the differences observed because the presence of a CBM can increase enzyme efficiency by helping the enzyme stay close to the substrate (Din et al., 1994). My results support this possibility since *T. reesei* Eg2/Cel5A and AfumEgl2010 had the highest degree of synergism, 3.3 and 2.9 respectively, and AnidEgl2020 had the lowest

degree of synergism, 1.3, when using Avicel. Since there are several factors, such as the amount of enzyme loaded, the reaction time, presence or absence of inhibitors, which can affect the degree of synergism, we can vary these factors to assess their effect on synergy (Zhang and Lynd, 2004). For example, we can add β-glucosidases into the synergy assays to see whether β-glucosidases can increase the degree of synergism. Also, we can use a greater number of concentrations when we vary the amount of each enzyme that was used in the synergy assays.

4.6 Future perspectives

Analysis of the signal peptides of functionally expressed endoglucanases in *A. niger* and other highly expressed secreted proteins could be informative in order to find the common signal peptide feature that support efficient secretion. Once the conserved features are identified, we can perhaps construct an optimized signal peptide. The optimized signal peptide could then be used to replace the native signal peptides of target proteins that are poorly secreted by *A. niger.*

Currently, the expressed endoglucanases in *A. niger* are not truly purified. There may be other secreted proteins present that may affect the results of initial screening and biochemical characterization. To solve this problem, a high throughput method for heterologous protein purification could be developed. Since the CBM can bind with the insoluble cellulose, it might be possible to fuse target proteins to a CBM and use it as a tag to purify heterologous proteins expressed in *A. niger*. Another approach that could be used to obtain pure proteins is to eliminate background enzyme activity from the culture filtrates. This could be done using mass spectrometry to identify the secreted proteins

produced by our expression host and then deleting their genes by targeted gene replacement.

REFERENCES

Albersheim P. 1975. The Walls of Growing Plants. Sci. Am. 232: 80-95.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215: 403-410.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 22: 195-201.

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

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340: 783-795.

Bernard P, Couturier M. 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol. 226: 735-745.

Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 14: 1023-1031.

Carrez D, Janssens W, Degrave P, van den Hondel CA, Kinghorn JR, Fiers W, Contreras R. 1990. Heterologous gene expression by filamentous fungi: secretion of human interleukin-6 by *Aspergillus nidulans*. Gene 94: 147–154.

Champe PC, Harvey RA, Ferrier DR. 2007. Lippincott's Illustrated Reviews: Biochemistry. Lippincott Williams & Wilkins. 4: ISBN:0781769604.

Crawford RL. 1981. Lignin biodegradation and transformation. Wiley (New York). ISBN 10: 0471057436

Davies GJ, Dauter M, Brzozowski AM, Bjørnvad ME, Andersen KV, Schülein M. 1998. Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 A and its cellobiose complex at 2.0 A resolution. Biochemistry. 37: 1926-1932.

Debets AJM, Bos CJ. 1986. Isolation of small protoplasts from *Aspergillus niger*. Fungal Genetics Newsletter 33: 24.

Decelle B. 2006. Cloning, functional expression and characterization of three *Phanerochaete Chrysosporium* endo-1, 4-beta-xylanases. Concordia University. Master thesis

Din N, Damude HG, Gilkes NR, Miller RC Jr, Warren RA, Kilburn DG. 1994. C1-Cx revisited: intramolecular synergism in a cellulase. Proc Natl Acad Sci U S A. 91: 11383-11387.

DOE/EIA-0383(2007). Annual energy outlook 2007 with projections to 2030. Report No. DOE/EIA-0383(2007).

DOE/EIA-0384(2008). Annual Energy Review 2008. Report No. DOE/EIA-0384(2008)

DOE/GO-102001-1102. Renewable Energy: An Overview. Energy Efficiency and Renewable Energy Clearinghouse (EREC) Brochure. Report No. DOE/GO-102001-1102.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19: 4008.

Doner LW, Irwin PL. 1992 Assay of reducing end-groups in oligosaccharide homologues with 2,2'-bicinchoninate. Anal Biochem. 202: 50-53.

Eun H-M. 1996. Enzymology Primer for recombinant DNA technology. ISBN: 0122437403 (academic press)

Fowler T, Berka RM, Ward M. 1990. Regulation of the glaA gene of *Aspergillus niger*. Curr. Genet. 18: 537-545.

Galbe M, Zacchi G. 2007. Pretreatment of lignocellulosic materials for efficient bioethanol production. Adv Biochem Eng/Biotechnol. 108: 41-65.

Grishutin SG, Gusakov AV, Markov AV, Ustinov BB, Semenova MV, Sinitsyn AP. 2004. Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. Biochim Biophys Acta. 1674: 268-281.

Gruno M, Väljamäe P, Pettersson G, Johansson G. 2004. Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. Biotechnol Bioeng. 86: 503-511.

Hata Y, Kitamoto K, Gomi K, Kumagai C, Tamura G. 1992. Functional elements of the promoter region of the *Aspergillus oryzae* glaA gene encoding glucoamylase. Curr. Genet. 22: 85–91.

Henrissat B, Driguez H, Viet C, Schülein M. 1985. Synergism of Cellùlases from *Trichoderma reesei* in the Degradation of Cellulose. Bio/Technology 3: 722-726.

Jeenes DJ, Marczinke B, MacKenzie DA, Archer DB. 1993. A truncated glucoamylase gene fusion for heterologous protein secretion from *Aspergillus niger*. FEMS Microbiol. Lett. 107: 267–271.

Käfer E. 1977. Meiotic and mitotic recombination in *Aspergillus* and its chromosomal aberrations. Adv. Genet. 19: 33-131.

Karlsson J, Siika-aho M, Tenkanen M, Tjerneld F. 2002. Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. J Biotechnol. 99: 63-78.

Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009 The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 37: D387-392.

Kleywegt GJ, Zou JY, Divne C, Davies GJ, Sinning I, Stahlberg J, Reinikainen T, Srisodsuk M, Teeri TT, Jones TA. 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. J. Mol. Biol. 272: 383-397.

Lamed R, Kenig R, Morag E, Calzada JF, de Micheo F, Bayer EA. 1991. Efficient cellulose solubilization by a combined cellulosomes-β- glucosidase system. Appl Biochem Biotechnol 27: 173–183.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. ClustalW and ClustalX version 2. Bioinformatics. 23: 2947-2948.

Lo Leggio L, Larsen S. 2002. The 1.62 A structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. FEBS Lett. 523: 103-108.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev. 66: 506-577.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH. 2009. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 37: D205-210.

Master ER, Zheng Y, Storms R, Tsang A, Powlowski J. 2008. A xyloglucan-specific family 12 glycosyl hydrolase from *Aspergillus niger*: recombinant expression, purification and characterization. Biochem J. 411: 161-170.

Miki T, Park JA, Nagao K, Murayama N, Horiuchi T. 1992. Control of segregation of chromosomal DNA by sex factor F in Escherichia coli. Mutants of DNA gyrase subunit A suppress letD (ccdB) product growth inhibition. J Mol Biol. 225: 39-52.

Nidetzky B, Hayn M, Macarron R, Steiner W. 1993. Synergism of *Trichoderma reesei* cellulases while degrading different celluloses. Biotechno Lett. 15: 71-76.

Peitsch MC. 1995. Protein modeling by E-mail. Bio/Technology. 13: 658-660.

Percival Zhang YH, Himmel ME, Mielenz JR. 2006 Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv. 24: 452-481.

Purves WK, Orians GH, Heller HC, Sadava D. 1998. Life, the science of biology. Sinauer Associates, Inc, W.H.Freeman and Company. 5: ISBN:0716728699.

Qin Y, Wei X, Liu X, Wang T, Qu Y. 2008. Purification and characterization of recombinant endoglucanase of *Trichoderma reesei* expressed in *Saccharomyces cerevisiae* with higher glycosylation and stability. Protein Expr Purif. 58: 162-167.

Rose JKC. 2003. The plant cell wall. Annual Plant reviews. CRC press. Blackwell publishing. Oxford. 8: ISBN: 084932811X

Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol. 30: 279–291.

Saloheimo M, Lehtovaara P, Penttilä M, Teeri TT, Ståhlberg J, Johansson G, Pettersson G, Claeyssens M, Tomme P, Knowles JK. 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. Gene. 63: 11-22.

Sandgren M, Shaw A, Ropp TH, Wu S, Bott R, Cameron AD, Stahlberg J, Mitchinson C, Jones TA. 2001. The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 A resolution. J. Mol. Biol. 308: 295-310.

Santerre Henriksen AL, Even S, Muller C, Punt PJ, van den Hondel CA, Nielsen J. 1999. Study of the glucoamylase promoter in *Aspergillus niger* using green fluorescent protein. Microbiology. 145: 729–734.

Semova N, Storms R, John T, Gaudet P, Ulycznyj P, Min XJ, Sun J, Butler G, Tsang A. 2006. Generation, annotation, and analysis of an extensive *Aspergillus niger* EST collection. BMC Microbiol. 6: 7.

Storms R, Zheng Y, Li H, Sillaots S, Martinez-Perez A, Tsang A. 2005. Plasmid vectors for protein production, gene expression and molecular manipulations in *Aspergillus niger*. Plasmid. 53: 191-204.

Sun Y, Cheng J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour Technol. 83: 1-11.

Teeri TT. 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol. 15: 160–167.

Tomme P, Van Tilbeurgh H, Pettersson G, Van Damme J, Vandekerckhove J, Knowles J, Teeri T, Claeyssens M. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur J Biochem 170: 575-581.

Tsang A, Powlowski J, Butter G, Storms R, Mohrman L, Musaers JHGM. 2008. Novel xylanase enzymes XYL001 and XYL002 and uses thereof. Patent number WO 2008/074884.

Updegraff DM. 1969. Semimicro determination of cellulose in biological materials. Anal Biochem. 32: 420-424.

Urbanowicz BR, Bennett AB, Del Campillo E, Catalá C, Hayashi T, Henrissat B, Höfte H, McQueen-Mason SJ, Patterson SE, Shoseyov O, Teeri TT, Rose JK. 2007. Structural organization and a standardized nomenclature for plant endo-1,4-beta-glucanases (cellulases) of glycosyl hydrolase family 9. Plant Physiol. 144: 1693-1696.

Van Petegem F, Vandenberghe I, Bhat MK, Van Beeumen J. 2002. Atomic resolution structure of the major endoglucanase from *Thermoascus aurantiacus*. Biochem Biophys Res Commun. 296: 161-166.

Verdoes JC, Punt PJ, Schrickx JM, van Verseveld HW, Stouthamer AH, van den Hondel CA. 1993. Glucoamylase overexpression in *Aspergillus niger* : molecular genetic analysis of strains containing multiple copies of the glaA gene. Transgenic Res. 2: 84–92.

Wang M, Wu M, Huo H. 2007. Life-cycle energy and greenhouse gas emission impacts of different corn ethanol plant types. Environ. Res. Lett. 2: 024001

Wernars K, Goosen T, Wennekes BM, Swart K, van den Hondel CA, van den Broek HW. 1987. Cotransformation of *Aspergillus nidulans*: a tool for replacing fungal genes. Mol Gen Genet. 209: 71-77.

Withers JM, Swift RJ, Wiebe MG, Robson GD, Punt PJ, van den Hondel CA, Trinci AP. 1998. Optimization and stability of glucoamylase production by recombinant strains of *Aspergillus niger* in chemostat culture. Biotechnol. Bioeng. 59: 407–418.

Wood PJ, Erfle JD, Teather RM. 1988. Use of complex formation between Congo Red and polysaccharides in detection and assay of polysaccharide hydrolases. Methods Enzymol. 160: 59–74.

Wood TM. 1988. Preparation of crystalline, amorphous, and dyed cellulase substrates. Methods Enzymol. 160: 19-25

Woodward J. 1991. Synergism in cellulase systems. Biores. Technol. 36(1): 67-75

Xiao Z, Storms R, Tsang A. 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol Bioeng. 88: 832-837.

Zhang YH, Lynd LR. 2004 Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng. 88: 797-824.

Zorov IN, Dubasova MY, Sinitsyn AP, Gusakov AV, Mitchenko AA, Baraznenok VA, Gutyerres B, Popova NN. 1997. Application of the bicinchoninic method of the assay for the reducing sugars to determine carboxymethylcellulase activity of cellulases using a microplate reader. Biochemistry (Mosc.). 62: 704-709.