Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Biomass to biofuel: Towards the bioengineering of *Saccharomyces* species for cellulose degradation.

Thesis submitted for the Degree of Doctor of Philosophy

Trinity College Dublin.

(2011)

By James Fitzpatrick

Declaration

I, James Fitzpatrick declare that this thesis has not been submitted as an exercise for a degree at this or any other University. The experimentation and work recorded herein represents my own work. I hereby agree that the Library may lend or copy the thesis upon request. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

James Fitzpatrick

James Fitzpatrick.

Summary:

Cellulose is the most abundant polysaccharide on earth and therefore represents a major reservoir of sugar that could be potentially converted to alcohol and used as a fuel source. The filamentous fungus *Trichoderma reesei* possesses an array of enzymes which can degrade cellulose to glucose. These enzymes are known as cellulases. The aim of this project was to combine the fermentative capacity of *Saccharomyces* yeast species with the cellulolytic ability of the *T. reesei* to simultaneously saccharify and ferment a cellulose substrate into ethanol. Three cellulolytic genes of *T. reesei egl1*, *cbh2* and *bgl1* were cloned separately into the haploid *S. cerevisiae* strain S150 and also into the polyploid *S. pastorianus* strain C10-51. Experiments were carried out in parallel to created cellulase encoding cassettes of *egl1*, *cbh2* and *bgl1* using either *T. reesei* genomic DNA (gDNA) or complementary DNA (cDNA). The gDNA inserts were amplified by polymerase chain reaction (PCR) using gene specific oligonucleotide primers. The cDNA inserts were synthesised from *T. reesei* RNA by reverse-transcription PCR (RT-PCR). The aim of inserting *T. reesei* genomic DNA genes into yeast was to investigate if the host *Saccharomyces* splicing machinery could recognise and correctly remove foreign introns to enable gene expression. The cellulase genes generated from *T. reesei* nucleic acids were transferred into the target organism via *in vivo* homologous recombination cloning using the pGREG vector. The *S. cerevisiae* transformants were selected by a *URA3* auxotrophic marker, whereas *S. pastorianus* clones were selected by antibiotic resistance to the kanMX cassette present on the pGREG vector. The pGREG vectors contained the powerful but regulated galactokinase, GAL1, promoter. In order for the recombinant yeast to utilize and ferment cellulose, the inserted cellulase encoding genes must be constantly producing cellulase enzymes. The GAL1 promoter of pGREG was replaced with the constitutively expressed phosphoglycerate kinase, PGK1, promoter of *S. cerevisiae*.

Functional cellulase activity was observed only in the yeast clones containing *T. reesei* cDNA inserts. This suggested that *Saccharomyces* yeast could not remove the introns within *T. reesei* genomic DNA inserts. Cellulase activity was detected in the growth culture supernatant using enzyme assays.

The resultant cellulase producing yeast strains were co-cultured with 100 g l^{-1} phosphoric acid swollen cellulose (PASC) at 25°C for 10 days. The *S. cerevisiae* S150 co-culture generated a maximum ethanol yield of approximately 8.15 mg l^{-1}. This was statistically significant compared to the control non-cellulolytic S150 culture. The *S. pastorianus* co-culture yielded 1.77 mg l^{-1} and was not statistically significant.

This thesis reports that co-culturing *S. cerevisiae* strains expressing individually *T. reesei* proteins EGI, CBHII or BGLI could be used to produce ethanol from a purely cellulose substrate. This is the first report of bioengineering *S. pastorianus* yeast to secrete recombinant cellulase enzymes. The *S. pastorianus* co-culture was not statistically significant, however optimizing the reaction conditions and strain selection could increase ethanol content from future fermentations. *S. pastorianus* is the dominant yeast strain used in industrial brewery. Its potential use in the bioethanol industry is significant due to its stress-tolerance.

Acknowledgements

Firstly, I would like to thank Dr Ursula Bond for allowing me the opportunity to join her laboratory and carry out research for a most interesting project. I am indebted to Dr Bond for all of her helpful advice and guidance throughout the project.

I also owe a debt of gratitude to Dr. Tharappel C James (Jimmy) for helping me with all my practical and theoretical issues I encountered in the lab.

Thank you to Dr Tadhg O Croinin, formerly of the Dorman lab, for the gift of an aliquot of HIS tagged FIS protein.

A huge thank you to all the members of the Bond lab present and past. I am grateful and thankful for the help of Dr Jane Usher who assisted me in settling in the lab and getting started on my project.

A special word of appreciation goes to my fellow PhD alumni Suzanne and Christina. The PhD process is lonely and tough; you have the good day followed by the bad months. Thank you to Sue and Chris for being there with me to celebrate the good day and share the burden to help me through the bad months. I wish you all the best in your future stellar careers!

To Joanne, I wish you good luck with the rest of your work! To Will, the newest member of the Bond lab, I wish you best of luck with the project and I hope this thesis may serve as a guide to your work.

Thank you to all in the Dorman lab, you provided the soundtrack of my time in the bunker! Thank you to everyone else in the Moyne for their assistance and in particular a huge thank you is owed to the members of the Prep-room.

I would like to thank my funding body the Environmental Protection Agency (EPA).

Finally I must thank my family and girlfriend Naomi. Simply put, without their unending support, belief and love I would not have reached this point.

List of conferences and publications

Irish Fungal Meeting: June 2007 NUI Maynooth: poster presentation

EPA STRIVE conference: November 2007: poster presentation

British Yeast Group meeting: March 2008: NUI Maynooth: poster presentation

Irish Fungal Meeting: June 2008: NUIGalway: poster presentation

Society General Microbiology: September 2008: poster presentation: Trinity College Dublin

EPA STRIVE conference: November 2009: poster

Irish Fungal Meeting: June 2009: UCD: poster

EPA STRIVE conference: December 2009: oral presentation

Posters and talk all shared the same title:

Biomass to Biofuel: Generation of cellulose-based biomass degrading strains of brewery yeast.

Abbreviations

ADH	Alcohol dehydrogenase
BGL	β-glucosidase enzyme
bgl1	Gene encoding β-glucosidase
CBH	Cellobiohydrolase enzyme
cbh	Gene encoding cellobiohydrolase
CBM	Carbohydrate-binding module
CD	Catalytic domain
CMC	Carboxymethyl cellulose
DP	Degree of polymerisation
EG	Endoglucanase
egl	Gene encoding endoglucanase
ER	Endoplasmic reticulum
GAL1	Galactokinase 1
GH	Glycoside hydrolase
GHG	Greenhouse gas
HIS	Histidine
PASC	Phosphoric acid swollen cellulose
PCR	Polymerase chain reaction
PGK1	Phoshoglycerate kinase
SC	Synthetic complete
SSF	Simultaneous saccharification and fermentation
URA	Uracil
X-glu	X-glucoside
YPD	Yeast extract peptone dextrose
WT	Wildtype

List of Figures

List of Tables

Table of contents

Chapter 1: General Introduction

1.1 Origins of biofuels

The world is entering the post-fossil fuel era. Alternative energy sources are now needed to power industry and transportation of the future. The realisation of climate change means that any new fuel source must be environmentally sustainable. One such green energy source is biofuels. The term biofuel is a catch-all phrase that describes any fuel that is derived from recently living natural biological sources. The two most common examples of biofuels are bio-ethanol and bio-diesel. Ethanol is one of the oldest products obtained by traditional biotechnology. Although largely used as a beverage or food preserver, ethanol has long been used as a fuel source. Indeed, Henry Ford in the 1880s originally designed the model T car to run on "farm ethanol". However, the relative abundance and cheapness of fossil fuel-derived petroleum overtook the use of ethanol as a fuel source (Zaldivar et al., 2001). It was the advent of the oil crises of the 1970's that renewed interest in alternative energies. Countries such as Brazil and the USA took the lead in biofuel production back in the 1970's (Gray et al., 2006). Recent figures show that total world ethanol production was 41 billion litres (Carere et al., 2008), with Brazil and the USA accounting for 37% and 33% of global production levels, respectively. Of all the total ethanol produced worldwide each year, over 66% is used for fuel, the reminder is used by the beverage industry.

1.2. Benefits of biofuel usage

Biofuels are considered the cleanest liquid fuel alternative to fossil fuel (Lin and Tanaka, 2006). The incomplete combustion of fossil fuels releases toxic gases e.g. sulphur dioxide, carbon dioxide (CO_2) and methane (CH_4) into the environment. These so called Greenhouse gases (GHG) contribute to global warming. The GHG create a layer in the earth's atmosphere which can prevent heat reflected off the Earth's surface from escaping into Space. Approximately 60% of greenhouse gases are attributed to anthropogenic sources. The combustion of fossil fuels for transportation is by far the largest contributer to the increase in CO_2 in the atmosphere (Carere et al., 2008). The US consumes approximately 20 million barrels of crude oil daily, 70% of which is used in transportation (Gray et al., 2006). In the European Union, 28% of all energy consumed is for transportation, of which 80% is due to road transportation (Antoni et al., 2007). A reduction in the use of fossil fuels in transportation

could significantly reduce GHG emissions. Brown et al., (1998) concluded that replacing fossils fuels with bioethanol could decrease CO_2 emissions by 60-90%. Biofuels are an attractive alternative to fossil fuels because they can be used as transportation fuels with little change to existing technology (Carere et al., 2008). There has been an increase in the production of flexi-fuel vehicles (FFV). These vehicles contain engines that can accept a variety of different fuels be it fossil fuel derived petrol or bioethanol. In Ireland, bioethanol usage has been lead by Maxol (www.maxol.ie). Motorists can avail of E5 or E85 mixes; these consist of petrol blended with 5 or 85% bioethanol respectively. The bioethanol is produced by the Carbery Group based in Cork. Whey and by-products from the cheese industry are fermented to ethanol.

The CO_2 released by consumption of bioethanol is recycled by photosynthetic systems of plant biomass and hence leading to a sustainable virtually carbon neutral cycle (Brown et al., 1998) (Fig 1.1).

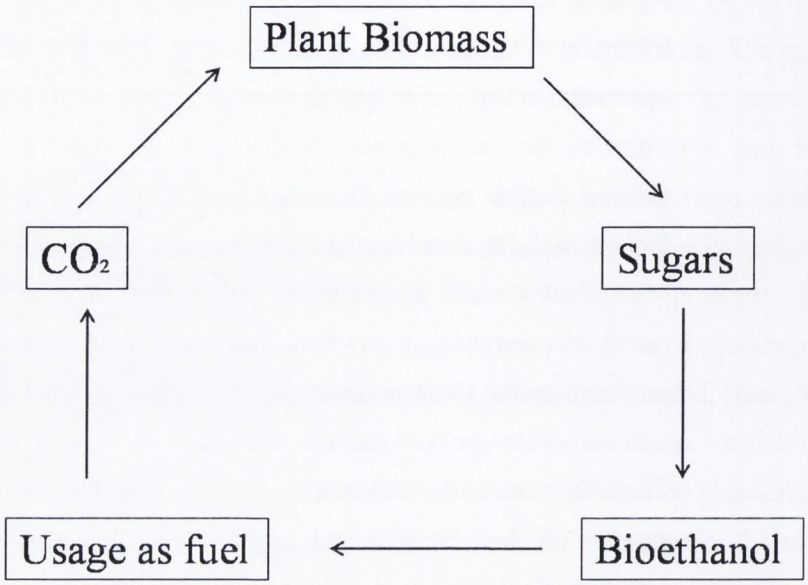


FIG. 1.1. Schematic overview of biomass to biofuel carbon cycle

1.3. Legal mandates for biofuel usage

The Kyoto protocol mandates the developed world to reduce greenhouse gas emissions. The EU agreed to reduce its GHG emissions by 8% below 1990 levels during the period 2008-2012. Ireland was required to limit its annual total national GHG emissions to 62.8 Mtonnes of CO_2 emissions. The Environmental Protection Agency (EPA) monitors Ireland's GHG emissions annually; in 2006 it reported that Ireland's GHG emissions were 69.77 million tonnes (www.epa.ie). In recent years Ireland's GHG emissions have begun to decrease, the latest report by the EPA predicts that Irelands GHG emissions will be 63 Mtonnes of CO_2 per annum between 2008-2012 (www.epa.ie). The decrease in GHG, coupled with carbon credits purchased by the government will ensure that Ireland will meet its Kyoto commitment. However, during the 2008-2012 period, GHG emissions for the Irish transport sector will increase by 25% to 17.8 Mtonnes of CO_2 per annum. A factor in this increase in emissions could be due to the continued low levels of biofuel penetration into the transportation fuel sector. The EC Biofuels directive 2003/30/EC (http://eur-lex.europa.eu/) requires member states to replace 5.75% of fossil fuels with biofuels by 2010 and 10% by 2020. In 2007 the Irish government adopted the EC directive in the Energy White Paper entitled "Delivering a sustainable energy future for Ireland". The monitoring agency Sustainable Energy Agency of Ireland (SEAI) reported that biofuels currently accounted for 2% of Ireland's total energy usage (www.seai.ie). The latest SEAI report entitled "Energy forecast for Ireland to 2020" predicts that biofuel usage will increase 10.1% annually between 2008-2020 (www.seai.ie). The SEAI report forecasts that by 2020; biofuels will account for 3% of Ireland's transportation fuels. In 2009 the EU Parliament and Council introduced the "EU Climate Change Package" proposals to increase the share of renewable energies to 20% of the total level of energy consumed within the EU (406/2009/EC) (http://eur-lex.europa.eu/). Based on current trends, Ireland will not meet its biofuel usage requirements.

1.4. Substrate for bioethanol

The production of biofuels can be characterised into different generations based upon fermentation strategies and biotechnology developments (Fig 1.2).

The main substrates used for bioethanol production in Brazil and the USA are sugarcane and corn respectively. The major carbohydrate in sugar cane is

sucrose and in corn, is starch. These carbohydrates are readily fermentable into ethanol by microorganisms such as the yeast *Saccharomyces cerevisiae* and represent the 1st generation of biofuels (Fig. 1.2.A). A disadvantage to using these substrates as energy sources is the fact that they are food crops and hence controversy regarding the ethics of exploiting "food for fuel" surrounds the established bioethanol industry. Over 90% of the world's bioethanol is produced from food crops (Carere et al., 2008). Other disadvantages include limits of supply of these crops and also that not every climate is suited to their growth. A cheaper more readily available substrate that could make biofuel economically viable is biomass.

Biomass can be grouped into roughly four categories, wood residues, municipal solid waste, agricultural waste and dedicated energy crops (Lin and Tanaka 2006). Biomass is not readily fermentable. Expensive pre-treatments are required to increase access to the sugars within the biomass. Dilute acid pretreatment is extensively used in bioethanol production (Gray et al., 2006). The sugars released from pretreatment must be further hydrolyzed via enzymatic actions to yield fermentable glucose. The glucose is fermented by yeast into ethanol which is recovered and can be used for energy. Currently generating biomass to bioethanol uses a process called separate hydrolysis and fermentation (SHF). This process gives rise to the 2nd generation biofuels (Fig 1.2.B). The enzymes required to degrade the structure of biomass to glucose are not found naturally in the fermenting organism *S. cerevisiae* and must be supplied *ex vivo*. Therefore with this process different micro-organisms are required to convert biomass into bioethanol (Fig 1.2.B). Ideally, it would be more economically efficient if one microbe could simultaneously saccharify and ferment (SSF) biomass ethanol in one reaction (Fig 1.2.C). This 3rd generation of fuels is known as consolidated bioprocessing (CBP).

An alternative description exists for the characterising of biofuels in terms of feedstocks. The 1st generation feedstocks include food crops containing reservoirs of readily fermentable sugars. The 2nd generation feedstocks comprise cellulosic agricultural waste by-products. The sugars within these substrates are not readily fermentable and require physical and chemical pre-treatment. Algae, which contain approximately 20-30% cellulose and other sugars, have been suggested as the 3rd generation biofuel feedstock due to its abundance (Lee et al., 2011).

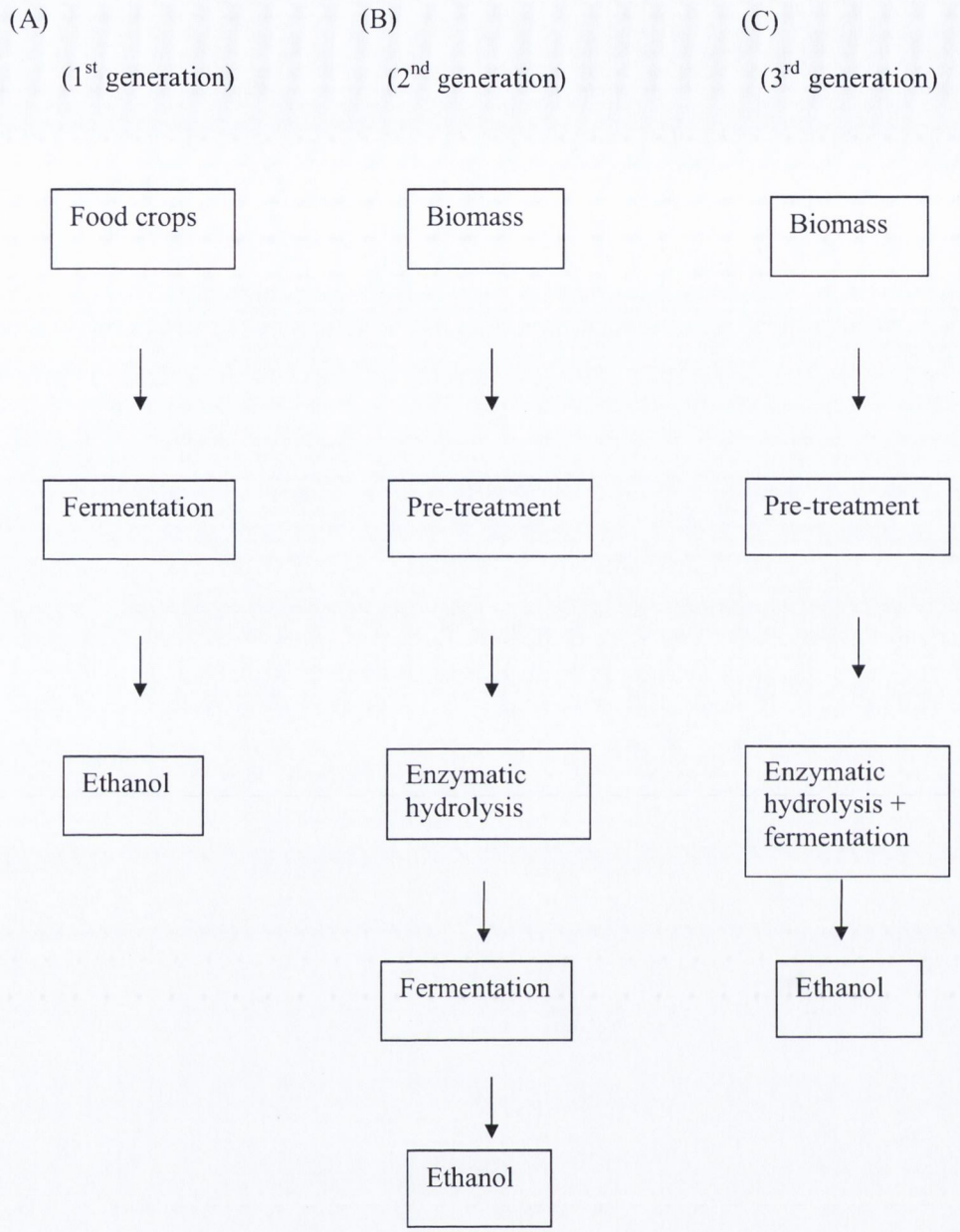


FIG 1.2. Schematic flow diagram of production of bioethanol [a]

(A) 1st generation biofuels. Using food crops such as sugar cane as a substrate

(B) 2nd generation biofuels. Current strategy of biomass to bioethanol using separate hydrolysis and fermentation (SHF)

(C) 3rd generation biofuels. Potential efficient strategy of consolidated bioprocessing (CBP)

[a] Characterisation of generations based upon fermentation strategies and biotechnology developments.

1.5. Lignocellulose composition

Biomass is known more correctly as lignocellulose. Lignocellulose is the main structural component of biomass comprising between 50 to 90% of all plant matter (Lin and Tanaka, 2006; Sanchez, 2009). The global production of plant biomass amounts to approximately $2x10^{11}$ Mt per annum, of which between 8-20 x 10^9 is potentially accessible for processing (Lin and Tanaka, 2006). Thus lignocellulose represents the most abundant renewable source for fuel. Lignocellulose is composed of three major components; cellulose, hemicellulose and lignin. The relative amounts of each component vary with different plant types; on average the composition is cellulose 30-50%, hemicellulose 20-30% and lignin 15-25% (Gray et al., 2006). The typical lignocellulose composition from three groups of plants namely, hardwoods (Hybrid poplar), softwoods (Pine) and grasses (Switchgrass) that could be used as a substrate for biofuel production is shown in Table 1.1 (Hamelinck et al., 2005).

Cellulose is the chemically simplest component of lignocellulose and is the most abundant polysaccharide on Earth (Bayer et al., 1998). The structure of cellulose is complex and still not fully understood (Arantes and Saddler, 2010). In its simplest chemical form, cellulose is a linear polymer of D-glucose monomers linked by β-1,4-glycosidic bonds (Sanchez, 2009). Such polysaccharides are termed glucans. The average degree of polymerisation (DP; the average number of glucose residues per molecule) of natural cellulose is 3500 (Cullen and Kersten, 1992). The basic repeating subunit of cellulose is the glucose dimer cellobiose (Zaldivar et al., 2001) (Fig 1.3.A.). At the macroscopic level, cellulose exists as two distinct forms, tightly packed crystalline and non-organised amorphous regions. Furthermore, crystalline cellulose can exist in several allomorphs. The crystalline region of natural cellulose is almost exclusively the allomorph cellulose I, which itself has different subtypes such as cellulose Iα and Iβ (Weimer et al., 1991). Plant celluloses are primarily Iβ and can be irreversibly converted into alternative crystalline allomorphs, namely cellulose II, III or IV (Weimer et al., 1991) by chemical treatment of lignocellulose with strong alkali or ammonia. The amorphous regions of cellulose are believed to be a result of surface shaving caused by natural erosion. At the nanoscale, the amorphous or crystalline

forms are made up of cellulosic fibres which are made up of microfibrils (Fig. 1.3.B). At the molecular level, microfibrils are composed of approximately 30 β-glucan chains (Arantes and Saddler, 2010). The microfibrils are formed by the coupling of adjacent cellulose chains by a combination of hydrogen bonds, hydrophobic interactions and van der Waals forces that leads to the parallel alignment of the crystalline structure and increase in tensile strength (Dashtban et al., 2009) (See Fig 1.3.B). In higher plants microfibrils, the dominant structural feature of cellulose, are cross-linked by other cell wall components such as xyloglucans. The chain ends of crystalline cellulose have one of two orientations either reducing or non-reducing. The end of a chain containing a free anomeric carbon is called the reducing end; if the anomeric carbon is not exposed it is a non-reducing end (Fig 1.3.A).

Hemicellulose is the second most abundant polysaccharide within lignocelluloses (See Table 1.1). Hemicellulose has a lower molecular weight than cellulose (Sanchez, 2009). It is a highly branched heteropolymer composed of pentoses, such as xylose and arabinose, hexoses, mainly mannose, glucose and galactose and sugar acids. The composition of hemicellulose is variable in nature and depends upon the plant source (Dashtban et al., 2009) (See Table 1.1).

The third main component of lignocellulose is lignin. The polymer is the condensation product of three aromatic alcohols; coniferyl alcohol, p-coumaryl alcohol and sinapyl alcohol. Lignin is the most recalcitrant component of lignocellulose to degrade (Dashtban et al., 2009). Lignin links both hemicelluloses and cellulose together forming a physical barrier in the plant cell wall (Sanchez, 2009). The more lignin present the more difficult the plant will be to degrade. Lignin is resistant to most microbial attacks and oxidative stress. It is the most abundant aromatic polymer in nature but unlike cellulose and hemicellose cannot be fermented. The reason for the latter is due to the phenolic compounds that lignin is composed of, inhibit fermentation.

Table 1.1. Typical composition of lignocelluloses biomass for three types of plant[a]

Component	**Main sugar**	**Hardwood[b]**	**Softwood[c]**	**Grass[d]**
Cellulose	Glucose	44.70	44.55	31.98
Hemicellulose	Xylose	14.56	6.30	21.09
	Arabinose	0.82	1.60	2.84
	Galactose	0.97	2.56	0.95
	Mannose	2.20	11.43	0.30
Lignin	Phenolic compounds	26.44	27.67	18.33

[a] Table adapted from Hamelinck et al., (2005)

[b] Hybrid poplar

[c] Pine

[d] Switchgrass

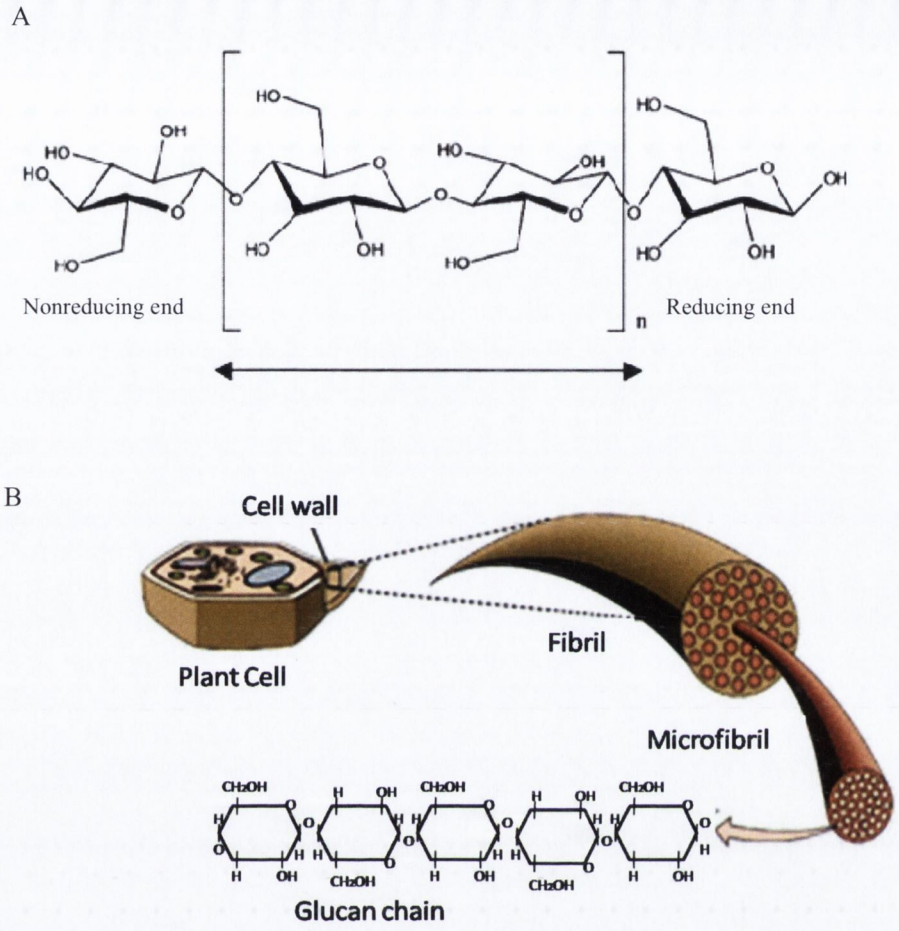


FIG.1.3. Structure of cellulose

(A) Cellobiose repeating unit of cellulose. The disaccharide consists of two glucose molecules linked by a β-1,4 glycosidic bond. The two orientations of chain ends are indicated.

(B) Overview of cellulose from typical plant cell wall to glucan chains. (Adapted from www.mhhe.com)

1.6. Cellulase enzymes

Cellulases are enzymes that aid in the hydrolysis of cellulose to glucose. Cellulases belong to the O-glycoside hydrolases (Miettinen-Oinonen, 2004). The glycoside hydrolases (GHs) are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (Miettinen-Oinonen, 2004). Cellulases are distinct from other glycoside hydrolases by their ability to cleave β-1,4-glucosidic bonds between glucosyl residues (Lynd et al., 2002). Cellulases are not found naturally in yeast; however, *S. cerevisiae* does produce endogenous saccharolytic enzymes (Lynd et al., 2002). These enzymes are β-1,3-glucanases and are active against β-1,3 glucan. β-1,3-Glucan is the main structural polysaccharide responsible for rigidity of the yeast cell wall. It was suggested that endogenous yeast β-1,3-glucanases play a role in controlled autolysis of cells (Lynd et al., 2002). Strict β-1,3-glucanases are inactive toward β-1,4-linked glucans and hence cannot hydrolyse cellulose (Van Rensburg et al., 1998).

There are three types of enzymes within the cellulase class; endoglucanases (EG), cellobiohydrolases (CBH) or exoglucanases and β-glucosidases (BGL). The insolubility of cellulose prevents it from being directly transported into the cell for digestion. This means that the hydrolysis of cellulose must occur outside the cell. For microorganisms to degrade and metabolise cellulose they must produce extracellular cellulases. There are two different cellulase systems; complexed and non-complexed.

1.6.1. Complexed systems

Microbes using complexed cellulase systems are mostly found in anaerobic environments (Lynd et al., 2002). Anaerobic bacteria are the primary organisms that use the complexed system; systems from the species of the *Clostridium* and *Ruminococcus* genera have been studied in great detail. Cellulases within the complexed system are produced as protuberances on the cell wall of the cellulolytic bacterium (Lynd et al., 2002). All the different types of cellulases are anchored to the bacterial cell well and not secreted out into the extracellular media. These protuberances are better known as the cellulosome. Cellulosomes are firmly bound to the cell wall but also have sufficient flexibility to interact with cellulose structures. The architecture of

the cellulosome is conserved amongst different cellulolytic bacterial species; however, the cellulase ratio may differ depending on the substrate. The cellulosome consists of a large non-catalytic scaffoldin protein. The scaffoldin is anchored to the bacterial cell wall by type II cohesion domains (Lynd et al., 2002). The cellulosome is composed of 22 saccharolytic proteins; at least nine display endoglucanase activity, four exoglucanases, and five hemicellulases amongst other enzymes. Cellulosomes are remarkably stable considering their size can range from 2-16 MDa. Electron microscopy studies have indicated that the cellulosome resembles a "fist"-like structure that opens out when in contact with cellulose, allowing the local spread of catalytic domains. A benefit of complexed cellulase systems is that the physical entity of the cellulosome minimises the distance over which the products of cellulose hydrolysis must diffuse in order to be taken up by the bacterial cell, which results in more efficient uptake.

1.6.2. Non-complexed systems

Non-complexed cellulase systems have been better characterised than the complexed system. Filamentous fungi and certain actinomycete bacteria such as *Cellulomonas* are known to use non-complexed cellulase systems. Non-complexed cellulases are "free" cellulases, i.e. they are secreted into the extracellular environment and not attached to the cell surface. Filamentous fungi have the ability to physically penetrate cellulosic structures through hyphal extensions. This permits the organism to secrete cellulases within cavities of the cellulose. The production of "free" cellulases would be sufficient to degrade cellulose under these conditions. Anaerobic bacteria lack the physical ability to penetrate cellulose and therefore must seek an alternative strategy, hence the complexed system (Lynd et al., 2002). The most extensively studied cellulolytic organism that uses the non-complexed system is the filamentous fungus *Trichoderma reesei*; cellulases from this organism were utilized in this project.

1.6.3 *Trichoderma reesei* cellulases:

The filamentous fungus *T. reesei* was originally known as *T. viride* QM6a. However, by the 1960s the unique biological nature of "*T. viride*" QM6a was recognised and the organism was renamed *T. reesei* in honour of its major investigator Elwyn T. Reese. As a result of his pioneering research on *T. reesei*'s biosynthesis, structure and mechanism of cellulose degradation, *T.*

reesei is today the leading cellulolytic organism used in industry. *T. reesei* is a mesophilic soft-rot ascomycete fungus and is an anamorph of the pantropical ascomycete *Hypocrea jecorina* (Martinez et al., 2008). *T. reesei* shares its most recent common ancestry with *Fusarium graminearum*, anamorph *Gibberella zeae* (See Fig 1.4). The filamentous ascomycetes and budding yeasts diverged from one another approximately 900-1000 million years ago (Cornell et al., 2007). *T. reesei* synthesizes an array of cellulases, including at least five EGs, two CBHs and two BGLs (Foreman et al., 2003). Interestingly, the sequencing of the *T. reesei* genome revealed that the fungus encodes fewer cellulases than other fungi capable of hydrolysing plant cell wall polysaccharides such as *Aspergillus* species (Martinez et al., 2008) (See Table 1.2). Experiments in this project focused primarily on *T. reesei* enzymes EGI, BGLI, CBHII and briefly CBHI.

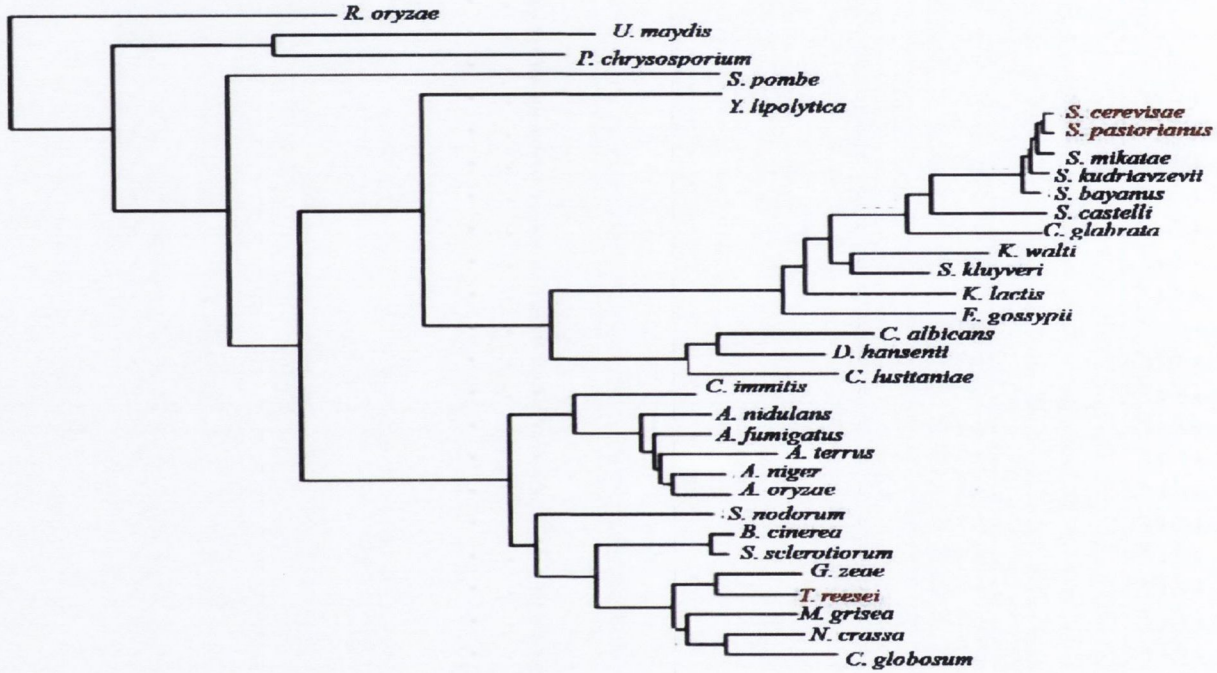


FIG 1.4. Species tree of various fungi. Locations of *S. cerevisiae, S. pastorianus and T. reesei* are indicated in red. Figure adapted from Cornell et al., (2007)

Table 1.2. Number of copies of cellulolytic enzymes in *T. reesei* and several species of *Aspergillus*.

Organism	**EGI**	**EGII**	**EGIII**	**EGIV**	**EGV**	**CBHI**	**CBHII**	**BGLI**	**BGLII**	**BGLIII**	**BGLV**
T. reesei	1	2	1	3	1	1	1	1	1	0	0
A. fumigatus	2	3	3	7	1	2	1	1	0	0	0
A. niger	1	2	1	9	1	2	2	2	0	1	0
A. orzyae	1	2	2	8	0	2	1	1	0	1	1

Table adapted from Martinez et al., (2008).
Additional information taken from Enzyme Database-BRENDA (www.brenda-enzymes.info)

1.7. Classification of cellulases

The O-glycosyl hydrolases (GH) are a group of enzymes that hydrolyse the glycosidic bond between two or more monosaccharides (Henrissat et al., 1998). Many GH are produced by microorganisms for degradation of polysaccharides in plant cell walls. Cellulases are GHs. In the past, GHs were assigned names based on the International Union of Biochemistry and Molecular Biology (IUB-MB) enzyme nomenclature (EC number) in the order of their discovery (Henrissat et al., 1998). Endoglucanase, cellobiohydrolase and beta-glucosidase were identified as EC 3.2.1.4, EC 3.2.1.91 and EC 3.2.1.21 respectively. The IUB-MB enzyme nomenclature of GHs is based on their substrate specificity. An alternative classification was established based on amino acid sequence similarity. This system was more appropriate for characterisation of each GH protein because it reflects the 3-D structure of the protein. Comparison of amino acid sequence reveals evolutionary relationships between gene families. GHs are now differentiated by hydrophobic cluster analysis (HCA) of their catalytic domain (CD). Amino acid sequence alignment of the catalytic domain enables prediction of key residues involved in hydrolysis. However this classification scheme cannot distinguish between endo or exo-glucanases. EGI, CBHI and CBHII are reclassified as Cel7B, Cel7A and Cel6A respectively, where Cel denotes cellulolytic activity, the number indicates the GH family that the protein belongs to and the letter after the family number corresponds to the order in which the enzymes were first discovered (Henrissat et al., 1998). Using the classification system based on HCA of the CDs, 2500 GHs have been separated into over 100 families. Cellulases are found in at least 13 of these families, indicating the relative conservative nature of cellulases. BGLI belongs to GH family 3, CBHII belongs to GH family 6 and both EGI and CBHI belong to GH family 7.

For clarity the *T. reesei* nomenclature of cellulase enzymes will be retained for the rest of this thesis. Accordingly, the name of gene is designated by italicised lowercase text and protein is represented by uppercase text; for example the genes *egl1*, *cbh1*, *cbh2* and *bgl1* encode for the proteins EGI, CBHI, CBHII and BGLI, respectively.

1.8. Cellulase structure

Cellulases have modular structures. All have an N-terminal secretory signal and a catalytic domain (CD) where the enzyme active site is located. Most

cellulases also possess an independently folded carbohydrate binding module (CBM) that is connected to the CD by a flexible linker usually rich in serine and threonine amino acids and believed to be highly O-glycosylated (Palamarczyk et al., 1998). A schematic diagram of the basic gene structure of the major cellulase enzymes of *T. reesei* is displayed (Fig 1.5).

1.8.1 Catalytic domain

The CD contains the enzymatic active site, its function is to hydrolyse its substrate of interest. The CD is the largest of the domains within cellulases. On average, the CD accounts for 70% of total protein (Bhat and Bhat, 1997). The structures of CDs are remarkably different between cellulases and significant differences in CDs exist between celluases that are active on the same substrate. Both CBHI and II of *T. reesei* hydrolyse crystalline cellulose; however the CD of CBHI and CBHII share only16.3% sequence homology. Consequently, cellulases and hemicellulases are classified according to the structural features of their CDs (Rabinovich et al., 2002). The classifications were based upon hydrophobic cluster analysis. Most of the CDs that have been crystallized contain structures related to the well known folds of the immunoglobulin-like "jelly-roll" (Rabinovich et al., 2002). The "jelly-roll" is a complex structure in which four pairs of anti-parallel β sheets are wrapped in three dimensions to form a barrel shape. CBHII is a member of GH family 6. The structure of the CD of this GH family 6 member is derived from the triosephosphate isomerase-fold (TIM) barrel. The TIM barrel is a conserved protein fold that contains 8 α-helices and 8 parallel β-strands. However, the CD of CBHII contains 5 α-helices and 7 β-strands, making it an incomplete TIM barrel.

EGI and CBHI share significant amino acid sequence homology (43% of overall sequence). The CDs of EGI and CBHII are included in the same family GH 7. This family contains only fungal cellulases. The structure of the CD within this family is folded in a distorted β-sandwich with its concave site formed by 7 anti-parallel β-strands and convex site formed by eight anti-parallel β-strands. The β-strands are connected with short α-helices (Rabinovich et al., 2002).

BGLI is a member of GH family 3. The CD conserved within this family was proposed to have a (β/α)8-barrel fold (Bhatia et al., 2002).

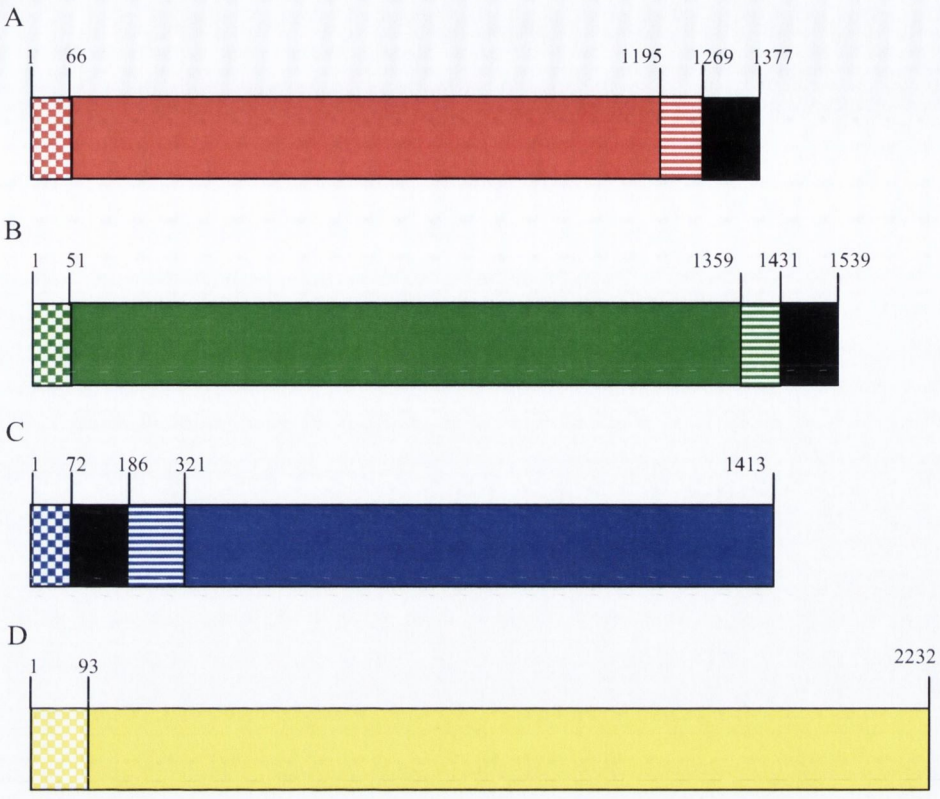


FIG.1.5. Schematic diagram of structure of major *T. reesei* cellulase genes. Genes representing the individual enzymes are colour co-ordinated. Chequered pattern represents the base pair region encoding the secretory signal. Solid colour pattern represents catalytic domain (CD). Linker is represented by horizontal lines. Carbohydrate binding module (CBM) represented by black region. CBMs of EGI, CBHI and CBHII share 70% amino acid sequence homology. Nucleotide positions of domains are indicated. Note locations are based on cDNA sequence therefore will be different from genomic sequence due to absence of introns.

(A) EGI (pink). (B) CBHI (green). (C) CBHII (blue). (D) BGLI (yellow)

1.8.2 Carbohydrate Binding Module

A carbohydrate binding module (CBM) is defined as a region within an enzyme with a discreet fold having carbohydrate-binding activity (Carbohydrate-active enzymes database; www.cazy.org). The cellulose-binding are CBMs are responsible for binding the enzyme to the cellulose surface and hence concentrate enzyme on the substrate surface. It has also been suggested that the CBM may aid in the hydrolysis of crystalline cellulose by physically disrupting the structure of the fibrous cellulosic network and release chain ends without showing any detectable cellulolytic activity (Rabinovich et al., 2002). The exposed chain ends can then bind to the CD for hydrolysis. The CBM was formerly referred to as the cellulose binding domain based on the initial discovery of domains that bind to cellulose (Miettinen-Oinonen, 2004). CBMs have been classified into 39 families based on amino acid similarities; only 13 of these are reported in fungi. The amino acid sequences of the CBMs of *T. reesei* CBHI, CBHII and EGI show 70% sequence homology (Knowles et al., 1987). This is quite remarkable considering that endoglucanase and cellobiohydrolases bind to different cellulose substrates. EGI has been demonstrated to have little to no activity against crystalline cellulose and CBH I or II has no activity against amorphous celluloses such as carboxymethylcellulose (CMC). The CBMs of EGI, CBHI and CBHII are classified in the same CBM family, CBM1, which contains a 33-40 amino acid conserved motif although the location of the CBM within the protein can vary (Fig 1.5). The CBMs found in bacterial cellulolytic species average approximately 100 amino acids. Nuclear magnetic resonance of the CBM of the *T. reesei* CBHI CBM elucidated the structure for the CBM1 family (Rabinovich et al., 2002). The CBM family members have a wedge-like fold with exposed hydrophilic and hydrophobic planes which contain as a base structure a distorted β-sheet of three short anti-parallel strands. The structure is stabilised by two conserved disulfide bonds and contains four highly conserved aromatic residues on its hydrophobic plane thought to be essential for cellulose binding (Rabinovich et al., 2002). The spacing of these residues corresponds to the spacing of every second glucose ring on the glucan chain and it has been postulated that the aromatic residues form van der Waals and aromatic polymerization interactions with the pyranose rings on the surface of the

cellulose (Zhang and Lynd, 2004). The removal of the CBM from *Trichoderma* cellulases results in a decrease in the rate of hydrolysis of insoluble crystalline cellulose but has little effect on the degradation of soluble cellulose (Linder and Teeri, 1997).

T. reesei BGLI does not possess a CBM and consequently cannot bind to cellulose directly (Fig 1.5D). This may be because the main substrate that β-glucosidases reacts on is the soluble cellobiose. Cellobiose is not the only substrate for β-glucosidases, these enzymes also act on glucooligosaccharides such as flavonoid and cyanoglycosides. Wood and Bhat (1988) suggested that BGLs are not technically cellulases. However such is the importance of BGL's conversion of cellobiose and cellooligosaccarides, the major products of the action of EGs and CBHs to glucose that the enzyme is referred to as a cellulase.

1.8.3 Linker

Most CDs and CBMs of cellulases are separated by a linker sequence. Linkers are flexible disordered chains usually rich in proline and hydroxyl amino acids residues such as serine and threonine. The length can vary from 5-100 amino acid residues, although most are limited to 20-50. Their crystal structure cannot be obtained due to the amount of O-glycosylation occurring at the numerous threonine residues. The linker region within cellulases is believed to provide spatial separation of CDs from CBMs to allow their autonomous function on the surface of the insoluble substrate. Deletion mutagenesis of the linker of *T. reesei* CBHI found that while the binding affinity of the enzyme was undiminished, the activity of the enzyme on ordered cellulose was reduced (Rabinovich et al., 2002).

1.9. Cellulase active site location

Crystal structure analysis of CBHI indicated that its active site is concealed within a 50 Å long tunnel formed by four surface loops adjacent to a β-sandwich structure (Stahlberg et al., 1996; Zhang and Lynd, 2004). Due to steric hindrance as a result of the tunnel, CBHs can only hydrolyse cellulose from the exterior chain ends, hence the description of CBHs as exo-glucanases. The glutamic acid residues at positions E229 and E234 act as the nucleophile and proton donor respectively (Kleywegt et al., 1997; Stahlberg et al., 1996).

The active sites within CBHII also occur within a tunnel. Two surface loops form a 20 Å long tunnel adjacent to a α/β-barrel structure. CBHII degrades non-reducing ends. The tunnel of CBHII can accommodate up to six glucosyl groups. The binding of four glucosyl groups contributes to transition state stability (Koivula et al., 2002). Studies of the structure of CBHII revealed a network of interacting residues contributing to formation of the active site. Aspartic acids at positions 199 and 245 (D199 and D245) were identified as the residues closest to the scissile bond. The two carboxyl/carboxylate groups of these residues are within hydrogen bond distance of each other. The carboxylate group of D245 interacts with glycosidic oxygen. This residue was shown to be the proton donor and proposed to act as the acid catalyst. Mutagenesis of this residue by Koivula *et al* resulted in an inactive enzyme (Koivula et al., 2002). The D199 was found to stabilize a uniquely electron-deficient transition state through electrostatic interactions. However the same study also found that other residues may be involved in the CBHII hydrolysis reaction. The mutation of the tyrosine at position 169 suggested that it distorts the glucose ring into a more reactive configuration. Thus ring alteration is an important aspect of the catalytic mechanism of CBHII. An aspartic acid residue at position 425 has been postulated to act as a catalytic base. Numerous hydrogen bond acceptor interactions were observed with nearby amino acid side chains (Koivula et al., 2002).

EGI is structurally related to CBHI; both are members of GH family 7. The EGI active site residues were implicated by their absolute sequence conservation when compared to CBHI. EGI, like CBHI, has glutamic acid (E) residues acting as the catalytic nucleophile and proton donor at positions E196 and E223 respectively (Kleywegt et al., 1997). The major structural difference in EGI compared to CBHI is that the active site of EGI forms an open cleft groove rather than a tunnel (Lynd et al., 2002). This enables EG1 to cleave at random internally within the cellulose chain. Structural analysis identified a number of deletions in EGI that were mapped to tunnel forming loops (Kleywegt et al., 1997). The absolute functional distinction between EGs and CBHs based on substrate specifities is sometimes called into question (Zhang and Lynd, 2004). It has been suggested that cellobiohydrolases could exhibit endoglucanase activity by altering the conformation of their tunnel structures so that the active sites are more exposed (Zhang and Lynd, 2004). Kleywegt *et*

al, also observed that disrupting the loops of the tunnel of an exoglucanase increased the enzymes endoglucanase activity (Kleywegt et al., 1997).

The role of conserved aspartic acid (D) residues in GH family 3 β-glucosidases was established following studies of a β-glucosidase from *Aspergillus wentii*. Nucleophilic activity was found to be associated with the aspartic acid residue contained within a conserved S/TDW motif. In *T. reesei* BGLI the putative nucleophile site was identified as the aspartic acid residue at position D302. Histidine residues have been proposed as proton donors in family 3 BGLs. The active histidine is usually conserved within a motif of KHY/F/L. In *T. reesei* BGLI the proton donor has been implicated as the histidine residue at position 205 (Iwashita et al., 1999).

1.10. Induction of cellulases

The generally accepted view is that cellulases act sequentially and synergistically. It was traditionally believed that cellulose degradation began when endoglucanase cleaved randomly at amorphous sites along the cellulose fibre. This generated a rapid decrease in the degree of polymerisation and exposed new chain ends. The cellobiohydrolases act processively on reducing and non-reducing chain ends to release mainly cellobiose. This results in a gradual decrease in the degree of polymerisation. β-Glucosidases hydrolyse the β-1,4 glycosidic bond of cellobiose and cellooligosaccharides to release monomeric glucose (Fig 1.6).

Cellulase encoding genes are activated in response to the presence of cellulose. Cellulose is an insoluble polymer which cannot transverse the cell membrane and the mechanism of activation in response to cellulose is not fully elucidated. It is generally believed that soluble oligosaccharides released from the polymers function as inducer molecules to trigger cellulase gene expression (Mach and Zeilinger, 2003). The best known soluble inducer of cellulase expression in *T. reesei* is the disaccharide sophorose (Ilmen et al., 1997). Sophorose is composed of two β-1,2 glucose units. Sophorose is believed to be formed from cellobiose. It has been postulated that the transglycosylation activity of BGLI converts the β-1,4 glucose units of cellobiose to the β-1,2 glucose units of sophorose (Mach and Zeilinger, 2003). Studies have found that sophorose has been isolated from cultures of *T. reesei* grown in cellobiose. The sophorose is taken up by a cellobiose permease (Kubicek et al., 2009). This still does not answer the question, "what induces the expression of BGL1

to transglycosylate cellobiose to sophorose and induce general cellulase gene expression?" There are a number of models proposed to explain the initial induction of cellulases (Mach and Zeilinger, 2003). Early studies suggested that there was low basal level of cellulases produced at all times under all conditions. However, Ilmen et al., (1997) found that this was not the case. Cellulase expression was repressed in the presence of glucose. The addition of sophorose to a *T. reesei* culture containing glucose failed to induce cellulase expression and glucose was found to inhibit sophorose uptake. However, prolonged culturing in glucose-depleted media demonstrated cellulase expression albeit at very low levels (Ilmen et al., 1997).

Physical contact between the conidia of *T. reesei* with insoluble cellulose substrate has been proposed as the cause of initial degradation of cellulose thereby releasing inducer molecules to initiate general cellulase expression (Mach and Zeilinger, 2003). *T. reesei* conidia contain an array of enzymes hydrolysing a wide range of polysaccharides. CBHI and CBHII were identified on the surface of the conidia. Twice as much CBHII as CBHI was observed at the conidial tip. Interestingly EGI was not observed. A *T. reesei* strain with *cbh2* gene deleted displayed a lag in growth on cellulose and cellulase formation (Kubicek et al., 2009). A double mutant of *cbh1* and *cbh2* was unable to grow on cellulose. However, during growth on lactose, a known cellulase inducer, cellulase was produced. These findings indicate that cellobiohydrolase particularly CBHII plays a pivotal role in cellulase induction in the presence of insoluble cellulose. The most likely scenario for cellulase induction in *T. reesei* is that cellulose comes in contact with conidia, the CBHII present on the surface performs partial degradation of the crystalline cellulose to release cellobiose. The disaccharide is taken up into the fungal cell via the cellobiose permease. This may result in the expression of intracellular BGL which could transglycosylate cellobiose into sophorose and thereby act as an inducer for general cellulase expression of all endoglucanases, cellobiohydrolases and β-glucosidases. An intracellular BGL, termed BGLII has been identified in *T. reesei* and may fulfil such a role (Saloheimo et al., 2002).

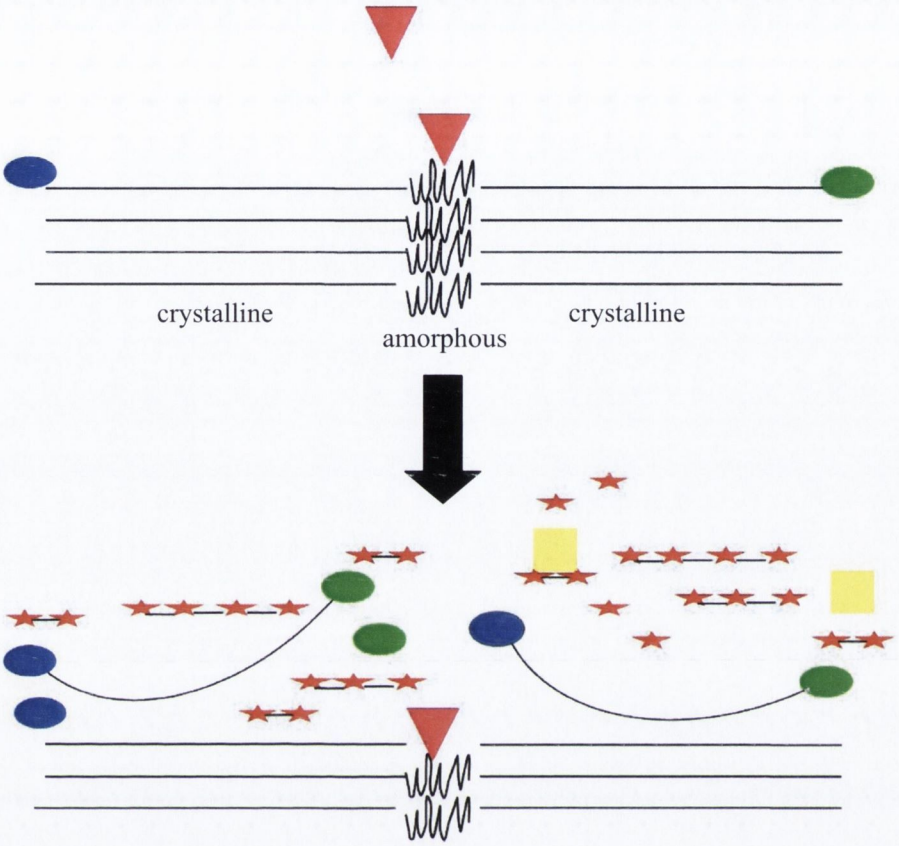


FIG 1.6. Schematic diagram of synergistic activity of cellulase enzymes degrading cellulose.
Pink triangles: EGI. Green oval: CBHI. Blue oval: CBHII. Yellow square: BGL1. Red star: glucose monomer. Tethered double stars: cellobiose. Tethered multiple stars: cellooligosaccharides. Crystalline and amorphous cellulose regions are indicated

The soluble sugar lactose is able to induce cellulase gene expression. Interestingly, lactose occurs naturally only in the milk of mammals, therefore it is unlikely that lactose is a carbon source normally found in the habitat of *T. reesei* (Kubicek et al., 2009). The catabolism of lactose in *T. reesei* is initiated by the extracellular hydrolysis of the sugar to its monomers galactose and glucose by β-galactosidase. These sugars are then taken into the cell and cellulases are induced. Interestingly, culturing *T. reesei* with either galactose, glucose or mixes of both does not induce cellulase expression.

1.11 Secretion of cellulases

It is generally accepted that the secretion pathway in filamentous fungi does not differ greatly from yeast (Palamarczyk et al., 1998). Differences do exist, particularly the mycelial growth phenotype of filamentous fungi that is absent in *S. cerevisiae*. The capacity for protein secretion in filamentous fungi is higher than yeast, for example mutants of *T. reesei* have been shown to secret up to 40 g l^{-1} of extracellular protein (Keranen and Penttila, 1995). Yeast can rarely reach gram-per-litre quantities (Conesa et al., 2001). The reference mutant *T. reesei* strain is RUT30. This strain is characterised by its high levels of protein secretion. It was generated by random mutagenesis by exposing the parental NG14 strain to UV followed by selection for the ability to hydrolyse cellulose in catabolite repressing conditions (Le Crom et al., 2009). Electrophoretic karyotyping of RUT30 indicated chromosomal rearrangements but the precise genetic changes that occurred are poorly understood (Le Crom et al., 2009).

Fungal secretory proteins are synthesized on ribosomes and begin the journey to the extracellular medium by entering the endoplasmic reticulum (ER). Secretion is mediated by signal peptides that are located near the N-terminal of secretory proteins. Studies of *S. cerevisae* indicated that two routes exist for protein targeting to the ER membranes; the signal recognition particle (SRP) - dependent pathway and a SRP-independent pathway (Conesa et al., 2001). It is postulated that filamentous fungi may use similar pathways. The hydrophobicity of the signal sequence determines the targeting route of each protein. Proteins with a less hydrophobic signal sequence are translated through the SRP-independent route whereas both routes can be followed when more hydrophobic signals are present. The signal peptides of the major

cellulases of *T. reesei* contain a higher percentage of hydrophobic amino acids than hydrophilic residues (see Table 1.3). Proteins have to fold and mature into their native forms to be functional; this process is facilitated by chaperones and foldases. In the ER proteins undergo co- and post-translational modifications. The ER-related events contribute to the folding of proteins. An important function of the ER is "quality control". The cell must ensure that only correctly folded proteins are secreted. Quality control involves two cellular mechanisms, unfolded protein (UPR) and the ER-associated protein degradation (ERAD). UPR detects the presence of unfolded protein in the ER and ERAD degrades proteins that fail to reach conformation (Conesa et al., 2001).

Correctly folded proteins exit the ER to be targeted to the Golgi apparatus. The characteristic Golgi structure is not always seen in filamentous fungi, however Golgi-associated functions are present. The transport from the ER to the Golgi involves transport via coated vesicles. Post-translational modifications occur after vesicular transfer of proteins to the Golgi apparatus such as the proteolytic cleavage of the prepro-protein amino acid sequence. The KEX2 proteases are responsible for the excision of the mature protein from the larger precursor proprotein (Julius et al., 1984). The importance of Kex2 cleavage of *T. reesei* secretory proteins was demonstrated when the inhibition of Kex2p resulted in a decrease in secretion and increase in intracellular accumulation of non-cleaved protein (Palamarczyk et al., 1998).

N- and O-glycosylation events occur within the Golgi. Oligomannose N- and O- glycans are predominant in filamentous fungi whereas in *S. cerevisiae* hyperglycosylation often occurs. Most cellulases contain high mannose-type oligosaccharide structures attached through N-glycosidic linkages. The CBHs and EGs also contain O-linked oligosaccharides. CBHI is believed to have at least 4 potential N-glycosylation sites, CBHII, EGI and BGLI are predicted to have 3, 6 and 7 N-glycosylation sites respectively (Palamarczyk et al., 1998). *Trichoderma reesei* N-linked glycosylated oligosaccahrides or glycans are small compared to *S. cerevisiae* N-linked glycans. The importance of N- and O-glycans for proteolytic stability and efficient secretion is different for every fungal glycoprotein. Oligosaccharides are generally accepted as the determining factor for the conformation of a glycoprotein. A previous study indicated that the inhibition of O-glycosylation resulted in a decrease in the

secretion of endoglucanases from *T. reesei* (Palamarczyk et al., 1998). This suggested that O-mannosylation of proteins may be essential for protein secretion (Palamarczyk et al., 1998). The effects of hyper-glycosylation of cellulase enzymes have been studied in yeast (Cummings and Fowler, 1996; Okada et al., 1998). In each case enzyme activity did not seem to be affected. After passage through the Golgi compartment, proteins are targeted to the plasma membrane via vesicles for extracellular secretion (Idiris et al., 2010) (Fig 1.7).

Table 1.3. Summary of secretory signal peptides features of major *T. reesei* cellulases

Enzyme	S.S. length	Secretory signal sequence[a]	%Hydrophobicity
EGI	22	MYRKLAVISAFLATA**RA**	77%
CBHI	17	MIVGILTTLATLATLAASVPLE**ER**	65%
CBHII	24	MAPSVTLPLTTAILAIARLV**AA**	67%
BGLI	31	MRYRTAAALALATGPFARADSHSTSGASA**EA**	55%

[a] Residues in bold indicate sites for cleavage by KEX-like proteases

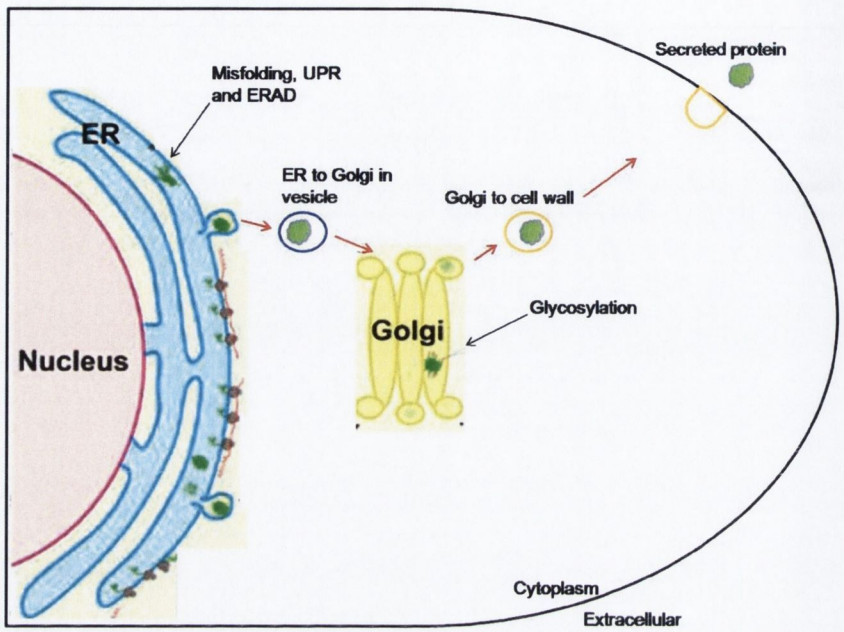


FIG 1.7. Schematic diagram of secretory pathway of heterologous proteins in yeast. The main pathways are indicated. Figure adapted from Idiris et al., (2010).

1.12. Previous research in generating recombinant cellulolytic organisms

The ultimate goal of research into generating bioethanol from lignocellulose is to produce an organism capable of SSF. A substantial proportion of bioethanol research has focused on the expression of cellulases in *S. cerevisiae* (Lynd et al., 2002). The Finnish Institute VTT performed pioneering work on the expression of functional cellulases in this yeast. Penttïlla and colleagues (1987) cloned and expressed cDNA copies of *T. reesei* genes *egl1* and *egl3* in *S. cerevisiae.* The recombinant genes were expressed under the control of the *S. cerevisiae* constitutive promoter phosphoglycerate kinase 1 (PGK1) and their native respective signal sequence. The study discovered that endoglucanase activity was present in the yeast clone supernatant in stationary phase after 40 h. This proved that EGI was secreted by the host secretory mechanisms. The study also indicated that the recombinant EGI enzyme present in the culture medium was of an apparent larger molecular weight than the native enzyme secreted by *T. reesei.* The reduction in size of the recombinant protein when treated with endoH enzyme suggested that EGI was hyper N-glycosylated by *S. cerevisiae.* The hyper glycosylation did not seem to block enzyme activity (Penttila et al., 1987).

Penttïla et al., (1988) also demonstrated the expression of *T. reesei* CBHI and CBHII by recombinant yeast. The study indicated that CBHII was expressed at higher levels than CBHI by *S. cerevisiae.* The recombinant CBH enzymes had lower specific activities than the native *T. reesei* enzymes. This was attributed to hyperglycosylation by the yeast host. A recent study compared the expression of four fungal cellobiohydrolase encoding genes, *T. reesei cbh1* and *cbh2*, *A. niger cbhB* and *Phanerochaete chrysosporium cbh1-4* in *S. cerevisiae.* The *cbh1* insert contained its native secretory signal and was expressed under the control of the *S. cerevisiae* alcohol dehydrogenase (ADH) 2 promoter. The other gene inserts had their native signal sequence replaced by the secretory sequence of the *T. reesei xyn2* gene and were expressed under the control of the *S. cerevisiae* enolase I promoter. The study indicated that while all of the cellobiohydrolase inserts were expressed by the yeast, the extracellular titres of secreted enzyme were low (Den Haan et al., 2007a). The activity of CBHII was demonstrated to be significantly higher than the other CBHs. Cellobiohydrolase enzymes represent a challenge for recombinant enzyme

expression in yeast due to their lower specific activity than endoglucanase or β-glucosidase (Den Haan et al., 2007a).

BGL1 enzyme activity in *T. reesei* is lower than in other cellulolytic organisms such as *Aspergillus* species (Lynd et al., 2002). *T. reesei* β-glucosidase has been expressed in *S. cerevisiae* (Cummings and Fowler, 1996). Cummings *et al* inserted the *T. reesei* gene *bgl1* under the control of the *S. cerevisiae* galactokinase GAL1 promoter. A study by van Rooyen and co-workers (2005) inserted a variety of fungal β-glucosidase genes into *S. cerevisiae* to enable growth on cellobiose. The β-glucosidase genes examined were *bglA* from *A. kawachii*, *bglB* from *Candida wickerhamii*, *bgl1* from *Saccharomycopsis fibuligera* and *bgl1* from *T. reesei*. The native secretory signal of each gene was replaced with the signal sequence of the *T. reesei xyn2* gene and expression was controlled by the *S. cerevisiae* PGK1 promoter. The results of the study found that no BGL activity was observed in the culture supernatant for any of the clones; however, active BGL was detected within the periplasmic space of the recombinant yeast strains. The yeast strain Y294[SFI] containing *bgl1* from *S. fibuligera* grew in media containing cellobiose as the sole carbon source (van Rooyen et al., 2005). The strains containing β-glucosidase genes from either *T. reesei* or *C. wickerhamii* did not grow in cellobiose-enriched media. The strain Y294 [SFI] grew in cellobiose without secreting extracellular BGL indicated that cellobiose must have been taken up by the yeast and metabolised internally. *S. cerevisiae* does not naturally possess a cellobiose transporter but perhaps the hexose transporters (HXT) could facilitate uptake.

Cho and colleagues (1999) claimed to be the first group to develop a *S. cerevisiae* strain that expressed a full set of cellulolytic genes. In this study the bi-functional endo/exo-glucanase gene from *Bacillus sp. D04* and the β-glucosidase gene of *B. circulans* were integrated into the chromosome of *S. cerevisiae*. Each gene contained the killer toxin secretory signal and expression was controlled by the *S. cerevisiae* ADH1 promoter. The resultant yeast strain L2612δGC successfully expressed and secreted cellulase enzyme and was able to ferment 15 g l^{-1} of crystalline cellulose derived cellodextrin to 4 g l^{-1} of ethanol (Cho et al., 1999). The latter study was a breakthrough for the simultaneous expression of different enzymes. Fujita et al., (2002) also expressed multiple cellulase enzymes in yeast. The genes to be expressed were

tethered to the yeast cell surface. The *T. reesei* endoglucanase gene *egl2* and the *A. aculeatus* β-glucosidase gene *bgl1* were anchored to the cell wall by fusing mature protein with the C-terminal of the α-agglutinin protein of *S. cerevisiae*. Each of the cellulolytic encoding genes had its native secretory signal replaced by the secretory sequence of the glucoamylase gene from *Rhizopus oryzae* and expression was controlled by the *S. cerevisiae* glyceraldehyde-3-phosphate dehydrogenase (GAPDH) constitutive promoter. The resultant strain transformed with plasmids containing the cellulolytic inserts was found to have enzyme activity located on the cell surface and not the extracellular supernatant. Immunofluroescence microscopy confirmed that the two cellulolytic enzymes were co-displayed on the cell surface of *S. cerevisiae*. The engineered yeast was able to ferment 45 g l^{-1} of cereal or mixed β-glucan to 16.5 g l^{-1} ethanol. β-glucan is a linear, soluble polysaccharide composed of 70% β-1,4-Glucan and 30% β-1,3-Glucan (Fujita et al., 2002). A continuation study by Fujita et al., (2004) engineered a *S. cerevisiae* strain to contain the *T. reesei egl2*, *cbh2* genes the and the *A. aculeatus* gene *bgl1*. The inserts contained the same signal sequence and promoter as described in Fujita's previous study (Fujita et al., 2002). The recombinant yeast was found to co-display all three cellulase enzymes on the cell surface and was able to ferment at 30°C 10 g l^{-1} of phosphoric acid swollen cellulase to produce 3g l^{-1} of ethanol. This study was the first to report the co-display of 3 recombinant proteins on a yeast cell surface. The goal of this strategy of anchoring cellulases to the cell wall was to mimic, in yeast, the cellulosome found in cellulolytic bacteria. Tethering cellulases to the yeast surface could concentrate enzymes rather than possibly diluting their activity by secreting them into the supernatant. Fujita and colleagues (2004) did not report their recombinant yeast being able to sustain growth on pure cellulose.

Wen et al., (2010) also favoured engineering *S. cerevisiae* to display a mini-cellulosome. In this study *T. reesei* gene products EGII and CBHII along with *A. aculeatus* BGLI were connected via a cellulosome miniscaffoldin containing a cellulose binding domain and three cohesion modules. This was tethered to the yeast cell surface through the α-agglutinin adhesion receptor. The miniscaffoldin was encoded by CipA3 of *Clostridium thermocellum.* The expression of the cellulases was under the control of GAL promoters and therefore induced in media containing galactose. The *S. cerevisiae* strain

expressing all three cellulase was shown to grow on 1% phosphoric acid swollen cellulase (PASC) and produce 1.8 g l^{-1} of ethanol (Wen et al., 2010). Den Haan et al., (2007b) were the first group to achieve growth of recombinant yeast on PASC as the sole source of carbohydrate. PASC is a matrix of amorphous cellulose that has been converted from Avicel crystalline cellulose. The yeast strain Y294[Cel5] co-expressed *T. reesei* EGI and *S. fibuligera bgl1* enzymes. The *egl1* gene insert contained its native secretory signal and expression was controlled by the *S. cerevisiae* enolase 1 promoter. The *BGL1* gene insert had the signal sequence of the *T. reesei xyn2* gene and expression was controlled by the PGK1 promoter. Despite not containing a cellobiohydrolase, the recombinant yeast strain Y294[Cel5] managed to ferment 10 g l^{-1} of PASC to 1 g l^{-1} ethanol and importantly could sustain growth on media containing PASC as the sole carbon source. Unlike Fujita et al., (2004), Den Haan et al., (2007b) did not tether the enzymes to the cell surface.

Traditionally, the cellulolytic enzymes from filamentous fungi were the inserts of choice but the search for more efficient cellulases has led to the expression of recombinant bacterial cellulases in yeast. A study by van Wyk et al., (2010), recently inserted and expressed the cellulase gene *cel9A* of the thermophilic actinomycete *Thermobifida fusca*. The *cel9A* gene coded for Cel9A protein that was a processive endoglucanase enzyme which also had exo-glucanase activity (van Wyk et al., 2010). The *cel9A* gene had its secretory signal replaced with the signal sequence of *T. reesei xyn11B* gene and expression was under the control of the *S. cerevisiae* PGK1 promoter. The recombinant yeast strain Y294[CEL9A] displayed activity on both CMC and crystalline cellulose. The chromography analyses of soluble sugars released from reaction with Avicel revealed a cellobiose/glucose ratio of 2.5:1. The strain Y294[CEL9A] was capable of growth on media with PASC as the sole carbon source. This was the first demonstration of yeast growing on a cellulosic substrate as a sole carbohydrate source while only expressing one recombinant gene (van Wyk et al., 2010). This indicated that Cel9A produced relatively high titres of glucose to enable yeast grow on cellulose. Du Plessis et al., (2010) observed that recombinant yeast producing CBHI reacted with PASC to yield a cellobiose/glucose ratio of 6:1, which was not able to sustain growth on this cellulose substrate.

S. cerevisiae is not the only ethanologenic microbe used for bioethanol production from cellulose. Alternative yeasts are being considered for bioethanol production from cellulose. A key attribute to any potential ethanologen is thermotolerance. Cellulases are optimum at 50°C, whereas *S. cerevisiae* ferments optimally at a maximum of 30°C. *Kluyveromyces* yeast species are more thermotolerant than *Saccharomyces* yeast. *K. marxianus* can produce ethanol at temperatures above 40°C (Fonseca et al., 2008). Hong *et al* utilised *K. marxianus* as a host for expression of functional cellulases (Hong et al., 2007). The study displayed that the recombinant *K. marxianus* strain fermented a 10% (w/v) cellobiose minimal media to produce 43.4 g $l^{-1}$of ethanol. A disadvantage with *K. marxianus* is that it is less ethanol tolerant than *S. cerevisiae* (Fonseca et al., 2008).

Bioethanol production is not limited to eukaryotic hosts. *Zymomonas mobilis* is a widely used ethanol producing bacterium. *Z. mobilis* also has GRAS status and has a high ethanol tolerance of 120 g l^{-1}. The bacteria cannot naturally degrade cellulase but genetic engineering to introduce recombinant genes has been performed. Bresticgoachet et al., (1989) successfully introduced and expressed the endoglucanase gene *celZ* of *Erwinia chrysanthemi* in *Z. mobilis.*

An alternative approach employed by researchers was to introduce the entire ethanol producing pathway enzymes into natural cellulolytic organisms. The cellulase producing filamentous fungi *Trichoderma* and *Aspergillus* produce poor fermentation yields due to low resistance of these microbes to higher concentrations of ethyl alcohol (Lin and Tanaka, 2006). Heterologous gene expression of pyruvate decarboxylase and alcohol dehydrogenase from *Z. mobilis* in the cellulolytic bacterium *Clostridium cellulolyticum* was found to increase ethanol production by 53% (Carere et al., 2008).

An alternative area of research for bioethanol is the in *vitro* production of cellulase enzymes. This involves the production and purification of cellulase enzymes from native cellulolytic organisms. The *in vitro* production of cellulases is essential because it allows each enzyme to be extensively characterised in terms of optimal temperature and pH range for activity. This information is very important for ensure efficient degradation of cellulose. The Finnish group at the VTT institute have led the way in the characterisation

cellulases from *T. reesei* (Miettinen-Oinonen, 2004). However *in vitro* research of cellulases from other non-*Trichoderma* organisms is an area of great interest. The research performed on the filamenous fungus *Talaromyces emersonii* is particularly important. *T. emersonii* is an aerobic thermophilic fungus that produces a complete cellulase system (Tuohy et al., 2002). Studies have demonstrated that the cellulases of *T. emersonii* are thermophilic and function at temperatures 10-20°C higher than *T. reesei* (Tuohy et al., 2002; Fernandes et al ., 2010). These so-called thermozymes are optimally active at pH 4.5-5.0 and 70°C (Waters et al., 2010). Cellulase enzymes that can operate at increased temperatures are of interest due to their ability to degrade cellulose more efficiently. Indeed a recent study by Voutilainen et al., (2010) has cloned and expressed the *T. emersonii* cellobiohydrolase gene *cel7a* into *S. cerevisiae*.

1.12.1 Utilization of pentose sugars

Another area of interest in biofuel production is pentose fermentation. The conversion of biomass to useable energy will not be economically viable unless hemicellulose is used in addition to cellulose (Jeffries, 2006). Xylose is the second most abundant sugar in nature and is a major sugar present within hemicellulose and can account for up to 30% of sugar within lignocelluloses (Jeffries and Shi, 1999). *S. cerevisiae* cannot naturally metabolise xylose but there are natural xylose-fermenting yeasts such as *Pichia stipitis*, *Candida shehatae* and *Pachysolen tannophilus* (Ryabova et al., 2003). *P. stipitis* has the highest natural capacity for xylose fermentation (Jeffries et al., 2007). Large scale fermentations of xylose by the natural xylose fermenters are hampered by their sensitivity to ethanol concentrations greater than 40 g l^{-1}. Efforts have been attempted to engineer *S. cerevisiae* to ferment yeast. Xylose is degraded by the pentose phosphate pathway via a two-step reduction and oxidation mediated by the proteins xylose reductase, Xyl1p, and xylitol dehydrogenase, Xyl2p, respectively. The corresponding genes from *P. stipitis xyl1* and *xyl2* were expressed in *S. cerevisiae* but yielded poor ethanol production from xylose (Walfridsson et al., 1995). An improvement of xylose metabolism in *S. cerevisiae* was accomplished by overexpression of the homologous xylulokinase gene *xks1* along with *P. stipitis xyl1* and *xyl2* (Ho et al., 1998). Xylose fermentation by recombinant yeast is less efficient than that of glucose due to the bottle neck of redox imbalance of cofactors for xylose reductase and

xylitol dehydrogenase. Another limitation is the uptake of xylose into the cell. Xylose uptake into *S. cerevisiae* is mediated by low affinity hexose-transporters. Improved transport of xylose is an area of research interest (Hamacher et al., 2002; Sedlak and Ho, 2004; Saloheimo et al., 2007; Runquist et al., 2010).

Alternative hosts other than *S. cerevisiae* for xylose fermentation have been explored. The thermotolerant methylotrophic yeast *Hansenula polymorpha* is an interesting choice. *H. polymorpha* is known to ferment cellobiose and xylose as well as glucose (Ryabova et al., 2003). *H. polymorpha* ferments optimally at 37°C. However, *H. polymorpha* is less ethanol tolerant than *S. cerevisiae* (Ryabova et al., 2003).

1.13. Cellulase encoding enzymes chosen for project

T. reesei was chosen as the donor of the cellulase encoding genes for this project because it represented the paradigm for enzymatic breakdown of cellulose and its enzymes are well characterised (Martinez et al., 2008). The *egl1* gene was chosen because it is the chief endoglucanase produced by the fungus. The primary cellulase produced by *T. reesei* is CBHI. It accounts for 60% of protein secreted during induction (Takashima et al., 1998). However, a study of recombinant CBHI and CBHII secreted by *S. cerevisiae* observed that CBHII produced more activity than CBHI (Penttila et al., 1988). Consequently, for this reason, the gene *cbh2* which encodes the enzyme CBHII was chosen for the cellobiohydrolase of this current study. The chief β-glucosidase of *T. reesei* is BGLI encoded by the gene *bgl1* and is essential for converting cellobiose to monomeric glucose. This gene was chosen for this project.

As described in Section 1.12, the *T. reesei* genes *egl1, cbh2 and bgl1* have been individually expressed in *S. cerevisiae* (Cummings and Fowler, 1996; Penttïla et al., 1988; Penttïla et al., 1987) and *egl1* and *cbh2* have been co-expressed along with other non-*T. reesei* cellulases (Den Haan et al., 2007b; Fujita et al., 2004). However according to the literature there is no documented study featuring all of the above three *T. reesei* genes being expressed by recombinant yeast strains for the SSF of cellulose to ethanol in a co-culture fermentation.

In this project, the cloned *T. reesei* genes were expressed with their native secretory signal. Based on the studies of *egl1* and *cbh2* by Pentïlla et al., (1987; 1988) and *bgl1* by Cummings and Flowers (1996), the native secretory

sequence of the cellulase enzymes should be recognised by the *S. cerevisiae* secretory mechanism. An aim of this project was to replicate the non-complexed cellulase system of *T. reesei* in yeast and permit the extracellular secretion of the enzymes. Fujita et al., (2004) and Wen et al., (2010) took a different approach; their cellulases were tethered to the cell wall of the yeast and thus resembled a bacterial cellulosome. The merits of this approach would be that the enzyme would be concentrated at the cell and thus no loss of glucose would occur. This approach may be more suited to the use of cellulase enzymes designed for a cellulosome, such as Clostridial cellulases (Tsai et al., 2009). To anchor the cellulases to the cell wall, Fujita and colleagues fused the celluase genes to the 3'end of the α-agglutinin gene of *S. cerevisiae* (Fujita et al., 2004). *T. reesei* enzymes are naturally extracellularly secreted; the fusing of an additional 320 amino acids of α-agglutinin to the C-terminal of the cellulase enzymes could alter the natural fold of the enzymes and may affect activity. The non-complexed approach, taken by this project, allowed the cellulase enzymes to be folded to the natural conformation. Extracellular EGI can act as a scavenger for amorphous cellulose. In *T. reesei*, BGLI is found extracellularly; however, a large portion is naturally cell bound (Lynd et al., 2002).

1.14. Target organism: *Saccharomyces* yeast

T. reesei is not an efficient fermenter of sugars to ethanol. The best known fermenting microorganisms are yeast of the *Saccharomyces* genus. The *Saccharomyces* genus includes two types of species; *sensu stricto* and *sensu lato* (Rainieri et al., 2003). *Saccharomyces sensu stricto* species such as *S. cerevisiae*, *S. bayanus* and *S. pastorianus* are associated with the fermentation industry. The commercial production of bioethanol has been dominated by *S. cerevisiae* (Gray et al., 2006). *S. cerevisiae* has several advantages over other ethanologenic microbes. *S. cerevisiae* is generally regarded as safe (GRAS) and has been more extensively characterised biochemically and genetically than any other eukaryote. The genome sequence information has lead to an increased choice of promoter systems for yeast expression vectors. As an eukaryote, *S. cerevisiae* is a suitable host for high-level production and secretion of recombinant protein in its correctly folded form (Romanos et al., 1992). The most important characteristic of *S. cerevisiae* from an industrial point of view is its high fermentative capacity and robust tolerance to ethanol.

Strains of this yeast have been demonstrated to produce yields of 18% ethanol (Lin and Tanaka, 2006).

The brewing industry has distinguished two types of brewer's yeast; ale and lager. The yeasts used for the production of ales and lager were originally classified based on their flocculation properties. Ale yeasts float to the top of vat at the end of fermentation and are therefore known as top fermenters. Lager yeasts sediment to the bottom of vat and are known as bottom fermenters. The major ale yeast is *S. cerevisiae* whereas lager yeast is a natural hybrid and belongs to the species *S. pastorianus*. *S. pastorianus* was formed by the union of *S. cerevisiae* and *S. bayanus* related species (Dunn and Sherlock, 2008). The most likely *S. cerevisiae* ancestral parent was an ale yeast. *S. pastorianus* is believed to have arisen in response to selective pressures from cold brewing. *S. bayanus* strains are known to be more cold tolerant than *S. cerevisae*. Therefore *S. pastorianus* may have arose by instantaneous speciation due to an interspecies hybridization event that occurred during these selective conditions (Dunn and Sherlock, 2008). *S. pastorianus* is characterised by its low fermentation temperature of 6-14°C, whereas *S. cerevisiae* ferments optimally at 20-30°C. Fermentation with *S. pastorianus* accounts for 90% of world-wide beer production (Nakao et al., 2009). Studies involving competitive genome hybridisation (CGH) analysis demonstrated that *S. pastorianus* has two types of each chromosome from *S. cerevisiae* and *S. bayanus* (Bond et al., 2004; Dunn and Sherlock, 2008) Consequently, there are two divergent orthologous genes found in *S. pastorianus* for almost all genes found in the parental strains (Saerens et al., 2010). A CGH study performed by Dunn and Sherlock analysing 17 different *S. pastorianus* strains concluded that there are two independent origins of *S. pastorianus* strains. The two groups are characterised by different genome rearrangements, copy number and ploidy differences. In both groups the *S. bayanus* genome was more invariant than that of *S. cerevisiae* in terms of ploidy changes. The sequencing of the *S. pastorianus* strain Weihenstephen 34/70 revealed that it contained 36 chromosomes whereas the parental strains each possess 16 (Nakao et al., 2009). *S. pastorianus* strains are aneuploid, possessing unequal numbers of chromosomes. The genome of *S. pastorianus* is dynamic and is capable of adapting to environmental stresses. The stressful conditions of alcohol fermentation such as elevated levels of CO_2, ethanol and

low pH may have driven the evolution of *S. pastorianus*. A study of *S. cerevisiae* in continuous culture under glucose limitation revealed that selection pressures imposed by environmental conditions can result in the selection of a strain being aneuploid (Dunham et al., 2002). Subjecting *S. pastorianus* to fermentation conditions promotes recombination events to occur that generate mosaic chromosomes which can sometimes confer tolerance to stresses such as high ethanol concentration, high osmolarity and heat shock (James et al., 2008). *S. pastorianus* is more stress-tolerant than *S. cerevisiae*. Stress-tolerance is an important characteristic in an industrial setting because yeast collected from the end of fermentation are used again or "re-pitched" at the start of another (James et al., 2008). From an industrial perspective, it would be of great benefit to use yeast that can withstand stresses. *S. pastorianus* has been demonstrated to adapt to stressful environments due to its genomic plasticity which can undergo chromosomal rearrangements to confer resistance. In haploids, such as *S. cerevisiae*, chromosomal rearrangements may reduce fitness (James et al., 2008).

Previous work performed in the laboratory generated a set of stress tolerant *S. pastorianus* strains using a cross-tolerance strategy (James et al., 2008). Cells that acquire tolerance to one form of stress such as heat-shock can be tolerant to other unrelated stresses such as growth in high ethanol concentration. The parental lager yeast strain CMBS-33 was mutagenised with ethyl methanesulfonate. The surviving yeast were heat shocked at 55°C for 10 min followed by plating on high specific gravity agar. One such surviving yeast was designated C10. The genome of C10 was found not to differ significantly from the parental CMBS-33. Strain C10 displayed an increased fermentation rate compared to its parent (T.C. James, Bond unpublished data). A second generation of mutants was generated by exposing the first generation of mutants to repeated rounds of heat-shock treatment; one such mutant was C10-51. This strain was chosen as a result of its increased stress tolerance.

1.15. Aims of the project

The overall aim of this project was to generate yeast species capable of SSF of cellulose into ethanol. To date, most research on cellulase production has focussed in the yeast *S. cerevisiae* and there has been little attention paid to exploring the use of other *Saccharomyces* species such as *S. pastorianus,* which is more robust, for the SSF of cellulose into bioethanol. In this study the

lab strain *S. cerevisiae* S150 and stress tolerant lager strain *S. pastorianus* C10-51 were used as hosts for expression of *T. reesei* cellulases.

An additional aim of this study was to recreate the non-complexed cellulase system of *T. reesei* within recombinant yeast. Other groups have cloned *T. reesei* cellulase genes along with genes from different organisms into yeast however there are no reports of all 3 class of *T. reesei* cellulase expressed in yeast. *T. reesei* represents the prototypical cellulolytic organism. The fact that *T. reesei* encodes for fewer cellulase genes than other cellulolytic fungi (Table 1.2) demonstrates the efficiency of *T. reesei* cellulase system. It was decided to clone cellulase encoding genes from one organism and not from different sources; this would ensure that the same natural synergistic balance would be maintained as in nature.

The goals of the project were as follows:

1. To clone all three classes of cellulase genes (*egl1*, *cbh2* and *bgl1*) into two different yeast species, namely *S. cerevisiae* and *S. pastorianus*.
2. To engineer the expressed proteins so as to be secreted from the yeast cells into the medium.
3. To characterise the activity of the expressed cellulases and to examine the effects of mutations within the catalytic domain and the carbohydrate binding domain on the activity of cellulases.
4. Finally to determine if simultaneous saccharification and fermentation can be carried out by co-culture of strains expressing the individual cellulases and to compare the levels of alcohol produced from yeast grown on cellulose as a sole carbohydrate source.

Working hypothesis:

The bio-engineered yeast transformed with cellulase encoding genes would *in vivo*, express and produce recombinant cellulase enzymes. The extracellular secretion of the cellulases that contained the native fungal secretory sequence would confirm that *Saccharomyces* yeast recognise and correctly process foreign signal peptides. The production of recombinant cellulase enzymes should enable the yeast to utilise a cellulosic substrate as a source of carbohydrate and hence ferment cellulose into ethanol.

Chapter 2: Materials and Methods

2.1 Reagents, Strains and Media

All chemicals and reagents were of analytical grade and unless otherwise stated were purchased from Sigma-Aldrich (Dublin, Ireland). The water used throughout the research was sterile distilled water.

S. cerevisiae strain S150 (Bond laboratory, Trinity College Dublin) (*Mat* **a**, *leu*2-3, 112 *ura*3-52, *trp*1-289, *his*3) is a haploid laboratory yeast strain. *S. pastorianus* C10-51 (Bond laboratory, Trinity College Dublin) is a polyploid yeast strain that was been demonstrated to exhibit stress-tolerance (James et al., 2008). C10-51 was derived from the industrial strain *S. pastorianus* CMBS-33 but due to chromosomal rearrangements the genomic make up of C10-51 differs significantly from its parental strain.

T. reesei strain QM9123 was purchased from the Commonwealth Agricultural Bureaux International, CABI, (Surrey, UK).

Plasmid propagation for sequencing was carried out in *Escherichia coli* strain XL1 (*recA1*, *endA1m*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac*[*F'proAB*, *lacIZ_M15*, *Tn10*] (Stratagene, La Jolla, CA, USA).

The *S. cerevisiae* strains unless otherwise stated were grown at 30°C for 48h in yeast extract peptone dextrose (YPD; 10 g l^{-1} yeast extract, 20 g l^{-1} peptone, 20 g l^{-1} dextrose) or synthetic complete medium (SC; 1.7 g l^{-1} yeast nitrogen base w/o amino acids and ammonium sulfate (Foremedium, Norwich, England), 0.5g l^{-1} $(NH_4)_2SO_4$, 0.57 g l^{-1}drop out medium (Foremedium, Norwich, England), 20g l^{-1} glucose and supplemented with amino acids: 0.2 g l^{-1} adenine, 0.12 g l^{-1} lysine, 0.08 g l^{-1} tryptophan, 0.08 g l^{-1} leucine, 0.02 g l^{-1} histidine, and 0.03 g l^{-1} uracil. If uracil or histidine was omitted, the medium is referred to as SC ura^- or SC his^- respectively. Agar plates of the above media were made by addition of 17 g $l^{-1}$of agar. *S. pastorianus* strains were cultivated in the same media but at 25°C.

For the induction of yeast clones that contained the pGREGGAL1 vector, a medium containing galactose as a carbon source was used to promote expression of the cellulolytic genes. Cells were grown in YPGal (10 g l^{-1} yeast extract, 20 g l^{-1} peptone, 20 g l^{-1} galactose) or SC galactose (as described above except containing 20 g l^{-1} galactose). Clones that contained pGREGPGK1 plasmids were cultured in

SC ura$^-$ media containing glucose or YPD containing 0.2 g l^{-1} of G418 (geneticin) aminoglycoside antibiotic.

For fermentations, three different media were used; SC 10% glucose, SC 10% maltose and 10X SC phosphoric acid swollen cellulose (PASC). The former two media had the same composition as standard SC except with 200 g l^{-1} glucose or maltose, respectively. SC PASC (10X) was composed of 17 g l^{-1} yeast nitrogen base w/o amino acids and ammonium sulfate (Foremedium, Norwich, England), 5 g l^{-1} $(NH_4)_2SO_4$, 100 g l^{-1} PASC and supplemented with amino acids; 2 g l^{-1} adenine, 1.2 g l^{-1} lysine, 0.8 g l^{-1} tryptophan, 0.8 g l^{-1} leucine, 0.2 g l^{-1} histidine, and 0.3 g l^{-1} uracil.

Selection of the pGREG586 plasmid (referred to hereafter as pGAL1) was maintained by culturing in the presence of or via auxotrophic SC ura$^-$ media.

Trichoderma reesei fungus was grown in YPD at 25°C. For cellulase induction, *T. reesei* was grown in 3% lactose synthetic complete medium (L-SC; 1.7g l^{-1} yeast nitrogen base w/o amino acids and ammonium sulfate, 0.5 g l^{-1} $(NH_4)_2SO_4$, 30 g l^{-1} lactose and supplemented with amino acids; 0.2 g l^{-1} adenine, 0.12 g l^{-1} lysine, 0.08 g l^{-1} tryptophan, 0.08 g l^{-1} leucine, 0.02 g l^{-1} histidine, and 0.03 g l^{-1} uracil.

Cells were grown in Luria-Bertani medium (10 g l^{-1}Bacto tryptone, 5 g l^{-1} yeast extract, 5 g l^{-1} NaCl containing 25 μg carbenicillin ml^{-1} at 37°C overnight.

2.2 DNA extraction

Genomic DNA was extracted from *T. reesei* based on a protocol described by Cassago et al., (2002) with the following modifications. After 48 h cultivation in YPD, 5 ml of *T. reesei* culture was pelleted by centrifugation at 14,000rpm for 5 min. The pellet was then resuspended in 400 μl of DNA resuspension buffer (1% [wt/vol], sodium docecyl sulphate [SDS], 10 mM EDTA and 0.1 M Tris-HCl [pH 8]) using a vortex mixer. A 1/3 volume of zirconia glass beads was added and the mixture was vortexed virogously for 5 min. One volume of buffered phenol-chloroform [0.1 M Tris, pH 8] was then added to the mix, which was vortexed for 5 min. The suspension was centrifuged at 14,000 rpm for 5 min. The aqueous layer was extracted with one volume of phenol-chloroform, followed by an extraction

with one volume of chloroform. DNA was precipitated out of solution by adding 1/10 volume of 3M sodium acetate [pH 5.3], 3 volumes of absolute alcohol and incubating for 1h at –70°C. DNA was pelleted and washed in 70% (v/v) ethanol. The dried DNA pellet was finally resuspended in 50 µl of nuclease-free H_2O.

DNA was extracted from yeast cultures using the same protocol as described above with the exception that the yeast cultures were grown in SC ura$^-$ or in YEPD supplemented with 0.2 g l^{-1} G418. In general 5-50mL cultures, grown for 24hrs at 25-30°C, were used for DNA extraction.

2.3 RNA extraction

RNA was extracted from *T. reesei* as follows. The fungus was grown in 50 ml of 3% lactose SC medium for 48 h at 25°C on a rotary shaker 200rpm. The fungus was collected via filtration onto a glass fiber pad using a Millipore (Conrad, MA, USA) filtration kit. The resultant *T. reesei* "cake" was immediately immersed in liquid nitrogen and was ground on a mortar and pestle until a flour-like consistency was obtained. This material was transferred to a 50 ml tube where 5ml AE buffered (50 mM sodium acetate [pH 5.3], 10 mM EDTA,) phenol (68°C) and 1/3 volume of zirconia glass beads were added. The solution was vortexed vigorously for 5 min. The suspension was incubated at 68°C for 15 min with a brief vortex every 5 min. Thereafter 5 ml of AE buffer was added and mixture was vortexed and incubated for another 15 min at 68°C. Following centrifugation at 4000rpm, the aqueous layer was transferred to a new tube and extracted with one volume of AE buffered phenol [pH 5.3]. Following centrifugation for 5 min, the aqueous layer was removed and extracted with one volume of chloroform. The RNA was precipitated using the same procedure as above for DNA precipitation (Section 2.2). The resultant pellet was resuspended in 50 µl of diethyl pyrocarbonate [DEPC]-treated H_2O. RNA levels were quantified by using the A260/280 values on a Biophotometer spectrophotometer (Eppendorf, Germany).

2.4 DNase I treatment of RNA.

RNA (150 µg) was resuspended in 1X DNase buffer (4 mM TrisHCl [pH 8], 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) and incubated with 5 units of DNase I enzyme Promega (Madison, WI, USA) in a final volume of 50 µl at 37°C for 1 h. The enzyme was

inactivated by incubation at 65°C for 10 min. RNA was ethanol precipitated as in Section 2.2 and resuspended in DEPC-H_2O.

2.5 cDNA synthesis.

DNaseI treated RNA (2 µg) was reversed transcribed into cDNA using the one-step High-capacity cDNA Reverse Transcription (RT) Kit from ABI (Applied Biosystems, Foster City, CA, USA) as per the manufacturer's instructions. The RNA template was mixed in a 0.2 ml PCR tube with 1X RT buffer, 4mM dNTP, 0.4 µM of gene specific reverse primer and 50U of reverse-transcriptase enzyme in a final volume of 20 µl. The negative control had the reverse transcriptase enzyme omitted. The reaction was carried out at 25°C for 10 min, 37°C for 120 min and 85°C for 5 sec. The cDNAs were amplified using 1 µl of cDNA as template, under the same conditions described in Section 2.6. The fragment of interest was gel extracted and purified as described below in Section 2.7.

2.6 Polymerase chain reaction (PCR).

All reagents were supplied by YorkBiosciences (York, UK) unless otherwise stated. Oligonucleotide primers were supplied by Integrated DNA Technologies (IDT) [San Jose, CA, USA) or Sigma-Aldrich (Dublin, Ireland). The oligonucleotide primers were designed *in silico* using the ApE computer program. Sequences used for primer design were chosen to ensure that oligonucleotide primer pairs had similar melting temperatures. The *T. reesei* gene sequences used to design primers were obtained on www.ncbi.nlm.nih.org. The accession numbers for *egl1*, *cbh2* and *bgl1* are M15665.1, M55080.1 and U09580.1 respectively. The gene sequence for *cbh1* was obtained from the Joint Genome Institue (http://genome.jgi-psf/Tririe2/Trire2.home.html). The oligonucleotide primers used for PCR amplification are listed in Tables 2.1 and 2.2. PCR amplification was performed in a reaction volume of 25 µl, containing 0.4 µM of each primer, 1X PCR buffer (10 mM Tris-Cl [pH 8.8], 50 mM KCl, 1.5 mM $MgSO_4$, 0.1% Triton X-100), 0.06 µM of each deoxynucleoside triphosphate (dNTPs); dATP, dCTP, dGTP and dTTP, 2.5units of Verizyme™ polymerase and 200 ng of template DNA. The Verizyme PCR cycle used to amplify the DNA was 95°C for 1 min followed by 40 cycles of 95°C for 30sec, 67°C for 40sec and 70°C for 2min and a final elongation hold of 70°C for 12min. In some initial experiments the enzyme Taq DNA polymerase was

used, however this was discontinued due to the high error rate and replaced by the Verizyme enzyme which has a 3' to 5' proof-reading function. PCR was performed using a Robocycler Gradient 40 (Stratagene, CA, USA) machine

To connect PCR products sharing overlapping sequence homology the Allele cycle PCR was used (Erdeniz et al., 1997). Purified PCR products (100ng) with sequence overlap were mixed and used as the template for the PCR. The reaction cycle was 94°C for 3 min, followed by 10 rounds of 94°C for 30 sec, 60°C for 15 sec and 70°C for 4 min. This was then followed by 20 cycles of 94°C for 30sec, 67°C for 15sec and 70°C for 4min. The reaction ended with a final elongation step of 70°C for 5min. A 5 µl sample of each PCR product was electrophoresed on 1% (w/v) agarose gel in TAE buffer (40 mM Tris, 20 mM glacial acetic acid, 2 mM EDTA [pH 8]) containing ethidium bromide (1.25 µg ml^{-1}) and visualised on a UV transilluminator at 302 nm. DNA size ladders were included in each gel, either Hyperladder I (200bp-10000bp) or Hyperladder II (50bp-2000bp) (Bioline, Kilkenny, Ireland).

2.7 Gel extraction and purification.

Following electrophoresis, DNA bands were excised from the agarose gel using a blade. The agarose plug was transferred to a centrifuge tube and mixed with 500 µl of gel dissolving solution (5.5 M guanidine thiocyanate, 10 mM Tris-HCl [pH 7.5]). The tube was heated to 55°C until the agarose melted. Following vortexing, the agarose solution was transferred to a silica resin spin column (Epoch LifeScience, Texas, USA). The column was centrifuged at top speed for 1 min. The column was washed with 500 µl of wash buffer (60% (v/v) ethanol, 60 mM potassium acetate, 10 mM Tris-HCl [pH 7.5]). Following centrifugation at top speed the column was transferred to a new collection tube and centrifuged again to dry the column. The purified PCR product was then eluted in 20 µl of sterile distilled water.

2.8 Transformations.

Yeast transformations were carried out using the lithium acetate method (Jansen et al., 2005) with the following modifications. Yeast cells were cultured for 24 h in YPD; 2 ml of culture was centrifuged at top speed, and the yeast pellet washed in sterile distilled water. The pellet was then washed in 1 ml of 0.1 M lithium acetate.

Following centrifugation at top speed, the pellet was washed again in 0.1 M lithium acetate. The transformation mix consisted of 33% polyethylene glycol (PEG) 3300, 0.1 M lithium acetate and 100 µg of single stranded salmon sperm DNA. PCR products representing the gene of interest were mixed with linearised pGREG vector in a molar ratio of 10:1. Approximately 0.5 µg of vector was used in each transformation and based on this value, the molar ratios for inserts were calculated. Insertion of the PCR products into the vector was achieved via *in vivo* homologous recombination activity of the yeast.

The yeast cell pellet was mixed with 360 µl of transformation mix and was subjected to heat shock at 42°C for 60 min. Following the incubation, the cells were allowed to recover in pre-warmed YPD broth at 30°C for 3-6 h. *S. cerevisiae* and *S. pastorianus* transformants were selected on SC ura$^-$ and YPD containing 0.2 g l^{-1} G418 agar plates respectively which were incubated at 30°C for 48 h.

Transformations into *E. coli* strain XL1 for plasmid propagation were carried out using the electroporation method of the Fred Hutchinson Cancer Research Centre (http://www.fhcrc.org/science/labs). Total genomic DNA was extracted from yeast cultures as described in Section 2.2. An aliquot of 5 µl of DNA was mixed with 45µl of electrocompetent *E. coli* and chilled on ice for 30 min. The mixture was then transferred to a pre-chilled electroporation cuvette (YorkBio, York, UK) and subjected to a 2.5V charge for a time constant of between 4 to 5 sec. The cells were allowed to recover in pre-warmed L-broth at 37°C for 60 min. Transformants were selected for by growth on L-agar plates supplemented with 25 µg ml^{-1} of carbenicillin at 37°C overnight.

2.9. pGREG vector

The plasmid used throughout this study was the multi-purpose pGREG vector. The pGREG vectors are based on the pRS plasmid series (Jansen et al., 2005). The plasmid contains a galactokinase, GAL1 promoter flanked by SpeI and SacI restriction sites. Immediately downstream of the promoter is a 6xHIS tag sequence followed by a *HIS3* gene which is flanked by two recombination sites (rec1 and rec2). The introduction of DNA sequences with homology to the rec sites at the end of PCR fragments permits the cloning of genes into the SalI

linearised vector via *in vivo* homologous recombination. The pGREG plasmid contained a standard *bla* gene for ampicillin resistance but also an additional kanMX cassette which encodes for resistance to kanamycin. The vector also has a URA3 marker. The specific plasmid used in this study was pGREG586 which is 8.3kb in size. This will be referred to as pGAL1 throughout the rest of the thesis (Fig 2.1). A valuable asset of the pGREG vector is the ability to conveniently swap promoters. The GAL1 promoter is flanked by SacI and SpeI restriction sites. Cutting the pGAL1 plasmid with these enzymes will release the GAL1 promoter. Any promoter DNA sequence amplified with primers with homology to the flanking regions of the vector can be inserted by homologous recombination.

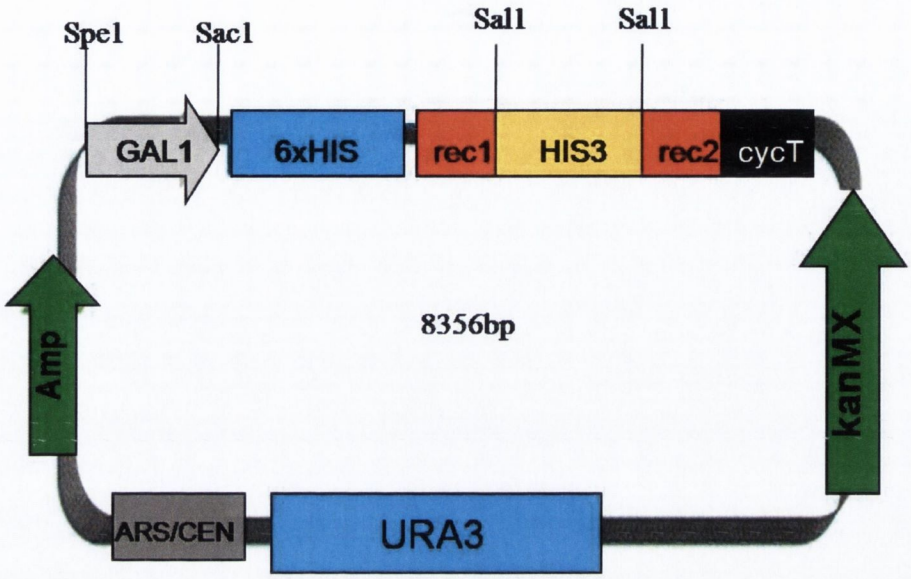


FIG 2.1. Schematic diagram of pGREG vector. GAL1 promoter: silver arrow. Rec1 and rec2 recombination site: red boxes. *HIS3* gene: yellow. Cyc terminator: black box. Antibiotic resistant markers: green arrows. ARS/CEN origin or replication: grey box. The *URA3* gene and 6xHIS tag are shown in blue. Additional tag and marker highlighted: sky blue. Restriction sites flanking the promoter and *HIS3* gene are shown. Figure adapted from Jansen et al., (2005).

2.10 Screening of yeast transformants.

Yeast transformants were screened for the presence of inserted DNA sequences via direct colony PCR. Colonies transformed by the addition of the vector were lysed by mixing with 10 µl of 20 mM NaOH and heating to 95°C for 10 min. A 1 µl volume of the lysed cell content was used for a PCR reaction as described above (Section 2.6) with fragment-specific and plasmid-specific primers (Tables 2.1 and 2.2).

2.11 Plasmid Extraction.

Yeast transformants were cultured in either SC ura$^-$ or YPDG418 media for 48h. Total DNA was extracted using the standard DNA extraction method as described above (Section 2.2). Plasmid DNA was extracted from *E. coli* following overnight cultivation in L-broth containing 25 µg ml^{-1} of carbenicillin, using the Sigma Genelute Plasmid Mini-Prep Kit (Sigma-Aldrich, Dublin, Ireland) as per the manufacturer's instruction. A full list of plasmids and strains generated is presented in Tables 2.4 and 2.5.

2.12 DNA Sequencing

Each plasmid generated in the project was sequenced by GATC (Konstadt, Germany). In each case, plasmid DNA was sequenced with gene-specific primers to ensure that the insert was present. The generated sequence was compared *in silico* to the wild type sequence using the ApE computer program. A list of sequenced inserts is available in the Appendix 1 (attached CD). Unless otherwise stated; all *egl1* inserts were sequenced using the primers egl1_F1, egl1_R1, egl1_F2, egl1_R2 and egl_R5. The *cbh2* inserts were sequenced using the primers cbh2_F1, cbh2_R1, cbh2_F2, cbh2_F3, cbh2_R4 and cbh2_R6. The *bgl1* insert was sequenced using the primers bgl1_F1, bgl1_R3, bgl1_F5, bgl1_R5, and bgl1_R12. Finally *cbh1* was sequenced using the primers cbh1_F1, cbh1_R1, cbh1_F2 and cbh2_R2 (Tables 2.1 and 2.2)

2.13 Promoter swap

Approximately 1 µg of pGAL1 was digested overnight at 37°C with 10 units each of SacI and SpeI to remove the GAL1 promoter (See Figure 2.1). The linearised

plasmid was gel extracted and purified as described in Section 2.5. The PGK1 promoter region was PCR amplified from *S. cerevisiae* DNA using sequence-specific primers which additionally contained sequences homologous to the regions flanking the SacI and SpeI sites (Table 2.3). PGK1 insert and linearised vector were mixed in a 10:1 ratio and transformed into *S. cerevisiae*. Transformants were selected on SC ura$^-$ plates. The DNA sequence of the recovered plasmid was obtained to confirm that the PGK1 promoter had been successfully recombined.

2.14. Concentration of culture supernatant.

Yeast culture supernatants were separated from cell pellet by centrifugation. The supernatants were then filtered through a 0.2 μM PVDF membrane to remove any remaining yeast cells. The supernatants were then transferred to a Amicon Ultra- 15 centrifugal filter unit spin column (Millipore, MA, USA) with a cut off of 3kDa and centrifuged at 4000 rpm at 4°C. The use of this column ensured that all molecules greater than 3kDa in size were retained. Supernatants were approximately 40-fold more concentrated relative to the starting sample volume.

2.15. Electrophoresis and Western immunoblotting.

Protein samples were examined by SDS-PAGE. The gel consisted of two phases; the upper stacking and lower resolving gels. The stacking gel consisted of: 5% (v/v) acrylamide, 0.126 M Tris[pH 6.8], 0.1% SDS (v/v), 0.1% ammonium persulfate (APS) and 0.1% TEMED in a final volume of 5ml. The resolving gel was composed of 10% (v/v) acrylamide, 0.375 M Tris [pH 8.8], 0.1% SDS (v/v), 0.1% ammonium persulfate (APS) and 0.04% TEMED. One volume of sample was boiled in one volume of loading buffer (0.2 M Tris-HCl [pH 6.8], 8% (w/v) SDS, 0.4% (v/v) glycerol, 0.4 M DTT and 0.04% (w/v) bromophenol blue) for 15 min. A 40μl aliquot of protein sample was loaded per well. The gel was electrophoresed at 100V for 90 min in Running buffer, 25 mM Tris, 0.187 M glycine and 0.1% (v/v) SDS. Gels were stained in Staining solution (50% (v/v) methanol, 10% acetic acid, 0.1% (Coomassie blue) and destained with Destain solution [40% (v/v) methanol, 7% (v/v) acetic acid].

For Western blots, proteins were transferred to a nitrocellulose membrane using 100V for 1 h in Transfer buffer: 48 mM Tris, 39 mM glycine, 0.037% (w/v) SDS and 20% (v/v) methanol. The membranes were blocked in Blocking buffer (5% (w/v) Marvel, 50 mM Tris [pH7.5] and 150 mM NaCl). Blocking was carried out at 4°C overnight. The primary antibody was monoclonal mouse anti-poly HIS (1/3000) (Sigma-Aldrich, Dublin, Ireland) which was added with Blocking buffer and incubated with membrane overnight at 4°C. The membrane was washed several times with Wash buffer (50 mM Tris [pH7.5], 150 mM NaCl and 0.1% Tween-20). The membrane was incubated with secondary antibody goat anti-mouse IgG conjugated with horseradish peroxidase (1/10000) in Blocking buffer for 1hr at room temperature with shaking. The membrane was washed as before and then washed with chemiluminence reagent (100 mM glycine [pH10], 0.4 M luminal and 8 mM iodophenol). The membrane was then exposed to X-ray film and developed.

2.16. ELISA

An indirect ELISA method was performed to detect HIS-tagged proteins present in concentrated supernatant samples. A small 11.2kDa HIS-tagged bacterial protein FIS was used as a positive control. A 50µl volume of a 1X carbonate buffer (15mM Na_2CO_3, 34 mM NaHCO3, [pH9.5] was added to separate wells of a Nunc Maxi-sorp 96 well plate (Nalge Nunc International). The same volume of protein sample was added to buffer containing wells. The protein was fixed overnight at 4°C. The wells were then washed several times with phosphate-buffer saline (PBS); 50mM KH_2PO_4 and 150 mM NaCl pH7.4. Addition of 0.05% Tween-20 to PBS resulted in PBST. The wells were blocked with Blocking buffer: 1% (w/v) bovine serum albumin (BSA) dissolved in PBST and incubated at room temperature with shaking for 2 hr. The wells were washed several times with PBST. A primary monoclonal mouse anti-poly HIS antibody diluted (1/3000) in Blocking buffer was added to the wells and incubated at 4°C overnight. The wells were washed several times with PBST. The secondary goat anti-mouse IgG conjugated with horseradish peroxidase antibody diluted (1/10000) in Blocking buffer was added to the samples and incubated for 2h at room temperature for 30min. Following further rounds of washing with PBST, each well was inoculated with 50 µl of 3,3′,5,5′-

tetramethylbenzidine (TMB). When sufficient colour development occurred the plate was scanned on a spectrophotometer at 620 nm. To intensify the colour change 50 µl of 4N H_2SO_4 was added to each and the plate was re-read at 540 nm.

2.17 Endoglucanase enzyme activity (Carboxymethylcellulose agar plate)

Yeast transformed with the *egl1* gene was grown in media containing galactose or glucose depending on whether the plasmid had the GAL1 or PGK1 promoter, for 48 h at 30°C. The culture was centrifuged and the supernatant was retained. The supernatant was filtered through a 0.2 µm PVDF membrane to remove any remaining yeast cells. A 20 µl volume of supernatant was spotted onto a 0.1% (w/v) carboxymethylcellulose (CMC) agar plate. After incubation at 30°C for 48 h, the plate was stained in 0.1% (w/v) Congo Red for 1 h and destained with 1 M NaCl. Endoglucanase activity was indicated by a zone of hydrolysis.

2.18. BGL1 activity assay

2.18.1 X-glucosidase agar assay.

X-glucoside (X-glu), also known as 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside is a chromogenic reagent used for detecting β-glucosidase activity. Yeast transformed with the *bgl1* insert was grown at 30°C for 48 h (Section 2.1). The culture was centrifuged. The supernatant was concentrated as described in Section 2.14. The yeast cell pellet was resuspended in 1 ml of 0.05 M citrate buffer (citric acid adjusted to pH 4.8 with NaOH). A 1/3 volume of zirconia beads was added to the resuspension which was vortexed vigorously for 20 min. The lysate was then centrifuged at 14500 rpm for 15 min. The supernatant was retained. Concentrated supernatants or cell lysates from cells were spotted onto a 1% (w/v) agar plate containing 40 µg ml^{-1} of X-glu (Melford, Suffolk, UK) dissolved in dimethyl sulfoxide (DMSO) 100% (v/v). The plate was incubated for 48 h at 30° C. The development of a blue halo indicated β-glucosidase activity.

2.18.2 X-glucoside liquid assay for BGL1 activity.

To semi-quantify BGL1 activity a liquid X-glucoside assay was devised. A volume (100 µl) of supernatant from yeast secreting the BGLI enzymes were mixed with an equal volume of 40 µg ml^{-1} X-glu in 0.05 M citrate buffer (0.05 M citric acid, pH

4.8). The solution was transferred to a well of a 96 well plate. The plate was incubated at 50°C. Activity was indicated by the presence of a blue hue in the reaction solution. The intensity of the blue colour was quantified by measuring the absorbance at 620 nm using a Multiskan Ascent spectrophotometer (ThermoLabsystems).

2.19 Generation of phosphoric acid swollen cellulose (PASC)

PASC was prepared as described by (Zhang et al., 2006) with modifications. Avicel PH-101 (0.2 g) was mixed with 0.6 ml of distilled water followed by the gradual addition of 5 ml of 85% phosphoric acid (14.8 M) to the cellulose. The mixture was vortexed. Another 5 ml of phosphoric acid was added followed by further mixing. The solution was kept at 4°C overnight. A volume of 20 ml of ice-cold distilled water was added followed by intense mixing. A further 20 ml of water was added. The solution was centrifuged at 3500rpm for 15 min at 4°C. The supernatant was removed and the pellet was washed with 40 ml of ice cold water. The mixture was centrigued at 3500 rpm for 15 min at 4°C. The pellet was washed twice more with ice cold distilled water. An aliquot (0.5 ml) of 2 M sodium carbonate and 45 ml of distilled water were added to mixture. The mixture was centrifuged and washed with ice cold water until a pH of 5-7 was obtained. The swollen cellulose solution was then autoclaved.

2.20 Assay for CBHII activity

Yeast clones containing the *cbh2* gene were cultured for 48 h (Section 2.1). The supernatant of the culture was retained and filtered through a nylon membrane of pore size 0.45 μm. PASC (1g) was added to supernatant and solution was left stirring overnight at 4°C. The purpose of this step was to allow extracellular CBHII enzyme to adsorb to the cellulose.

The solution was filtered through a nylon membrane. The cellulose was retained and transferred to a 15 ml tube where it was resuspended in 6 ml of 0.05 M citrate buffer. The tube was mixed thoroughly and then incubated at 50°C. Aliquots of mixture were removed at specified times and the level of total sugar was assayed via the phenol sulphuric method (Section 2.21).

2.21 Phenol-sulfuric method for estimating total sugar.

This assay is a modified version of the method found in Wood and Bhat (1988). A known volume (250 µl) of supernatant reacted with PASC was pipetted into a 2 ml centrifuge tube. The same volume of 5% (v/v) phenol was added to the tube. The solution was briefly mixed by vortexing followed by the addition of 1.25 ml of concentrated sulphuric acid (98%). The solution was mixed immediately and allowed to cool. A 100 µl volume of solution was pipetted into a well in a 96-well plate and the absorbance was read at 492 nm. Triplicates of each sample were analysed. The absorbance values were translated into glucose or cellobiose equivalents using a standard curve graph obtained by plotting known amounts of cellobiose, ranging between 10-200 µg, against absorbance at 492nm.

2.22 Fermentations on PASC

2.22.1 Control fermentation in carbohydrate rich media

A control fermentation was performed with yeast strains S150 and C10-51 containing empty PGK1 plasmid in either, SC 10% (w/v) glucose and SC 10% (maltose) (Section 2.1). The purpose of this was to examine the fermentation patterns of the two yeast species under different fermentation conditions. S150 and C10-51 were pre-cultured in SC glucose for 48 h. The cell pellets were retained and the number of cells within was determined with a haemocytometer. Approximately $1.5x10^8$ cells each from S150 and C10-51 were inoculated into 10 ml of each fermentation media to give a final concentration of $1.5x10^7$ cells ml^{-1}. An anaerobic environment was maintained by adding a layer of mineral oil on top of each fermentation mixture. The tubes are allowed to ferment for 10 days at 25°C. Aliquots of each mix were extracted and the alcohol was measured either in directly using a refractometer or indirectly using the ADH enzymatic assay. The specific gravity decrease during fermentation was estimated from the Brix values obtained from the refractometer. Brix must be converted to specific gravity using a formula as described by (James et al., 2008). The alcohol could then be determined by this formula:

% Alcohol = [(Initial specific gravity-final specific gravity)*(0.1281)*1000]

2.22.2 Fermentation of PASC

Yeast strains S150 and C10-51 producing either EGI, CBHII or BGLI from the PGK1 promoter were separately pre-cultured for 96 h in SC ura$^-$ media. The supernatant and the cell pellet of each strain was separated and retained. Each supernatant was concentrated 80-fold using a Amicon Ultra-15 centrifugal filter unit spin columns (Millipore, MA. USA) with a cut off of 3000 Da. This was performed to remove any unmetabolized sugars or ethanol produced during growth. The same cells from the pre-cultures were used for the subsequent fermentation. The number of cells in each pellet was determined using a haemocytometer. Approximately $0.5x10^8$ cells each from S150 or C10-51 strains expressing either EGI, CBHII or BGLI were pooled together in 10 ml to give a co-culture with a final concentration of $1.5x10^7$ cells/ml. The pooled concentrated supernatants from the pre-cultures were added to their respective yeasts, i.e. concentrated supernatant from C10-51 clones was added to C10-51 yeast. The 10X SC PASC fermentation medium was added to each reaction to leave a final reaction volume of 10 ml. All fermentations were performed as 10 ml cultures in 15 ml tubes (Sarstedt, Wexford, Ireland). An anaerobic environment was maintained by adding a layer of mineral oil on top of each fermentation mixture. The tubes were allowed to ferment at 25°C for 10 days. Aliquots of each mix were extracted for analysis at allotted time points.

2.23 Alcohol dehydrogenase assay for ethanol concentration

Aliquots (15 µl) of fermentation supernatant taken at time intervals were mixed with 400 µl of reaction mixture (0.5 M Tris, 65 mM succinic acid, 5 mM EDTA, 25 mM nicotinamide adenine dinucleotide (NAD), 6.4 IU alcohol dehydrogenase (ADH). The mixture was incubated at 30°C for 8 min. The absorbance resulting from the reduction of NAD to NADH was measured at 340 nm on a spectrophotometer. The absorbance values were compared to a standard curve plot of known concentrations of ethanol (Sigma-Aldrich, Dublin, Ireland) ranging between 0-0.8 g l^{-1}.

Table 2.1 Oligonucleotide primers used for amplification of *T. reesei bgl1*

Primer name	Oligonucleotide sequence (5'→3') [a,b,c]	Nucleotide position in gene [c,*]
bgl1_F1	**GAATTCGATATCAAGCTTATCGATACCGTCGCA** ATGCGTTACCGAACAGCAGCTGCGC	1-25
bgl1_R1	TGACTGTCTGCCCTAGCAAAGG	44-65
bgl1_F2	CCCTTTGCTAGGGCAGACAGTCACTC	43-65,135-137
bgl1_R2	GCCGAGGCCCCCGATGTTGAGTGAC	62-65,135-137
bgl1_F3	CTCAACATCGGGGGCCTCGG	135-154
bgl1_R3	GGTCCGTCTTGAAGGCATAGC	261-281
bgl1_R4	GCAGGCCCAGGTGGTATTGACC	780-801
bgl1_F4	TGCCATGGGTCAAACCATCAACGGC	579-603
bgl1_R5	CCATATCGTCGACTCTGCTCGTGGGG	1002-1027
bgl1_F5	AATAGCAATCAGGTCCCCACG	988-1008
bgl1_R6	GCCTTTGTCGTTGCACGAGGGCG	1256-1278
bgl1_F6	ACCACAAGACCAATGTCAGGGCAATTGCCAGGGACG GCATCGTTCTGCTCAAGAA	1115-1169
bgl1_R7	TTCTTGAGCAGAACGATGCCGTCCCTGGCAATTGCC CTGACATTGGTCTTGTGGT	1115-1169
bgl1_F7	GCCAACATCCTGCCGCTCAAGAAGCC	1174-1199
bgl1_R8	ACAGTCCATAGCCGAACTCGT	1874-1894
bgl1_F8	ACGAGTTCGGCTATGGACTGTCTTAC	1874-1894, 1959-1963
bgl1_R9	CGTGAGTAGTTGAACTTGGTGTAAGACAGT	1890-1894, 1959-1983
bgl1_F9	CTTACACCAAGTTCAACTACTCACGC	1959-1984
bgl1_F10	CCTGGTCAGAGCGGAACAGCAACGTTCAACATCCGACGAC	2207-2246
bgl1_R11	GCTGAGATCTCGTCGTCGGATGTTGAACGTTGCTGTT	2221-2257
bgl1_R12	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** CTACGCTACCGACAGAGTGCTCG	2346-2368
bgl1_R13	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** CTACGCTACCGACAGAGTGCTCGTCAGCCTGATATCCCG	2330-2368

[a] The sequences homologous to the pGREG plasmid sequence are highlighted in bold

[b] Genomic DNA nucleotide sequence

[c] Reverse primers are in reverse complement

* cDNA sequences will differ to the genomic DNA because of the absence of the introns, therefore oligonucleotide primer location in cDNA will be different.

Table 2.2 Oligonucleotide primers used to amplify *egl1*, *cbh1* and *cbh2*

Primer name	Oligonucleotide primer sequence (5'→3') [a,b,c]	Nucleotide position in corresponding gene [c]
egl1_F1	**GAATTCGATATCAAGCTTATCGATACCGTCGCA** ATGGCGCCCTCAGTTACAC	1-19
egl1_R1	TGCACGAGTTGTAGTTTGCGTCG	192-214
egl1_R2	TGGGCGCTGGGGATGTCGACGCC	965-987
egl1_F2	AACTCGAGGGCGAATGCCTTGACC	673-696
egl1_R3	GGTATCTCCGGGGCCGTAGTAGCTGGAAAACCATGGTT AGTATCAGCATGGAAAACCG	812-862
egl1_F3	CTTTCCTGGCGCTCTCGCGGTTTTCCATGCTGATACTA ACCATGGTTTTCCAGCTACT	795-845
egl1_R4	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** CTAAAGGCATTGCGAGTAGTCTGTAGTTAGTATTTTAGC GGT	1480-1507
egl1_R5	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** CTAAAGGCATTGCGAGTAGTCTG	1488-1507
egl1_R6	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** *CTAGTGATGGTGATGGTGATG*AAGGCATTGCGAGTAGTA GTCGT	1488-1504
cbh1_F1	**GAATTCGATATCAAGCTTATCGATACCGTCGCA** ATGTATCGGAAGTTGGCCG	1-19
cbh1_R1	CGGTATGGGTTCCAGTCGCAGCC	897-919
cbh1_F2	ATCTGCGAGGGTGATGGGTGCG	828-849
cbh1_R2	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** TTACAGGCACTGAGAGTAGTAAGG	1653-1676
cbh2_F1	**GAATTCGATATCAAGCTTATCGATACCGTCGCAA** ATGATTGTCGGCATTCTCACC	1-22
cbh2_F2	AACATCCCGGTCGAGCTCCGC	301-321
cbh2_R1	GCTAACTTCAGAGGCGTAATATGC	376-349
cbh2_R2	GGAATATTCCGATATCCGGACC	798-819
cbh2_F3	TGGCGAATACTCTATTGCCGATGG	723-746
cbh2_F4	AAGCTGTACATCCACGCTATTGGACGTCTTCTTGCCAAT CA	1246-1286
cbh2_R3	GGACCAGCCGTGATTGGCAAGAAGACGTCCAATAGCGT GGA	1256-1286
cbh2_R4	GGAAAGCAGCCTACCGGACAGC	1330-1351
cbh2_F5	GCTCCCAGATGCCTTGCAACCGGCGGCTCAAGCTGGTGC	1515-1553
cbh2_R5	GCTTGGAACCAAGCACCAGCTTGAGCCGCCGGTTGCAA GGC	1525-1565
cbh2_R6	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** TTACAGGAACGATGGGTTTGCG	1590-1611
cbh2_R7	**GCGTGACATAACTAATTACATGACTCGAGGTCGAC** *CTAGTGATGGTGATGGTGATG*CAGGAACGATGGGTTTGCG	1590-1608

[a] The sequences homologous to the pGREG plasmid sequence are highlighted in bold

[b] Italics represent the HIS tag sequence

[c] Genomic DNA nucleotide sequence

Table 2.3. Oligonucleotide primers used integration[a] and promoter swap[b]

Primer name	Primer sequence (5'→3')[a]
[a]I-GAL_F	GCGGGTATTTTGATGGTAAATCTACAAGCCCTCGG**TAGTACGGATTAGAAGCCGCCG**
[a]I-Kan_R	TAAGTAACTTTTTATGACCATGGTAATAATTAAAG**GGAAAAGACTCACGTTTCGAGGCC**
[a]Kan_F	**CGTTTTCGACACTGGATGGCGGCG**
[a]Kan_R	**GGAAAAGACTCACGTTTCGAGGCC**
[a]I_F1	TTATCCAGGAGCGCCTCTAA
[a]I_R1	CTACTAGGCGATGCCGCTAC
[a]I_F2	TGTGTTTGCTCGCTCATCTT
[a]I_R2	TATTCGTCGGTGTTCCTTCC
[a]YPR159_F	CTGAACGATGACGACAATGG
[b]PGK1_F	**AGGGAACAAAAGCTGGAGCTCGTTTAAACGGCGCGCC**CCAAGAATTACTCGTGAGTAAGG
[b]PGK1_R	**GTGATGCGATCCTCTCATACTAGTGCGGCCGCTCTAGA**CCGCGGTGTTTTATATTTGTTG

The sequences homologous to the pGREG plasmid sequence are highlighted in bold

Table 2.4 *S. cerevisiae* S150 strains generated in study[a,b].

Strain name	Plasmid name	Plasmid Source
S150GAL1	pGAL1	Jansen *et al* [c]
S150PGK1	pPGK1	Present study
S150GAL1egl1gDNA	pGAL1egl1gDNA	Present study
S150PGK1egl1 ^{TACTAAC}	pPGK1egl1 ^{TACTAAC}	Present study
S150GAL1egl1cDNA $^{T1360A::G1361}$	pGAL1egl1cDNA $^{T1360A::G1361}$	Present study
S150PGK1egl1cDNA $^{T1360A::G1361}$	pPGK1egl1cDNA $^{T1360A::G1361}$	Present study
S150PGK1egl1HIS cDNA ^{wt}	pPGK1egl1HIScDNA ^{wt}	Present study
S150GAL1cbh2gDNA	pGAL1cbh2gDNA	Present study
S150GAL1cbh2cDNA $^{G1076C, G1345C}$	pGAL1cbh2cDNA $^{G1076C, G1345C}$	Present study
S150PGK1cbh2 cDNA $^{G1076C, G1345C}$	pPGK1cbh2cDNA $^{G1076C, G1345C}$	Present study
S150PGK1cbh2cDNA ^{A896C}	pPGK1cbh2cDNA ^{A896C}	Present study
S150PGK1cbh2 cDNA ^{G142C}	pPGK1cbh2cDNA ^{G142C}	Present study
S150PGK1cbh2HIS cDNA ^{wt}	pPGK1cbh2cDNAHIS ^{wt}	Present study
S150GAL1bgl1gDNA	pGAL1bgl1gDNA	Present study
S150GAL1bgl1gDNA ^{recombined}	pGAL1bgl1gDNArecombined	Present study
S150GAL1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$	pGAL1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$	Present study
S150PGK1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$	pPGK1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$	Present study
S150PGK1bgl1 cDNA $^{C64A, A65G, \Delta 1521-2098}$	pPGK1bgl1cDNA $^{C64A, A65G, \Delta 1521-2098}$	Present study
S150PGK1bgl1cDNA ^{wt}	pPGK1bgl1cDNA ^{wt}	Present study
S150GAL1cbh1cDNA ^{wt}	pGAL1cbh1cDNA ^{wt}	Present study
S150PGK1cbh1cDNA ^{wt}	pPGK1cbh1cDNA ^{wt}	Present study

[a] Table legend: p: plasmid.

[b] Superscript indicates missence mutations, letter in italic indicates wildtype base, number denotes nucleotide position and letter in bold indicates mutant nucleotide

:: denotes an insertion of a nucleotide base

Δ denotes a deletion event

wt: indicates that wildtype coding sequence present, ie no missense errors.

[c] (Jansen et al., 2005)

Note strains grouped together depending on which cellulase gene they contain.

Table 2.5 *S. pastorianus* strains C10-51 generated in study

Strain name [a,b]	**Plasmid name**	**Plasmid Source**
C10-51GAL1	pGAL1	Jansen *et al*
C10-51PGK1	pPGK1	Present study
C10-51GALegl1cDNA $^{T1360A::G1361}$	pGAL1egl1cDNA $^{cDNA\ T1360A::G1361}$	Present study
C10-51PGK1egl1cDNA $^{T1360A::G1361}$	pPGK1egl1cDNA $^{cDNA\ T1360A::G1361}$	Present study
C10-51PGK1 egl1HIS cDNA ^{wt}	pPGK1egl1cDNAHIS ^{wt}	Present study
C10-51GAL1egl1cDNA $^{T1360A::G1361integrated}$	putative-integratedGAL1eg1cDNA $^{T1360A::G1361integrated}$	Present study
C10-51PGK1cbh2HIScDNA ^{wt}	pPGK1cbh2cDNAHIS ^{wt}	Present study
C10-51PGK1bgl1cDNA ^{wt}	pPGK1bgl1cDNA ^{wt}	Present study

[a] Table legend: p: plasmid.

[b] Superscript indicates missence mutations, letter in italic indicates wildtype base, number denotes nucleotide position and letter in bold indicates mutant nucleotide

:: denotes an insertion of a nucleotide base

Δ denotes a deletion event

wt: indicates that wildtype coding sequence present, ie no missense errors.

Note strains grouped together depending on which cellulase gene they contain.

Chapter 3: Generation of cellulase encoding yeast strains.

3.1 Introduction:

The overall aim of the experiments in this chapter was to bioengineer yeast strains to express cellulase genes. Studies in the literature have shown many examples of *T. reesei* genes being cloned and expressed in *Saccharomyces cerevisiae*. (See Chapter 1, Section 1.12). A novel aspect of this current project was the intent to engineer industrial brewing yeast species *S. pastorianus* with cellulase encoding genes. *S. pastorianus* is a robust fermenter and has been demonstrated to be highly stress tolerant (James et al., 2008). For bioethanol to be produced on an industrial scale a more stress tolerant yeast may be desirable (as described in Chapter 1, Section 1.14). This chapter described the strategies used for the cloning and expression of cellulase encoding genes into two species of yeast *S. cerevisiae* strain S150 and *S. pastorianus* C10-51. Experiments were carried out in parallel to create cellulase encoding cassettes of *egl1*, *cbh2* and *bgl1* using either *T. reesei* genomic DNA (gDNA) or complementary DNA (cDNA). Genomic DNA contains introns that must be removed by splicing if correct gene expression is to occur. The aim of transforming *T. reesei* genomic DNA inserts into yeast was to investigate if the host *Saccharomyces* splicing machinery could recognise and properly remove foreign introns to enable gene expression.

The final aim of the experiments described in this chapter was to express cellulase encoding genes from both inducible and constitutive promoters. The *T. reesei* genes were transformed into yeast via the pGREG expression vectors (Jansen et al., 2005). The pGREG vectors contained the powerful but highly regulated galactokinase, GAL1, promoter. This promoter is induced only upon uptake of galactose and is repressed by the presence of glucose in the growth medium. In order for the transformed yeast to utilize and ferment cellulose, the inserted cellulase encoding genes must be constantly producing cellulase enzymes, therefore, the GAL1 promoter of pGREG was replaced with the constitutively expressed phosphoglycerate kinase, PGK1, promoter of *S. cerevisiae*.

3.2. Results

3.2.1 Identification of *T. reesei* cellulase encoding genes

A gene from each of the three cellulase class was chosen for cloning (Chapter1, Section 1.13). Previous studies have demonstrated that *T. reesei* cellulases can be expressed in yeast, Chapter 1, Section 1.12. Studies had indicated that the first cellulase required to initiate cellulose degradation is endoglucanase (Penttïla et al., 1987; Saloheimo et al., 1997). As summarised in Chapter 1 Table 1.2, *T. reesei* has at least eight genes encoding five different endoglucanase enzymes. The major endoglucanase enzyme was EGI and accounted for 10% of all cellulase produced by *T. reesei* under inducing conditions (Takashima et al., 1998). EGI, encoded by the *egl1* gene, was chosen as the endoglucanase for this study. Penttïla et al., (1988) studied the expression of recombinant *T. reesei* cellohydrolases CBHI and CBHII in *S. cerevisiae*. The study indicated that CBHII produced higher activity than CBHI. The *cbh2* gene of *T. reesei*, which encodes CBHII enzyme, was included in this project. The chief β-glucosidase of *T. reesei* is BGL1 encoded by the gene *bgl1*; thus, this gene was chosen for this study (See Chapter 1, Section 1.13).

Based on *in silco* studies (www.ncbi.nlm.nih.org; Accession number: U09580.1), the size of the *bgl1* gene was 2368bp from the start codon of its signal peptide to the stop codon of the gene (Fig 3.1.A). The signal peptide sequence is encoded in the first 162bp, and is interrupted by a 69bp intron between base pair 66-134. The mature protein-coding region begins at 163bp and is 2206bp in length. This region contains a 64bp intron at base pairs 1895-1958. The gene encodes a BGLI protein of molecular weight of approximately 75.3kDa, containing 713 amino acids (aa) (Miettinen-Oinonen, 2004).

The *cbh2* gene (NCBI accession number: M55080.1) is 1611bp in length (Fig. 3.1.B) including the signal peptide encoded in the first 72bp long. The protein-coding region (73bp to 1536bp) includes three introns of 49, 56 and 90bp located at 93-141, 529-584 and 833-922bp respectively. The mature CBHII protein is approximately 50 kDa in size and contains and 447 aa (Miettinen-Oinonen, 2004).

The *egl1* gene (NCBI accession number: M15665.1) is 1507bp in length and contains the sequences for the signal peptide in the first 66bp (Fig 3.1.C). The mature protein-coding region is 1441bp in length and contains two introns of 70 and 57bp at base pairs 771-840 and 1431-1487 respectively. Excluding the

introns, the mature EGI encodes a protein of 437a.a with a molecular weight of 50-55 kDa depending on glycosylation (Miettinen-Oinonen, 2004).

Using publicly available DNA sequences (www.ncbi.nlm.nih.org) oligonucleotide primers to amplify these three genes were designed using the computer program ApE. The list of primers are summarised in Chapter 2, Tables 2.1 and 2.2. In each case, the sequence of the gene to be amplified starts with the native secretory signal peptide and ends at the stop codon of the mature protein-coding region (See Fig 3.1A-C). The decision to include the native signal peptide was taken to investigate if an expressed cellulase enzyme within a transformed yeast cell would be correctly processed and secreted.

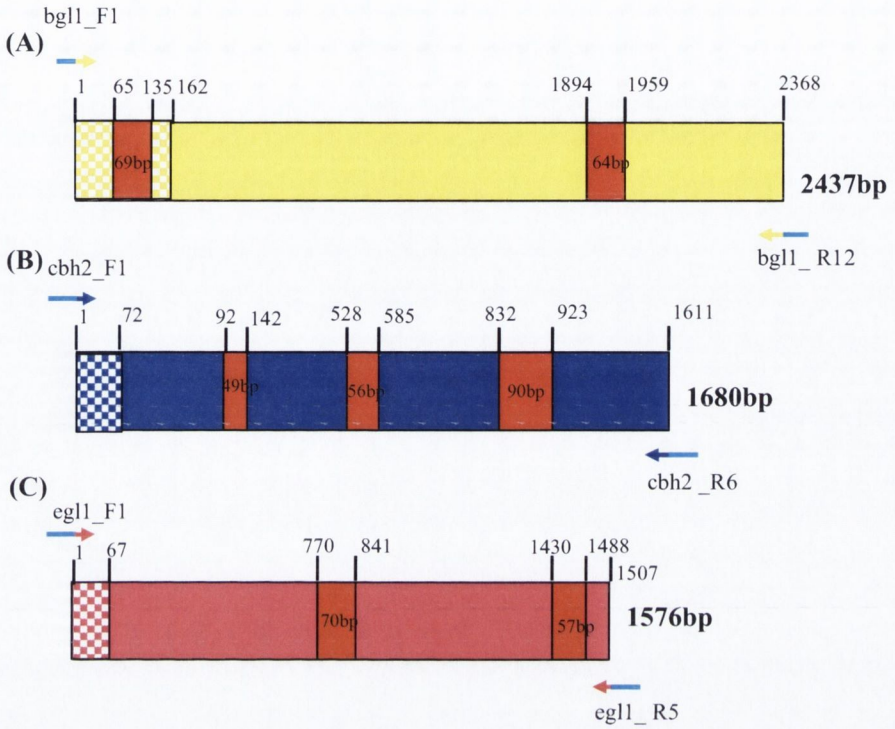


FIG 3.1. Schematic diagram of the three cellulase encoding genes.

The arrows used for PCR are coloured above their respective gene. The chequered region within each gene represents the signal sequence. The red boxes in each represent introns. The respective size and position of each intron is included. The turquoise tag at the end of each primer contains sequence homology to the pGREG plasmid. The expected size of each gene amplified by their respective primers is included (numbers in bold). (**A**): *bgl1* (yellow) gene. (**B**): *cbh2* (blue). (**C**): *egl1* (blue).

3.2.2 DNA extraction from *T. reesei.*

Genomic DNA was extracted from *T. reesei* as described in Chapter 2, section 2.2. The appearance of a characteristic high molecular mass band confirmed presence of intact DNA (Fig 3.2).

3.2.3 PCR reactions to amplify cellulase fragments.

Polymerase Chain Reaction (PCR) was carried out to amplify the genes of interest from the extracted *T. reesei* genomic DNA. The GC content of the three genes was considered when determining the PCR conditions. The GC% content of genes *bgl1*, *cbh2* and *egl1* is 57, 55 and 62% respectively. The GC content is quite high compared to genes in *S. cerevisiae*, which have an average GC content of 38% (Martinez et al., 2008). Consequently amplifying the cellulase encoding genes of interest required an elevated primer annealing temperature. A higher annealing temperature increases the stringency of the PCR and therefore decreases the likelihood of mis-priming. Temperature gradient PCRs were conducted to establish the maximum common annealing temperature that could be used to amplify all three genes (data not shown). The optimum temperature for annealing was determined to be 67°C (See Chapter 2, Section 2.6). Each gene was amplified using 200ng of *T. reesei* genomic DNA as template with gene-specific primers. A full list of oligonucleotide primers is provided in Tables 2.1 and 2.2. Since insertion into the pGREG vector occurred via homologous recombination, it was necessary to include sequences complementary to the site of recombination within the primers used for gene amplification. The gene-specific oligonucleotide primers included at least 33bp of sequence homologous to the rec (recombination) sites on the pGREG vector (Chapter 2, Fig 2.1). The gene *bgl1* was amplified with primers bgl1_F1 and bgl1_R12 (Fig.3.1.A). This yielded a band of approximately 2.5kb which corresponded to the expected size of 2437bp (Fig. 3.3.A, lane 2). The *cbh2* gene was amplified up with oligonucleotide primers cbh2_F1 and cbh2_R6. A PCR product of approximately 1.7kb was obtained using these primers (Fig 3.3.A, lane 3). The expected size band was 1680bp. The *egl* gene was amplified with oligonucleotide primers egl_F1 and egl_R5. A band of approximately 1.5kb corresponding to the expected size of 1576bp was generated (Fig. 3.3.A, lane 4). A small amount of low molecular weight products was evident in all lanes and most likely represented unincorporated oligonucleotides or primer dimers.

3.2.4 Transformation of *S. cerevisiae* S150 and screening of genomic DNA cellulase clones

The PCR products for all three genes were extracted from the gel and purified as described (Chapter 2, Section 2.7). The PCR products were mixed with linearised pGAL1 in a molar ratio of 10:1 and the mixture was introduced into *S. cerevisiae* strain S150 as described in Chapter 2, Section 2.8. Transformants were selected on SC ura$^-$ plates. To ensure the plasmids contained the inserted PCR fragment, the colonies were picked onto SC his$^-$ plates.

The resultant strains were named S150GAL1egl1gDNA, S150GAL1cbh2gDNA and S150GAL1bgl1gDNA where g denotes genomic DNA (Chapter 2, Table 2.4). To confirm the presence of gene-specific inserts, PCR amplifications were carried out on colony lysates using the same primers used to generate the original insert (See Section 3.2.3). In all cases a band of the expected size was obtained from cells transformed with the *bgl1*, *cbh2*, *egl1* genes respectively (Fig 3.3B lanes 2, 4 and 6). The presence of the pGREG plasmid was confirmed by PCR amplification using the primers Kan_F and Kan_R (See Chapter 2, Table 2.3) which amplified the 1.1kb KanMX antibiotic resistant cassette (Chapter 2, Fig 2.1). The presence of the 1.1kb band in lanes 3, 5 and 7 confirmed that plasmid was present in each colony (Fig 3.3B). Thus it would appear that the cellulase genes were successfully transformed into *S. cerevisiae* S150 strains. The DNA sequence of each insert was confirmed by sequencing plasmid DNA as described in Chapter 2, Section 2.12. A number of nucleotide mutations were detected in some inserts (See Table 3.1).

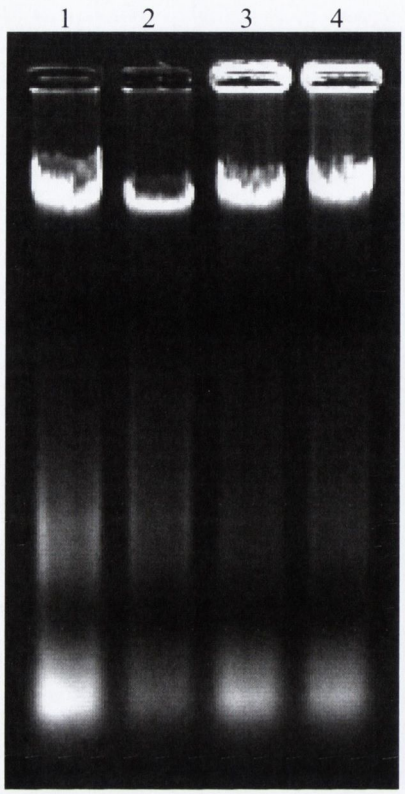
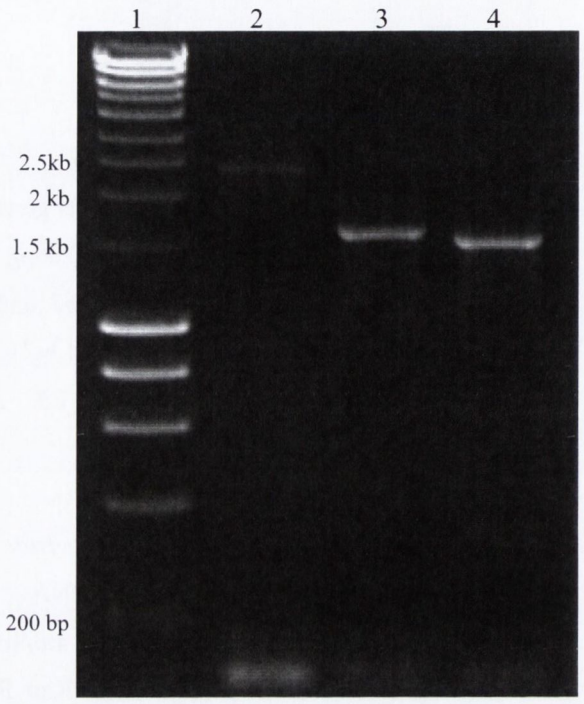


FIG. 3.2. Gel electrophoresis of genomic *T. reesei* DNA. Lanes 1-4 contain 5µl of genomic DNA from four separate extractions.

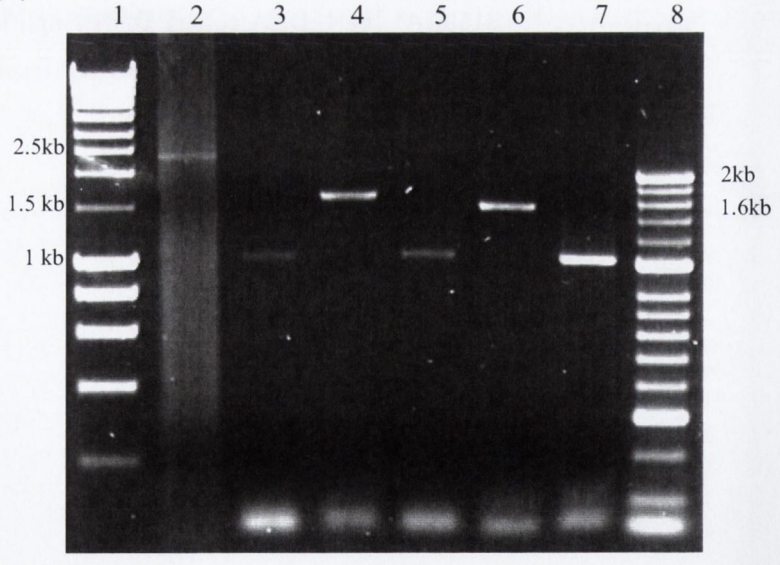
FIG.3.3. PCR amplification of *T. reesei* genomic DNA cellulase inserts.
(A). PCR products amplified from 200ng of *T. reesei* genomic DNA using primers specific for *bgl1*, *cbh2* and *egl1* respectively. Lane 1: Hyperladder I. Lane 2: bgl1_F1 and bgl1_R12. Lane 3: cbh2_F1 and cbh2_R6. Lane 4: eg1_F1 and eg1_R5. Appropriate size markers are indicated

(B). Direct colony PCR of *S. cerevisiae* strain S150 yeast colonies transformed with genomic DNA inserts. Lanes 2-3: S150GAL1bgl1gDNA colony DNA amplified using primers (bgl1_F1 and bgl1_R12) and (Kan_F and Kan_R) respectively. Lanes 4-5: S150GAL1cbh2gDNA colony DNA amplified using primers (cbh2_F1 and cbh2_R6) and (Kan_F and Kan_R) respectively. Lanes 6-7: S150GAL1egl1gDNA colony DNA amplified with primers (eg1_F1 and eg1_R5) and (Kan_F and Kan_R) respectively. Lanes 1 and 8: Hyperladders I and II respectively.

(A)



(B)



3.2.5 Replacing the GAL1 promoter of pGREG plasmid with a PGK1 promoter:

The cellulase encoding genes used in this study were cloned into the pGREG vector under the control of the highly regulated galactokinase, GAL1 promoter (Chapter 2, Section 2.9, Fig 2.1). Thus the genes could be induced in the presence of galactose and repressed in glucose. In order to utilize and ferment cellulose, it may be necessary to have continual expression of the transformed cellulase encoding genes within the yeast. Hence the GAL1 was replaced by the constitutively expressed phosphoglycerate kinase, PGK1 promoter of *S. cerevisiae* (Chapter 2, Section 2.13). The use of the pGREG vectors allowed promoter swapping to be done quickly and efficiently.

The GAL1 promoter from an empty pGREG vector (without gene insert) was removed from the vector by digestion of the restriction enzymes Sac1 and Spe1, which flank the promoter (Fig. 3.4.A). The double digestion released a 503bp fragment (Fig 3.4.B, lane 2). The linearised plasmid (approximately 7.8 kDa) was gel extracted and purified as described in Chapter 2, Section 2.7.

The PGK1 promoter was amplified from *S. cerevisiae* genomic DNA using oligonucleotide primers PGK1_F and PGK1_R (Chapter 2, Table 2.3) that had overhangs (35bps) with sequence homology to the regions flanking the promoter on the pGREG vector. The expected size of the PGK1 promoter with pGREG vector overhangs is 788bp and a band of approximately this size was obtained from the PCR amplification (Fig 3.4.C, lane 1).

Approximately 100ng of linearised purified vector was mixed with 1 µg of purified pGREGPGK1 PCR product and transformed into *S. cerevisiae* S150. Plasmid was recovered from transformants and was designated as pPGK1. Nucleotide sequencing of plasmid DNA with oligonucleotides PGK1_F and PGK1_R confirmed that the GAL1 promoter had been successfully replaced with the PGK1 promoter. The cloned copies of *egl1*, *cbh2* and *bgl1* were amplified and inserted into pPGK1vector via *in vivo* homologous recombination as described above (data not shown). The GAL1 promoter on the original cellulase-containing plasmids could not be swapped directly due to occurrence of a SacI restriction site within the coding region of each of the cellulase inserts.

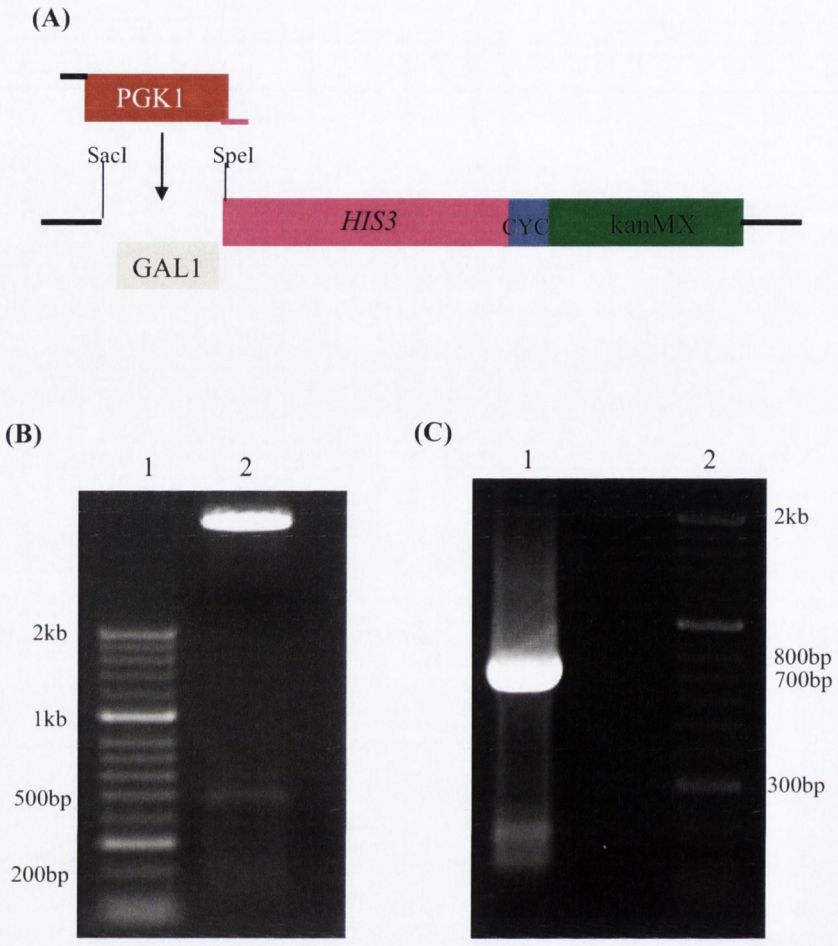


FIG. 3.4. Swapping of GAL1 promoter for PGK1 promoter

(A) Schematic diagram of promoter exchange within the empty pGREG vector. The restriction enzymes used to remove the GAL1 (silver) promoter are shown. The PGK1 promoter is shown in red. The regions of sequence homology to either side of the GAL1 promoter are shown in black and pink. The location for the *HIS3* gene is shown in pink, the CYC terminator and kanMX are displayed in blue and green respectively.

(B) Lane 1: Hyperladder II. Lane 2; pGAL1 plasmid (empty vector) cut with SacI and SpeI restriction enzymes.

(C) PCR amplification of S150 DNA by primers PGK1_F and PGK1_R. Lane 1: amplified PGK1 promoter. Lane 2; Hyperladder II.

3.2.6. Mutation of *egl1* introns

The cellulase encoding genes *egl1*, *cbh2* and *bgl1* each contained introns as described in Section 3.2.1. Analysis of intronic sequences of fungi indicated that consensus sequences involved in their removal, the 5' donor site, GT, and the 3' acceptor site, AG, are conserved (Alberts et al., 1994). These sequences are found in all introns of both *S. cerevisiae* and *T. reesei*. *S. cerevisiae* intronic sequences contain a characteristic TACTAAC branch site located upstream of the acceptor site (Thomas et al., 1999). The TACTAAC consensus sequence is highly conserved in *S. cerevisiae* introns but not present in *T. reesei* introns.

To improve the possibility that the intronic sequences in the *T .reesei* cellulase gene would be spliced in *S. cerevisiae*, the introns of *T. reesei egl1* were altered to insert a TACTAAC motif upstream of the AG acceptor site. Intron-specific primers were designed that contained the TACTAAC sequence. PCR amplification was performed to insert the new sequence into each intron of the *egl1* gene using primers egl1_F3, egl1_R3 and egl1_ R4 (Fig 3.5). The *egl1* gene was then amplified as two overlapping fragments called egl1-intron 1 and egl1-intron 2. Fragment eg1-intron 1 was amplified with primers egl1_F1 and egl1_R3. Fragment eg1-intron 2 was generated with primers egl1_F3 and egl1_ R4. This yielded two fragments of 900bp and 750bp (data not shown). Both fragments were then mixed together and co-transformed into S150 along with the linearised pPGK1. The resultant strain was designed S150PGK1 egl1^{TACTAAC} (Chapter 2, Table 2.4). The recovered plasmid pPGK1egl1^{TACTAAC} was sequenced as described in Chapter 2, Section 2.12.

FIG 3.5. Schematic diagram of introducing mutational insertions into the introns of *T. reesei* gene *egl1*.

(A) The entire 70bp sequence of *egl1* intron 1(Red box).

(B) The 57bp of *egl1* intron 2.

Primers indicated by colour coordinated arrows. Red; intron sequence. Pink: egl1 sequence. Turquoise; sequence homologus to pGREG vector.

Primer names and TACTAAC sequence to be inserted are shown.

3.2.7 Removal of *bgl1* introns via overlapping fragments.

The removal of introns is necessary for gene expression to occur. Should the *S. cerevisiae* splicing mechanism fail to remove the introns from the *T. reesei* inserts *in vivo*, an alternative scheme was devised to remove the introns *in vitro* via PCR amplification. This strategy was performed with the *bgl1* gene. The plan involved amplifying the *bgl1* gene as four independent fragments (B1, B2, B3 and B4) that excluded the introns (Figure 3.6.A). The subsequent *bgl1* fragments were assembled together and transformed into *S. cerevisiae*.

T. reesei genomic DNA was used as template to generate each *bgl1* fragment. Fragments B1, B2, B3 and B4 were generated using primer pairs (bgl1_F1 and bgl1_R1), (bgl1_F3 and bgl1_R5), (bgl1_F5 and bgl1_R8) and (bgl1_F9 and bgl1_R12) respectively (Fig 3.6.A). The expected size of each fragment was 98, 893, 907 and 443bp respectively. As shown in Fig 3.7A, (lanes 2 and 4) bands of the expected sizes for fragments B1 and B4 were obtained. Likewise, bands of the expected sizes were obtained for fragments B2, B3 (Fig 3.7B, lanes 2 and 4, respectively)

Sequence homology between the fragments was required in order to reassemble the gene. Each fragment was reamplified using primers containing overhangs that possessed sequence homology to the other *bgl1* fragments (Figure 3.6.B). Fragment B1 was reamplified with primers bgl1_F1 and bgl1_R2 yielding a band approximating to the expected 119bp (Fig 3.7.A, lane 3). The latter primer had an overhang comprising of 21bp of homology to B2. Fragment B2 was reamplified with primers bgl1_F2 and bgl1_R5 to yield the expected 916bp band (Fig 3.7.B, lane 3). Oligonucleotide primer bgl1_F2 had a 23bp overhang with sequence homologous to fragment B1. Overall there was 44bp of sequence homology between fragments B1 and B2. Fragment B3 was reamplified using primers bgl1_F5 and bgl1_R9 to generate a band of the approximate size 932bp (Fig. 3.7.B lane 5). The latter primer had 25 bp overhang of sequence homologous to B4.

FIG. 3.6. Schematic diagram of PCR based strategy to remove introns from *bgl1*

(A) Four bgl1 fragments (B1-B4) were generated using the primers indicated (coloured arrows). The introns are shown as red boxes. The predicted size of fragments is shown below the schema.

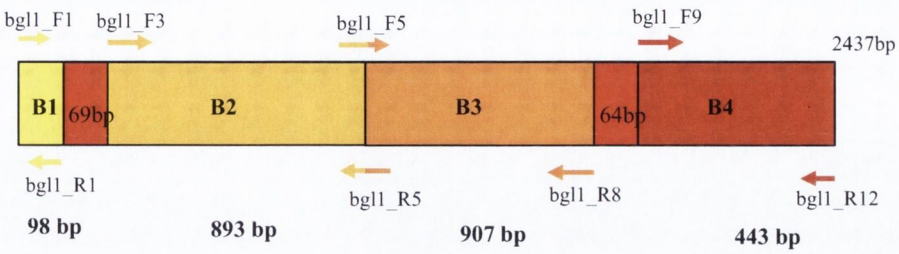
(B) Fragments were reamplified to generate overlapping sequence homology. The primers used to create overlap between fragments B1-B4 are indicated.

(C) Connecting fragments B1 with B2 and B3 with B4 using outlying primers. Vertical red lines delineate sequence homology

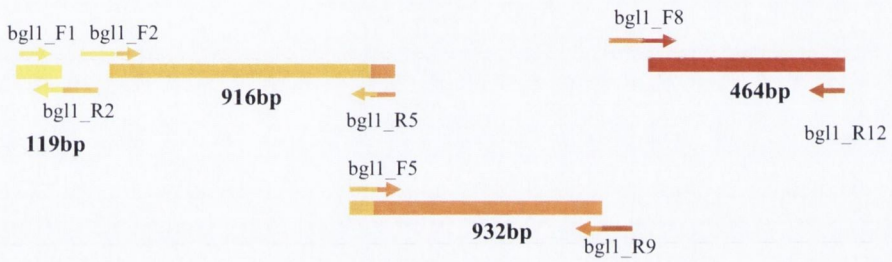
(D) Connecting fragment B1+2 with B3+4.

(E) Full length recombined intron-less *bgl1*

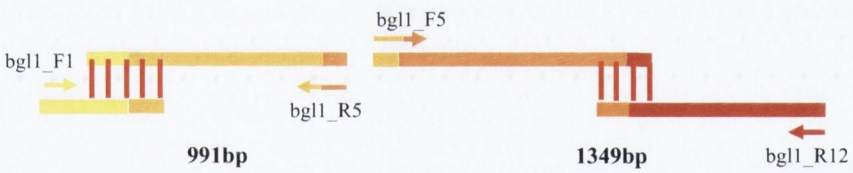
(A) PCR amplify four independent fragments



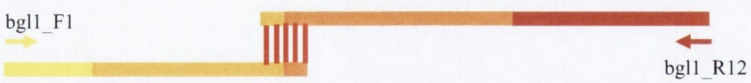
(B) Reamplify to generate overlapping sequence homology between fragments



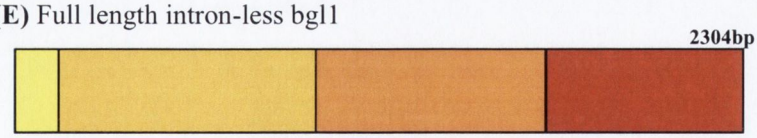
(C) Connecting fragments B1 with B2 and B3 with B4 outlying primers



(D) Connecting fragments B1+2 with B3+4



(E) Full length intron-less bgl1



Fragment B4 was reamplified using primers bgl1_F8 and bgl1_R12 to generate a band with a predicted size of 464bp. As shown in Fig 3.7.A lane 5, a band of ~500bp was obtained. Fragment B4 has a 46bp of homology with B3. In order to reassemble B1 with B2, 100ng of each was mixed together and amplified with the outer primers of the two fragments, i.e. primers bgl1_F1 and bgl1_R5 (Figure 3.6.C). The generated fragment was designated as B1+2 and yielded a band of approximately 1kb which corresponded to the expected size of 991bp band (Fig. 3.7.C, lane 2).

The reassembly of bgl1 fragments B3 and B4 was carried out by PCR using primers bgl1_F5 and bgl1_R12. This generated a band of approximately 1.4kb that corresponds to the expected size of 1349bp (Fig. 3.7.C, lane3). This was designated as bgl1 fragment B3+4. The bands of the correct size were gel extracted and purified.

A final round of PCR was performed to join B1+2 with B3+4 and thus yield an intron-less recombined *bgl1* fragment (Fig3.6.D). There is 39bp of sequence homology overlap between the 3' end of fragment B1+2 and the 5' end of fragment B3+4. The expected size of the full length intron-less *bgl1* amplified using primers bgl1_F1 and bgl1_R12 was 2304bp. A band corresponding to this size was observed in Fig 3.7.D, lane 2. Amplifying *T. reesei* genomic DNA using the same primers generated a band of approximately 2.5kb which corresponds to the expected size of 2436bp (Fig 3.7.D lane 3). The size difference between the recombined fragments B1-B4 and the genomic DNA reflected the removal of the two introns comprising 133nt in total. The resultant 2.3kb of lane 2 was gel extracted and purified then transformed into *S. cerevisiae*. The strain was named S150GAL1bgl1gDNA ^{recombined}.

FIG.3.7. Generation of intron-less *bgl1* from *T. reesei* genomic DNA

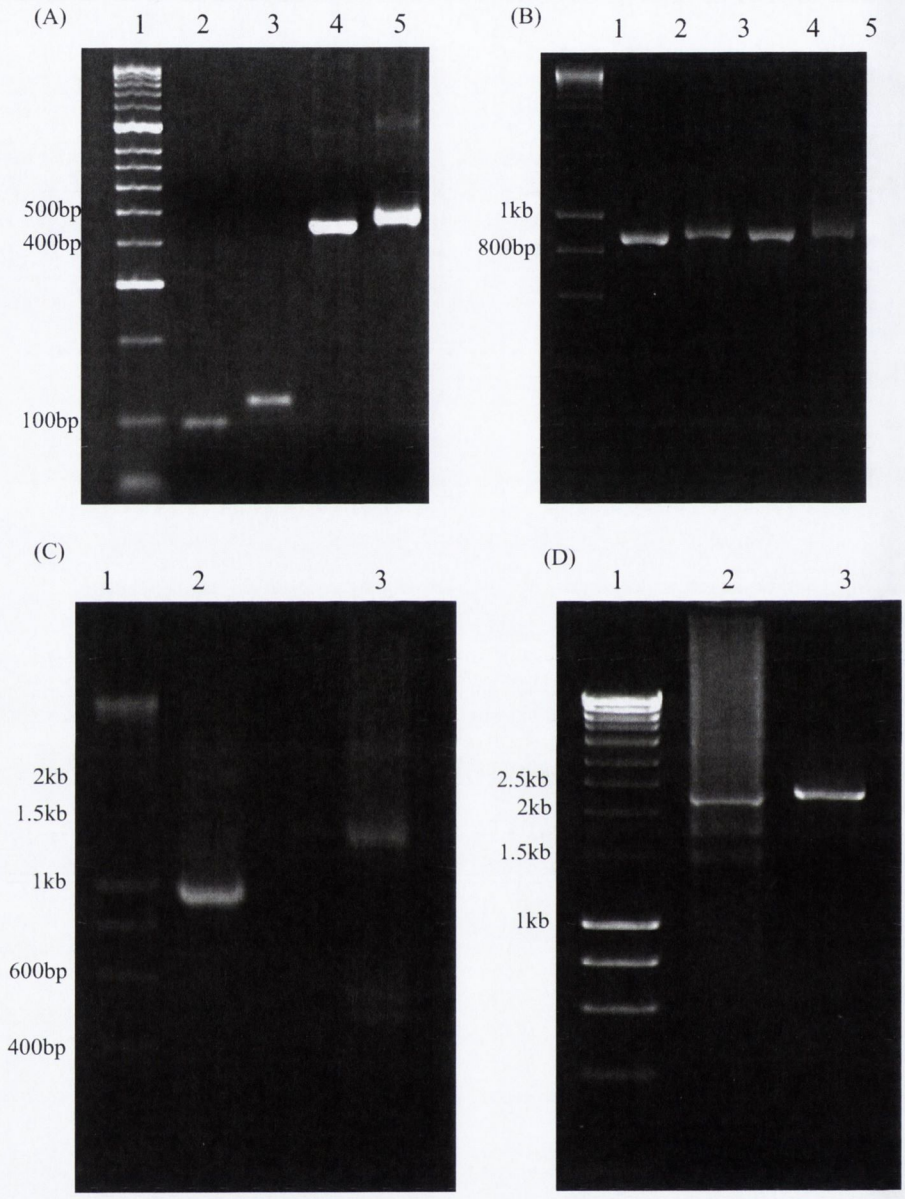
(A) PCR of *bgl1* fragments B1 (lanes 2 and 3) and B4 (lanes 4 and 5). Lane 1: Hyperladder II, the sizes of the 100, 400 and 500bp bands are shown. Lane 2: *T. reesei* DNA amplified by primers bgl1_F1 and bgl1_R1. Lane 3: reamplified fragment B1 (lane 2) using primers bgl1_F1 and bgl1_R2. Lane 4: *T. reesei* DNA amplified by primers bgl1_F9 and bgl1_12. Lane 5 reamplified fragment B4 (lane 4) using primers bgl1_F8 and bgl1_R12.

(B) PCR products of *bgl1* of Fragments B2 (lanes 2 and 3) and B3 (lanes 4 and 5). Lane 1: Hyperladder II with markers 800 bp and 1 kb displayed. Lane 2: *T. reesei* DNA amplified with primers bgl1_F3 and bgl1_R5. Lane 3: Reamplified fragment B2 (lane 2) using primers bgl1_F2 and bgl1_R5. Lane 4: *T. reesei* DNA amplified with primers bgl1F5 and bgl1_R8. Lane 4: Reamplified fragment B3 using primers bgl1_F5 and bgl1_R9.

(C) Allele PCR connecting *bgl1* Fragments B1 to B2 and B3 to B4. Lane 1: Hyperladder I, size markers 1kb and 1.5kb were shown. Lane 2: Fragments B1 and B2 amplified using primers bgl1_F1 and bgl1_R5. Lane 3: Fragments B3 and B4 amplified by primers bgl1_F5 and bgl1_R12.

(D) PCR assembly of *bgl1* fragments B1+2 with B3+4.

Lane 1: Hyperladder I with size markers 1.5, 2 and 2.5kb shown. Lane 2: 100 ng of each fragment amplified with primers bgl1_F1 and bgl1_R12. Lane 2: 200 ng of genomic DNA amplified using same primers.


(A)
1
2
3
4
5
500bp
400bp
100bp
(B)
1
2
3
4
5
1kb
800bp
(C)
1
2
3
2kb
1.5kb
1kb
600bp
400bp
(D)
1
2
3
2.5kb
2kb
1.5kb
1kb

3.2.8 cDNA synthesis of cellulase genes from *T. reesei.*

An alternative approach to obtaining intron-less genes was also pursued. This involved cDNA synthesis from *T. reesei* RNA. The fungus was grown in cellulase inducing conditions and RNA was extracted (See Chapter 2, Section 2.3) and treated with DNase I to remove contaminating DNA (See Chapter 2, Section 2.4). As previously reported, generating a full length cellulase cDNA clone can prove difficult (Cummings and Fowler, 1996). Due to the limitations in the processivity of reverse transcriptase enzyme, fragments greater than 1 kb are difficult to synthesise and may require specialised conditions. Problems can also arise due to the amount of template RNA. If a gene produces low amounts of mRNA there will be low levels of template to reverse transcribe into cDNA and hence synthesis will be difficult. To overcome these difficulties, it was decided to generate two cDNA fragments with overlapping homology and to connect the fragments using Allele PCR amplification (Chapter 2, Section 2.6).

3.2.8.1 Generating *egl1* cDNA

For *egl1*, fragment A was generated from cDNA using reverse primer eg1_R2 (Fig. 3.8A). For PCR amplification primers eg1_F1 and eg1_R2 were used (See Chapter 2, Table 2.2). This generated a band of approximately 950bp which corresponds to the expected size (Fig. 3.8.B, lane1). The Verizyme polymerase enzyme was used for PCR amplification. To ensure that the PCR product resulted from cDNA synthesis, a negative control was carried out in which the reverse transcriptase enzyme was omitted from the reaction (Fig 3.8.B, lane 2). As a positive control, genomic DNA was amplified with the same primers (Fig. 3.8.B, lane 3). This was expected to produce a band of 1020bp, due to the presence of an intron. This expected size difference between the cDNA and the genomic DNA was observed (Fig. 3.8.B, lanes 1 and 3). Thus, the cDNA reaction was successful and a partial *egl1* cDNA fragment was generated. The second *egl1* cDNA fragment, B, was synthesized using oligonucleotide primer eg1_R5 (Fig. 3.8.A). The 3' end of the mRNA was amplified using this same reverse primer and a forward primer eg1_F2. The expected size of fragment B was 741bp. A band of approximately 750 bp was evident in (Fig. 3.8.B, lane 4). The omission of RT in the reverse transcriptase reaction resulted in no products indicating that there was no genomic DNA contamination (Fig. 3.8.B, lane 5). A PCR amplification using genomic DNA and primers egl1_F2 and egl1_R5 produced a product of approximately 900bp (Fig. 3.8.B, lane 6). The expected

size was 868bp. The size difference reflects the lack of introns in the amplified cDNA.

Fragment A and B had an overlapping homology of 245 bp. To connect the two fragments together and yield full-length intron-less *egl1,* an Allele PCR amplification using 100ng each of purified fragment A and B as template was performed with primers eg1_F1 and eg1_ R5 (See Chapter 2, Section 2.6). The expected size of a full-length *eg1* cDNA clone was 1446bp (Fig. 3.8.C, lane 2). PCR amplification with the same primers and using genomic DNA as template produced a fragment of approx 1580 bp (Fig. 3.8.C, lane 3). The difference in size between the cDNA and genomic DNA PCR products reflects the absence of introns in the cDNA. The recombinant *egl1* fragment was then purified and transformed into yeast along with the linearised pGAL1 vector (Chapter 2, Sections 2.7 and 2.8). Plasmid DNA was recovered and sequenced by GATC (Konstadt, Germany) (See Chapter 2, Section 2.12). Sequence analysis indicated that the complete cDNA had been obtained and lacked the intronic sequences however, a silent mutation at base 69 and a missense mutation followed by a base insertion were found at bases 1360 and 1361. The resultant strain was named S150GAL1egl1cDNA $^{T1360\mathbf{A}::\mathbf{G}1361}$ (Table 3.2), where the superscript indicates the position of missense mutation. The letters in italics denoted the wild type nucleotide; the position of the error within the cDNA sequence is indicated by the number and the letter in bold represents the mutant nucleotide. The symbol :: indicated an insertion mutation had occurred at the numbered position. The cloning of this *egl1* cDNA insert into yeast along with linearised pPGK1 vector generated strain S150PGK1egl1cDNA $^{T1360\mathbf{A}::\mathbf{G}1361}$.

3.2.8.2. Generation of *cbh2* cDNA

Two overlapping *cbh2* cDNA fragments, 1 and 2, were synthesized using reverse primers cbh2_R2 and cbh2_R6 (See Chapter 2, Table 2.2) respectively (Fig. 3.8.D). The cbh2_R2 synthesised cDNA was amplified with primers cbh2_F1 and cbh2_R2 to give fragment 1 (Fig. 3.8.D). The expected size was 747bp and a band approximately 750bp was evident (Fig. 3.8.E, lane 1). Using the same oligonucleotide primers to amplify *T. reesei* genomic DNA (Fig 3.8.E, lane 2), a fragment of 850bp was expected due to the presence of a 105bp of intronic sequences. This expected difference in sizes between the cDNA and genomic DNA was observed (Fig. 3.8.E lanes 1 and 2). The cDNA for *cbh2* fragment 2 was synthesized using reverse primer cbh2_R6. The

cDNA was then amplified with the same reverse primer and cbh2_F2 (Fig. 3.8.D). A fragment of approximately 830 bp was generated (Fig 3.8.E, lane 3. Amplifying genomic DNA using the same primers generated a band of approximately 1000bp, which corresponds to the expected size of 920bp (Fig. 3.8E, lane 4). The *cbh2* fragments 1 and 2, which contained 92bp of overlapping homology, were purified and 100ng of each were mixed and connected together via Allele PCR amplification with primers cbh2_F1 and cbh2_R6 (Fig. 3.8.D). The expected size of intron-less *cbh2* was 1482bp (Fig. 3.8.F, lane 2). A PCR amplification of genomic DNA with the same primers generated a band of approximately 1.7kb which corresponds to the expected size of 1677bp (Fig. 3.8.F, lane 3). After purification, recombined *cbh2* cDNA was transformed into *S. cerevisiae* strain S150 along with linearised pGAL1 vector. The plasmid was recovered and sequenced as described (Chapter 2, Section 2.12). Sequence analysis indicated that the complete cDNA had been obtained and lacked the intronic sequences however, a silent mutation at base 561 and two missense mutations at positions 1076 and 1345 were detected (Table 3.2).

The resultant strain was named S150GAL1cbh2cDNA$^{G1076C,\ G1345C}$. The cloning of this *cbh2* cDNA insert into yeast along with linearised pPGK1 vector generated strain S150PGK1cbh2cDNA$^{G1076C,\ G1345C}$.

3.2.8.3 Generation of cbh1 cDNA

A cDNA copy of the *cbh1* gene was also generated. The *cbh1* was synthesized in two fragments X and Y (Fig.3.9.A). cDNAs were generated by reverse transcriptase using reverse primer cbh1_R1 and the resultant cDNA was amplified using cbh1_F1 and cbh1_R1 to yield fragment X. A band of approximately 900bp which corresponds to the expected size of 884bp was evident (Fig 3.9.B, lane 1). A second cDNA synthesis was carried out using the reverse primer cbh1_R2 and the resultant cDNA was amplified using primers cbh1_F2 and cbh1_R2 to generate fragment Y. A band of the expected size 821bp was generated (Fig. 3.9.B, lane 2). Each fragment was gel extracted and purified (Chapter 2, Section 2.7).

Fragment X and Y had 92bp of overlap. A 100 ng quantity of each purified fraction was mixed together and amplified with primers cbh1_F1 and cbh1_R2. This generated a full length *cbh1* fragment of approximately 1.6kb which

corresponded to the expected size of 1613 bp (Fig 3.9.C, lane 1). *T. reesei* of genomic DNA amplified by these primers generated a band of approximately 1750bp corresponding to the expected size of a 1744 bp band, a fragment of this approximate size was observed (Fig 3.9.C, lane 2). The size difference between the cDNA product and the genomic product results from the absence of the introns in the recombined *cbh1*. The plasmid was recovered and sequenced (Chapter 2, Section 2.12). The resultant strain was named S150GAL1cbh1cDNA ^{wt}. The wt indicated that the wild type *T. reesei cbh1* sequence was present with no errors. The cloning of this *cbh1* cDNA insert into yeast along with linearised pPGK1 vector generated strain S150PGK1cbh1cDNA ^{wt}.

3.2.8.4. Generation of bgl1 cDNA.

The generation of *bgl1* was more problematic due to the large size of the open reading frame (2235bp excluding pGREG overhangs). The full length cDNA was generated in three fragments α, β and γ. cDNA synthesis was performed using primers bgl1_R4, bgl1_R6 and bgl1_R12 (Chapter 2, Table 2.1) (Fig.3.10.A). The fragments were PCR amplified with a 10:1 mixture of Taq and Verizyme polymerases. Fragment α was amplified by primers bgl1_F1 and bgl1_R4 to generate a PCR product of 765 bp (Fig. 3.10.B, lane 2). The PCR amplification of genomic DNA using the same primers generated a fragment of approximately 800bp (Fig. 3.10.B, lane 3). This band corresponded with the expected size of 834bp. The size difference between the cDNA and genomic DNA PCR product indicated that the intron was absent in the cDNA (Fig 3.10.A).

The *bgl1* fragment β, was amplified with primers bg1_F4 and bgl1_R6 (Fig.3.10.A). A 700bp band was generated (Fig 3.10.C, lane 1). No band was observed in the RT negative control reaction (Fig 3.10.C, lane 2). This indicated that there was no genomic DNA contamination. The amplification of genomic *T. reesei* DNA with the latter primers produced a fragment of the same size as there are no introns in this section (Fig.3.10.C, lane 3).

The final *bgl1* cDNA fragment, γ, was amplified from cDNAs using primers bgl1_F7 and bgl1_R12 (Fig.3.10.A) to yield a band of approximately 1200bp, corresponding to the expected size of 1164bp (Fig 3.10.D, lane 1). Amplifying genomic DNA with the same primers generated a band of over 1200bp (Fig 3.10.D, lane 2), corresponded to the expected size of 1228bp. The RT negative

control displayed no amplified product (Fig 3.10.D, lane 3). The three PCR fragments generated from the cDNAs were gel extracted and purified (Chapter 2, Section 2.7).

Fragments α and β had 223bp of homologous overlap while fragments β and γ had 105bp of sequence homology. Previously when reconstructing a gene in several pieces, the fragments were connected together *in vitro* via PCR and transformed into yeast. An alternative approach was attempted with *bgl1* cDNA generated fragments. The three overlapping fragments were mixed together along with linearised plasmid pGAL1 in a molar ration of 10:1 inset to vector (Chapter 2, Section 2.8). The mixture was introduced into *S. cerevisiae* strain S150. The endogenous ability of yeast cells to undergo homologous recombination was used to reassemble the *bgl1* gene *in vivo*. Direct colony PCR was performed on transformants selected on ura$^-$ plates to determine if the three fragments had recombined to yield full length *bgl1*. The PCR amplification of DNA isolated from the transformants with primers bgl1_F1 and bgl1_R12 yielded a band of approximately 2.2kb (Fig 3.10.E, lane 1). PCR amplification genomic DNA using the same primers generated a fragment of approximately 2.5kb (Fig 3.10.E, lane 2). Thus the gene had been reassembled correctly by the yeast. The plasmid was recovered and sequenced as described. A number of missense mutations were identified in the cloned cDNA. The resultant strain was designated as S150GAL1bgl1cDNA$^{C64\mathbf{A},\ A65\mathbf{G},\ T1061\mathbf{C},\ T1085\mathbf{C},\ T2099\mathbf{C},\ A2212\mathbf{G}}$. The cDNA insert was also cloned into yeast along with linearised pPGK1 vector to generate strain S150PGK1bgl1cDNA$^{C64\mathbf{A},A65\mathbf{G},T1061\mathbf{C},T1085\mathbf{C},T2099\mathbf{C},A2212\mathbf{G}}$.

FIG.3.8. Generation of cDNA copies of *T. reesei egl1* and *cbh2*.

The arrows above each gene represent primers used for cDNA synthesis and are colour co-ordinated to its corresponding gene. The introns of each gene are highlighted in red.

(A) Schematic diagram of *egl1* cDNA insert amplified as fragments A (solid pink) and B (dashed pink and white). The primers used to amplify fragments A and B from cDNA are shown. The turquoise line on primers egl1_F1 and egl1_R5 indicate sequences homologous to the rec sites in pGREG plasmid. The introns are highlighted in red. The expected sizes of the amplified products are shown below the schema. Note the nucleotide positions of the primers will be different in the cDNA sequence compared to the gDNA sequence because of loss of introns. sequence homology to the pGREG plasmid.

(B) RT-PCR products of *egl1* fragments. Lane 1: Fragment A amplified using primers egl1_F1 and egl1_R2. Lane 3: Genomic DNA amplified using the same primers. Lane 4: Fragment B amplified using primers egl1_F2 & egl1_R5. Lane 6: Genomic DNA amplified by the same primers. Lanes 2 and 5: RT negative control, reverse transcriptase enzyme was omitted. Lane 7: Hyperladder II.

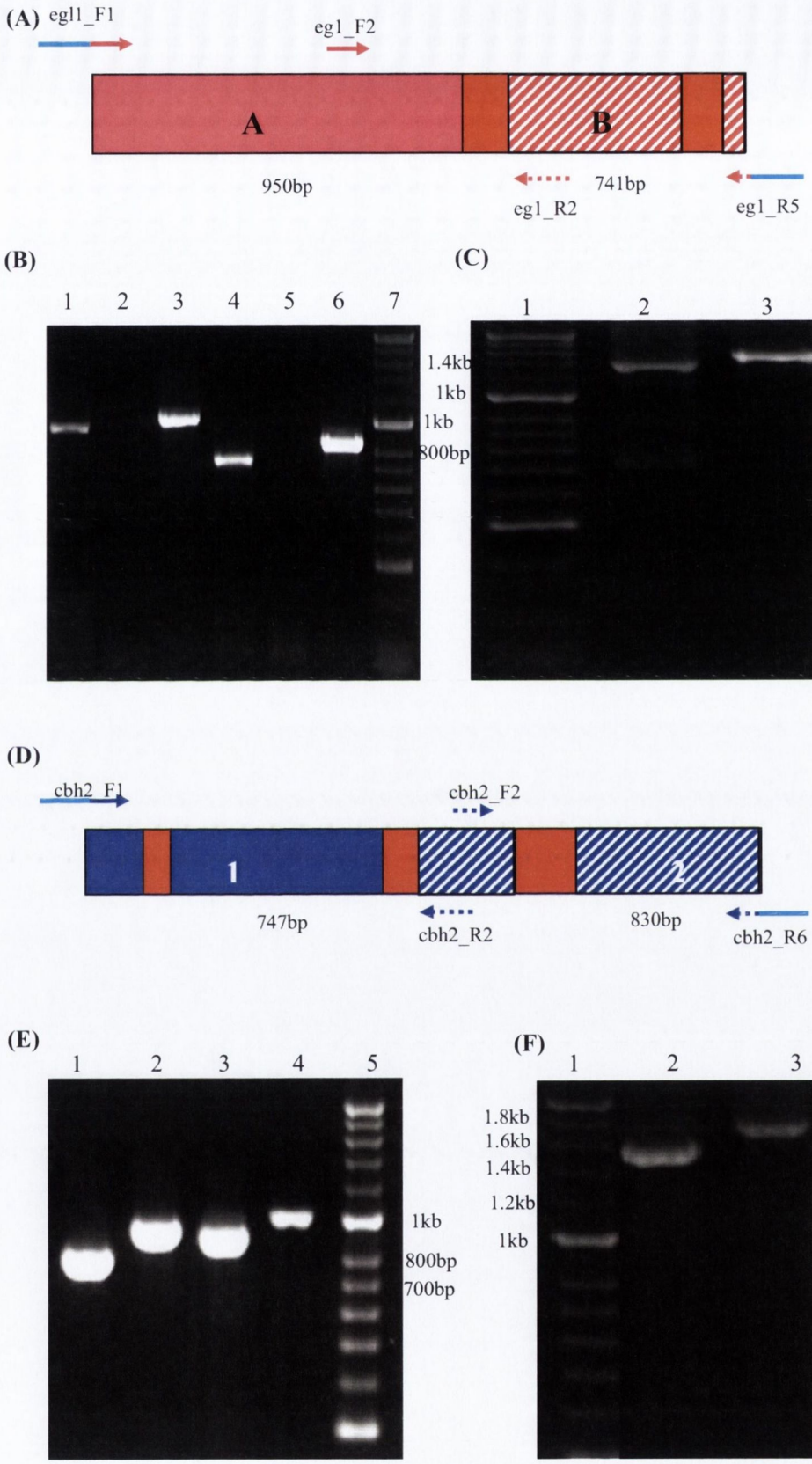
(C) Allele PCR joining *egl1* cDNA fragments A and B. Lane 2: Mixture of Fragments A and B amplified using primers eg1_F1 and egl1_R4. Lane 3; Genomic DNA amplified by the same primers. Lane 1; Hyperladder II.

(D) Schematic diagram of *cbh2* cDNA insert amplified as fragments 1 (solid blue) and 2 (dashed blue and white). The primers used to amplify fragments 1 and 2 from cDNA are shown. The turquoise line on primers cbh2_F1 and egl1_R6 indicate sequences homologous to the rec sites in pGREG plasmid. The introns are highlighted in red. The expected sizes of the amplified products are shown below the schema. Note the nucleotide positions of the primers will be different in the cDNA sequence compared to the gDNA sequence because of loss of introns.

(E) RT- PCR products of *cbh2* fragments. Lane 1: Fragment 1 amplified by primers cbh2_F1 and cbh2_R2. Lane 2: genomic DNA amplified by the same primers. Lane 3; *cbh2* cDNA fragment B amplified by primers cbh2_F2 and cbh2_R6. Lane 4: genomic DNA amplified by the same primers. Lane 5; Hyperladder II.

(F) Allele PCR connecting *cbh2* cDNA fragments 1 and 2.

Lane 1: Mixture of fragments 1 and 2 amplified by primers cbh2_F1 and cbh2_R6. Lane 2: Genomic DNA amplified by the same primers. Lane 3: Hyperladder II.


(A)
egl1_F1
eg1_F2
A
B
950bp
741bp
eg1_R2
eg1_R5
(B)
1 2 3 4 5 6 7
1.4kb
1kb
1kb
800bp
(C)
1 2 3
(D)
cbh2_F1
cbh2_F2
1
2
747bp
cbh2_R2
830bp
cbh2_R6
(E)
1 2 3 4 5
1kb
800bp
700bp
(F)
1 2 3
1.8kb
1.6kb
1.4kb
1.2kb
1kb

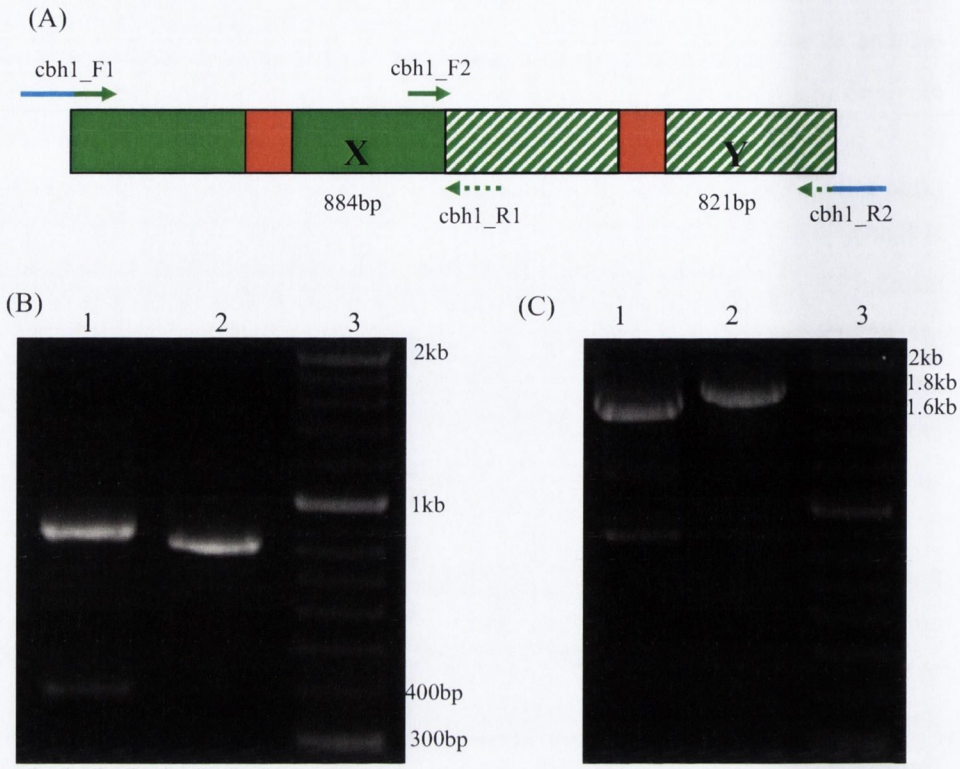


FIG. 3.9. Generation of a cDNA copy of *T. reesei cbh1*.

(A) Schematic diagram of *cbh1* gene. cDNAs of cbh1 were amplified as two fragments, X (solid green) and Y (dashed green and white). The turquoise line on primers cbh1_F1 and cbh2_R2 indicate sequences homologous to the rec sites in pGREG plasmid. The introns are highlighted in red. The expected sizes of the amplified cDNA products are shown below the schema. Note the nucleotide positions of the primers will be different in the cDNA sequence compared to the gDNA sequence because of loss of introns.

(B) RT-PCR products of *cbh1* fragments. Lane 1: Fragment X amplified by primers cbh1_F1 and cbh1_R1. Lane 2: Fragment Y amplified by primers cbh1_F2 and cbh1_R2. Lane 3 is Hyperladder II.

(C) Allele PCR cycle to connect *cbh1* fragments X and Y. Lane 1: Mixture of fragments X and Y amplified by primers cbh1_F1 and cbh1_R2. Lane 2: genomic DNA amplified by the same primers. Lane 3: Hyperladder II.

FIG. 3.10. Generation of a cDNA copy of *T. reesei bgl1*.

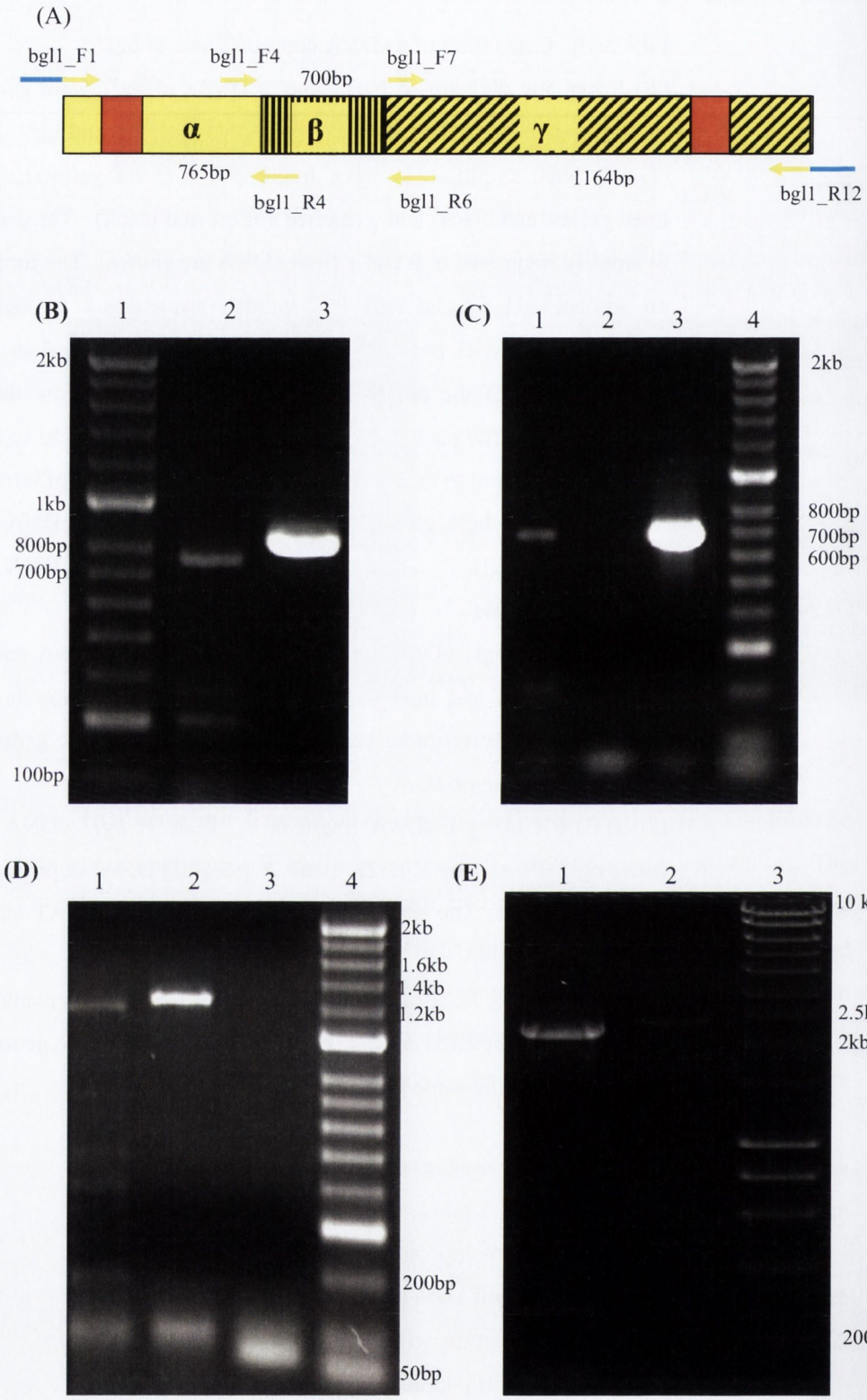
(A) Schematic diagram of *bgl1* gene. cDNAs of bgl1 were generated by reverse transcriptase using primers bgl1_R4, bgl1_R6 and bgl1_R12. The cDNAs were amplified as three fragments: α (solid yellow), β (vertical lined yellow and black) and γ (dashed yellow and black). The primers used to amplify fragments α ,β and γ from cDNA are shown. The turquoise line on primers bgl1_F1 and bgl1_R12 indicate sequences homologous to the rec sites in pGREG plasmid. The introns are highlighted in red. The expected sizes of the amplified products are shown below the schema. Note the nucleotide positions of the primers will be different in the cDNA sequence compared to the gDNA sequence because of loss of introns.

(B) RT-PCR of bgl1 cDNA fragment α. Lane 1: *bgl*1 cDNA amplified using primers bgl1_F1 and bgl1_R4. Lane 2: genomic DNA amplified using same primers.

(C)RT-PCR of bgl1 cDNA fragment β. Lane 1: *bgl1* cDNA amplified by primers bgl1_F4 and bgl1_R6. Lane 2: the same reaction as in lane 1 except reverse-transcriptase enzyme was omitted. Lane 3: genomic DNA amplified by same primers.

(D) RT-PCR of bgl1 cDNA fragment γ. Lane 1: *bgl1* cDNA amplified using bgl1_F7 and bgl1_R12. Lane 2: genomic DNA amplified by same primers. Lane 3: The same reaction as in lane 1 except RT enzyme was omitted.

(E) Direct colony PCR of S150 transformed with *bgl1* fragments. Lane 1: colony DNA amplified by bgl_F1 & bgl1_R12. Lane 2: genomic DNA amplified using the same primers.


(A)
bgl1_F1
bgl1_F4
bgl1_F7
700bp
α
β
γ
765bp
1164bp
bgl1_R4
bgl1_R6
bgl1_R12
(B)
1
2
3
2kb
1kb
800bp
700bp
100bp
(C)
1
2
3
4
2kb
800bp
700bp
600bp
(D)
1
2
3
4
2kb
1.6kb
1.4kb
1.2kb
200bp
50bp
(E)
1
2
3
10 kb
2.5kb
2kb
200bp

3.2.9 Site-directed mutagenesis PCR

DNA sequence analysis of the cDNA copies of *egl1*, *cbh2*, *cbh1* and *bgl1*, identified a number of nucleotide differences from wildtype *T. reesei* sequence. The number and location of each nucleotide change is summarised in Tables 3.1, 3.2 and 3.3. The silent errors were not corrected because they did not alter the amino acid sequence. The missense errors that changed the amino acid coding sequence within the cDNA inserts were corrected by site-directed mutagenesis PCR. This approach involved designing gene specific primers containing the corrected nucleotide (Fig. 3.11). The proof-reading polymerase Verizyme was used for each PCR. Each gene was amplified as overlapping fragments using the corrected gene specific primers (Chapter 2, Tables 2.1 and 2.2) and then re-transformed via the pGREG plasmid into *S. cerevisiae* as described in Chapter 2, Section 2.8. The resultant plasmid was then re-sequenced. If the errors persisted or if new errors were detected, the mutagenesis was repeated until a sequence that had the same amino acid coding region as the wild type was attained. This corrective action was performed only on the cDNA generated inserts.

3.2.9.1. egl1cDNA

The *egl1* insert within plasmid pPGK1egl1cDNA$^{T1360A::G1361}$ had three errors located at nucleotide 69, 1360 and 1361 (Table 3.2). The error at position 69 was a silent mutant of G to A and did not change the amino acid codon. At nucleotide 1360 a missense error of T to A occurred and was followed by an insertion of a G at 1361 (Fig. 3.11A). This caused a frame shift and knocked out the stop codon at nucleotide 1378. Oligonucleotide primer egl1_R6 (Chapter 2, Table 2.2) was designed to correct the errors at 1360 and 1361. This oligonucleotide also had a C-terminal 6xHIS tag attached. A 6xHIS tag polypeptide allows for quick purification of the fusion protein via a nickel column. The HIS tag has advantages over other tags such as GST or FLAG because it does not alter protein structure, function or secretion (Borsig et al., 1997). The primer, egl_R6, was designed to fuse a 6xHIS tag directly to the last coding amino acid of *egl1* (Fig. 3.12.A). A PCR was performed with oligonucleotide primers egl1_F1 and egl1_R6 (Table 2.2) using the pPGK1egl1cDNA $^{T1360A::G1361}$ plasmid as template (Fig 3.11.A). The expected

size band was detected and gel extracted (data not shown). The resultant PCR fragment, along with linearised pPGK plasmid, were transformed into *S. cerevisiae* as described in Chapter 2, Section 2.8. Plasmid DNA was recovered from the transformants. Nucleotide sequencing confirmed that the HIS tag was fused in-frame and that all missense errors had been corrected. The strain was called S150PGK1egl1HIScDNA ^{wt}.

FIG. 3.11. Schematic diagram of strategy to correct mutations within cDNA derived clones.

Missense errors are indicated by red. Silent mutations are indicated by green. The nucleotide positions of errors are based on the cDNA sequence of the respective gene. The brown line at the end of primers egl1_R6 and cbh2_R7 represent 6xHIS tag. The oligonucleotide primer locations described in Chapter 2 Tables 2.1 and 2.2 are based on the genomic DNA sequence, therefore the numbering difference is due to the absence of the introns in the cDNA inserts.

(A) egl1cDNA $^{T1360A::G1361}$. The plasmid pPGK11eg1cDNA $^{cDNA\ T1360A::G1361}$ was amplified with primers egl1_F1 and egl1_R6 to generate egl1HIScDNA^{wt} containing no missense errors.

(B) cbh2cDNA $^{G1076C,\ G1345C}$. The plasmid pPGK1cbh2cDNA $^{G1076C,\ G1345C}$ was amplified in three fragments with primers cbh2_F1 and cbh2_R4, cbh2_F4 and cbh2_R5 and cbh2_F5 and cbh2_R6. Fragments were connected to yield insert cbh2cDNA ^{A896C}.

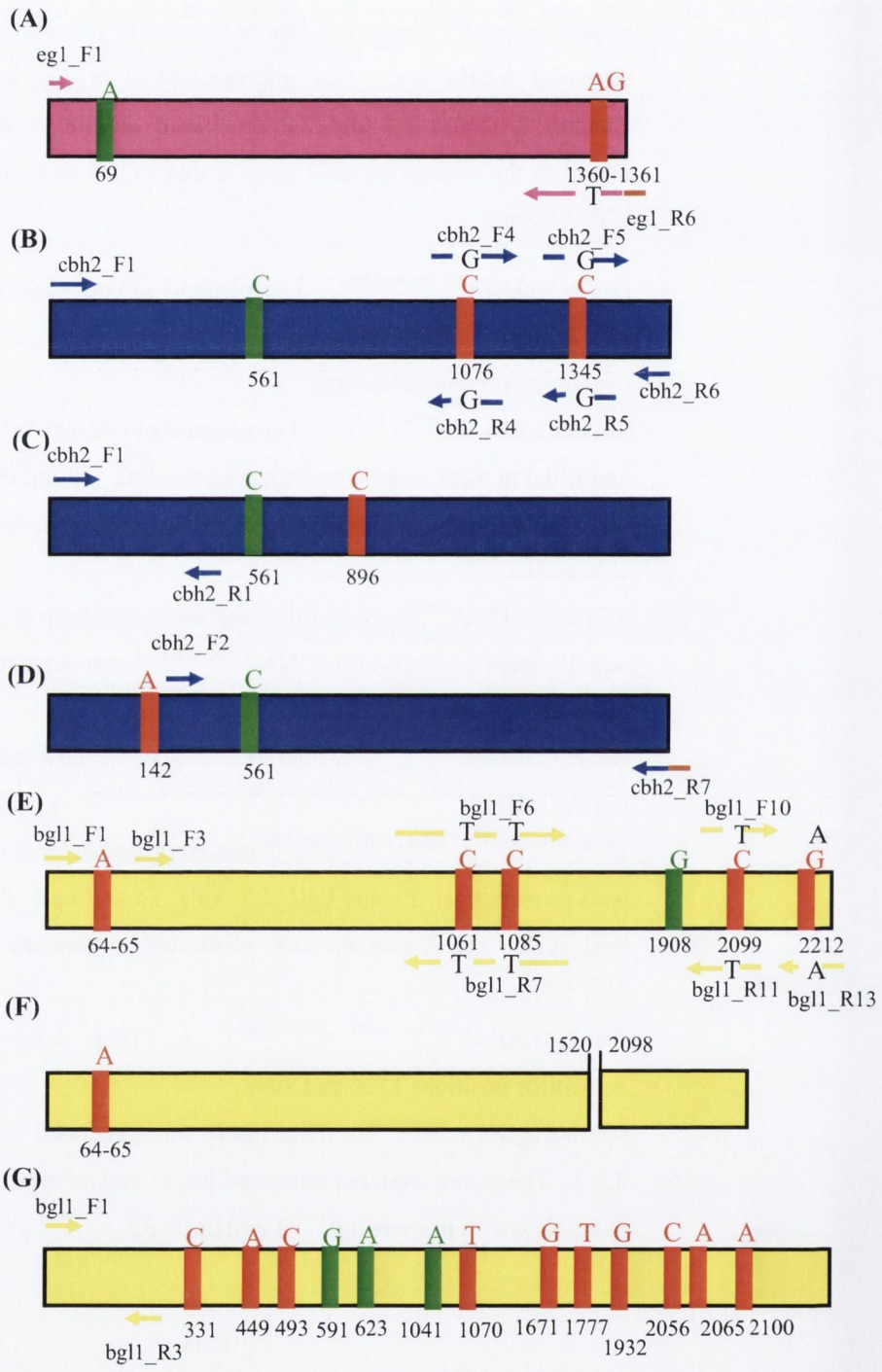
(C) cbh2cDNA^{A896C}. New missence error detected at nucleotide 896. The amplification of insert egl1cDNA $^{T1360A::G1361}$ was repeated as described in (A) to generated insert cbh2cDNA ^{G142C}.

(D) cbh2cDNA ^{G142C}. New missence error observed at nucleotide 142.

(E) bgl1cDNA $^{C64A,\ A65G,T1061C,\ T1085C,\ T2099C,\ A2212G}$. Plasmid pPGK1bgl1cDNA $^{C64A,\ A65G,T1061C,\ T1085C,\ T2099C,\ A2212G}$ was amplified in three overlapping fragments with primers bgl1_F1 and bgl1_R7, bgl1_F6 and bgl1_R11 and bgl1_F10 and bgl1_R13. The fragments were connected to generate bgl1cDNA $^{C64A,\ A65G,\ \Delta1521\text{-}2098}$.

(F) bgl1cDNA$^{C64A,\ A65G,\ \Delta1521\text{-}2098}$. A 578bp deletion occurred between nucleotide positions 1520 and 2098.

(G) bgl1gDNA ^{recombined}. This insert was generated as described in section 3.2.7. The errors were not corrected but 5' end of insert was used as template to amplify with primers bgl1_F1 and bgl1_R3.


(A)
eg1_F1
A
AG
69
1360-1361
T
eg1_R6
(B)
cbh2_F1
cbh2_F4
cbh2_F5
G
G
C
C
C
561
1076
1345
cbh2_R6
G
G
cbh2_R4
cbh2_R5
(C)
cbh2_F1
C
C
561
896
cbh2_R1
(D)
cbh2_F2
A
C
142
561
cbh2_R7
(E)
bgl1_F1
bgl1_F3
bgl1_F6
bgl1_F10
T
T
T
A
A
C
C
G
C
G
64-65
1061
1085
1908
2099
2212
T
T
T
A
bgl1_R7
bgl1_R11
bgl1_R13
(F)
A
1520
2098
64-65
(G)
bgl1_F1
C
A
C
G
A
A
T
G
T
G
C
A
A
331
449
493
591
623
1041
1070
1671
1777
1932
2056
2065
2100
bgl1_R3

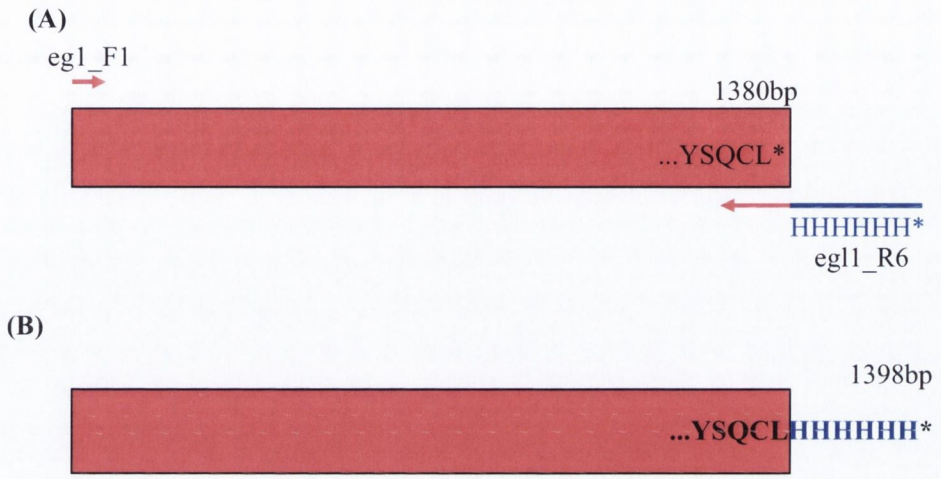


FIG. 3.12. Schematic diagram of attaching a 6xHIS tag to the C terminal of *egl1* coding sequence.

(A) The last six amino acids at the C terminal end of EG1 are displayed. The asterisk at the right hand end represents a stop codon. Primers used are highlighted as coordinated arrows. 6xHIS amino sequence indicated by blue font.

(B) Resultant EG1 fused with C- terminal 6xHIS tag.

3.2.9.2. cbh2cDNA

The *cbh2* insert within plasmid pPGK1cbh2cDNA$^{G1076C, G1345C}$ had three errors located at nucleotide 561, 1076 and 1345. In each instance a mutation of T to C occurred (Table 3.2). The error at 561 was a silent mutation and was not corrected. The errors at 1076 and 1345 changed their respective amino acid codon (Table 3.2). Oligonucleotide primers cbh2_F4 and cbh2_R4 were designed to correct the nucleotide change at position 1076 (Fig 3.11.B). The alteration at base 1345 was corrected in both strands of the DNA with primers cbh2_F5 and cbh2_R5 (Fig 3.11.B). The *cbh2* insert was amplified from plasmid pPGK1cbh2cDNA$^{G1076C, G1345C}$ into three overlapping fragments (Fig 3.11B) using oligonucleotide pairs cbh2_F1 and cbh2_R4, cbh2_F4 and cbh2_R5 and cbh2_F5 and cbh2_R6 (Chapter 2, Table 2.2). Fragments of the expected size were generated, gel extracted and transformed into *S. cerevisiae* S150 (data not shown). Nucleotide sequencing (as described in Chapter 2, section 2.12) of the recovered plasmid pPGK1cbh2cDNA^{A896C} revealed that the errors at 1076 and 1345 were corrected, however a new missense error was found at nucleotide position 896 (Fig 3.11C). The nucleotide at that position changed from A to C and changed the amino acid codon (Table 3.2). The resultant strain was called S150PGK1cbh2$^{cDNA\ A896C}$. The corrective PCR was repeated on the original pPGK1cbh2cDNA$^{G1076C, G1345C}$ plasmid with the same primers cbh2_F1 and cbh2_R4, cbh2_F4 and cbh2_R5 and cbh2_F5 and cbh2_R6. Fragments of the expected size were generated, gel extracted and transformed into *S. cerevisiae* S150. The resultant strain was called S150PGK1cbh2cDNA ^{G142C}. Nucleotide sequencing (Chapter 2, section 2.12) of the recovered plasmid pPGK1cbh2cDNA^{G142C} showed that the errors at 1076 and 1345 bp were again repaired and no error was detected at nucleotide 896 however a missense error was detected at nucleotide position 142bp (3.12.D). The nucleotide changed from G to A (Table 3.2). It was then decided to generate *cbh2* in fragments from different plasmid templates. The *cbh2* gene was amplified in two overlapping fragments using two mutant *cbh2* plasmids as templates. The N-terminus of plasmid pPGK1cbh2cDNA^{A896C} (Chapter 2, Table 2.4) was amplified using oligonucleotide primers cbh2_F1 and cbh2_R1 (Fig 3.11C) while the C-terminus of plasmid pPGK1cbh2cDNA^{G142C} was amplified using primers cbh2_F2 and cbh2_R7 (Fig 3.11D). Oligonucleotide primer cbh2_F7 had the sequence of a 6xHIS tag

incorporated (Chapter 2, Table 2.2). Bands of the expected size were generated (data not shown), gel extracted and co-transformed into S150 along with linearised pPGK1 plasmid (Chapter 2, Section 2.8). The resultant strain was called S150PGK1cbh2HIScDNA ^{wt}. The sequence of this plasmid had no errors other then the silent error at nucleotide position 561 (Table 3.2) and confirmed that the HIS tag had been fused in frame with *cbh2*.

3.2.9.3 pGREGbgl1cDNA

The original cloned *bgl1* cDNA insert had seven errors, six of which were missense and one was a silent mutation (Table 3.2). The nucleotides at position 64-65 changed from CA to AG. This missense occurred within the signal sequence of *bgl1*. Nucleotide 1061 changed from T to C. The same nucleotide alteration occurred at nucleotide 1085 (Table 3.2). Both alterations changed their respective amino acid codes. The nucleotide change at 1908 from A to G was a silent mutation and was not corrected. Nucleotide 2099 had a changed from T to C and the final missense error at 2212 changed from A to G. The primers used to change errors at 1061 and 1085 were bgl1_F6 and bgl1_R7 (Fig 3.11E). Primers bgl1_F10 and bgl1_R11 were used to change the error at 2099. Primer bgl1_R13 was used to correct the alteration at nucleotide 2212. Three overlapping bgl1 fragments; bgl1-correctA, bgl1-correctB and bgl1-correctC were created using primer pairs bgl1_F1 and bgl1_R7, bgl1_F6 and bgl1_R11 and bgl1_F10 and bgl1_R13 respectively (Fig 3.11E). Fragments of 1100, 1080 and 197bp were generated respectively (data not shown). The latter two fragments were connected by primers bgl1_F6 and bgl1_R13. This generated a 1225bp fragment (data not shown). The fragments were then co-transformed into yeast along with linearised pPGK1 (Chapter 2, Section 2.8). Inadvertently the errors at 64-65bp were uncorrected. The sequence of this insert indicated that the mutations at position 1061 and 1085 were corrected. Interestingly, there was a 578bp section of *bgl1* missing from the plasmid. The section 1520-2099bp of *bgl1* cDNA sequence was deleted (Fig 3.11F and Table 3.2). Based on the *bgl1* cDNA sequence, fragment bgl-correctB lies between nucleotide positions 1046-2124. The bgl1-correctC fragment runs from position 2074-2235bp. Within bgl1-correctC fragment, there is an 11bp sequence AACAGCAACGT that occurs from 2088 to 2098. However this 11bp run also occurs within fragment bgl1-correctB at

nucleotides 1510-1520. The deletion occurs between bases 1521 to 2098. It would appear that during recombination within yeast the 11bp sequence at 1510-1520bp annealed to the homologous sequence at 2088-2098bp, thereby excluding *bgl1* nucleotides between 1521-2098bp, which yielded a gap of 578bp. The strain was called S150PGK1bgl1cDNA $^{C64A,\ A65G,\ \Delta 1521\text{-}2098}$ based on the sequence analysis of the recovered plasmid (Chapter 2, section 2.12).

To correct the final error in *bgl1*, it was decided to re-amplify bgl1 from different plasmid sources. Nucleotide sequencing of the bgl1 insert with pGAL1bgl1gDNArecombined (Section 3.2.6) indicated that the 5' end of the insert had no errors from nucleotides 1-330 (Table 3.3). Thus the 5' end of *bgl1* was amplified from this clone with primers bgl1_F1 and bgl1_R3 (Fig 3.11G). The 3' end was reamplified from plasmid pPGK1bgl1cDNA $^{C64\mathbf{A},\ A65\mathbf{G},T1061\mathbf{C},\ T1085\mathbf{C},\ T2099\mathbf{C},\ A2212\mathbf{G}}$ with primer pairs bgl1_F3 and bgl1_R7, bgl1_F6 and bgl1_R11, bgl1_F10 and bgl1_R12 (Fig 3.11E)

Bands of the expected size were generated, gel extracted and co-transformed into *S. cerevisiae* S150 along with with linearised pPGK1 vector. The resultant strain was called S150PGK1bgl1cDNA ^{wt}. Nucleotide sequence analysis indicated that the *bgl1* insert had no remaining missense errors (Table 3.2).

Table 3.1 Summary of mutations and alterations within genomic DNA inserts

Insert	Location of error within genomic DNA sequence	Nature of mutation/alteration	Basepair change (from/to)	Codon change
eg1_gDNA	ND	n/a	n/a	n/a
eg1_TACTAAC	827 (intron1)	Insertion	TACTAAC	n/a
	1489 (intron2)	Insertion	TACTAAC	
cbh2_gDNA	666	Silent	G to C	n/a
	758	Missense	A to G	K to R
	899 (intron)	Missense	G to A	C to Y
	1163	Missense	T to C	L to S
	1271	Missense	G to C	R to P
	1540	Missense	G to C	A to P
bgl1_gDNA	78 (intron1)	Insertion	C	I to H
	92 (intron1)	Insertion	G	Frameshift
	110 (intron 1)	Missense	A to G	
	402	Missense	C to T	
	564	Missense	C to T	
	662	Missense	G to A	
	1112	Missense	A to T	
	1141	Deletion	T	
	1742	Missense	A to T	
	1848	Missense	T to C	
	2191	Missense	C to T	

Table 3. 2 Summary of mutations within cDNA inserts

Insert	Location of error in cDNA sequence	Nature of mutant	Basepair change (from/to)	Codon change
egl1cDNA *T1360A::G1361*	69	Silent	G to A	n/a
	1360	Missense	T to A	Y to R
	1361	Insertion	G	frameshift
egl1HIS cDNA [wt]	69	Silent	A to G	n/a
cbh1_cDNA[wt]	nd	n/a	n/a	n/a
cbh2cDNA *G1076C, G1345C*	561	Silent	G to C	n/a
	1076	Missense	G to C	R to P
	1345	Missense	G to C	A to P
cbh2cDNA *A896C*	561	Silent	G to C	n/a
	896	Missense	A to C	N to T
cbh2 cDNA *G142C*	142	Missense	G to A	G to R
	561	Silent	G to C	n/a
cbh2HIS cDNA [wt]	561	Missense	G to C	n/a
bgl1cDNA *C64A, A65G, T1061C, T1085C, T2099C, A2212G*	64	Missense	C to A	H to S
	65	Missense	A to G	H to S
	1061	Missense	T to C	V to A
	1085	Missense	T to C	I to T
	1908	Silent	A to G	n/a
	2099	Missense	T to C	F to S
	2212	Missense	A to G	A to T
bgl1 cDNA *C64A, A65G, Δ1521-2098*	64	Missense	C to A	H to S
	65	Missense	A to G	H toS
	1521-2098	Deletion		Frameshift
bgl1cDNA[wt]	1908	Silent	A to G	n/a

Table 3.3 Summary of mutations within recombined genomic DNA bgl1 insert

Insert	Location of error within bgl1 cDNA sequence	Nature of mutation	Basepair change (from/to)	Codon change
bgl1gDNArecombined	331	Antisense	C to T	Q To STOP
	449	Missense	A to G	K to R
	493	Missense	C to T	P to S
	591	Silent	G to A	n/a
	623	Silent	A to G	D to G
	1041	Silent	A to T	Q to H
	1070	Deletion	T	Frameshift
	1671	Missense	G to T	
	1777	Missense	T to C	
	1932	Missense	G to T	
	2056	Missense	C to T	
	2065	Missense	A to G	
	2100	Missense	A to T	

3.2.10 Transformation of cellulase inserts into *S. pastorianus*:

Once all errors in the DNA sequences were repaired the plasmids containing the cellulase gene cassettes were cloned into *S. pastorianus*. Engineering industrial yeast species such as *S. pastorianus* is challenging (Saerens et al., 2010). *S. pastorianus* strains have allopolyploid (and sometimes aneuploid) genomes that show a very poor sporulation capacity and a low degree of spore viability. Many industrial yeast strains are homothallic, which makes it difficult to get stable haploids for mating (Saerens et al., 2010). *S. cerevisiae* is genetically well characterised and therefore easier to genetically engineer and more is known about recombination mechanisms. Cloning genes into *S. pastorianus* is further impeded by the lack of auxotrophic mutants which can be used for selection of transformants, therefore selection using antibiotic resistance is required. The pGREG vector contains a KanMX cassette and hence cells that have uptaken the plasmid will grow on the aminoglycoside antibiotic geneticin (G418). Secondary screening for transformants is also achieved by PCR amplification of transformant DNA using insert-specific primers (Chapter 2, Section 2.6).

S. pastorianus strain C10-51 as described in Chapter 1 Section 1.14 was chosen for this study because it was shown to be a stress tolerant yeast (James et al., 2008).

Transformation of strain C10-51 with plasmids containing the cellulase encoding genes was performed as described in Chapter 2, Section 2.8. Transformants were selected on YPD agar plates containing 0.2g l^{-1} G418 (See Chapter 2, Section 2.1). The resultant strains were listed in Chapter 2, Table 2.5. Total DNA was extracted from C10-51 transformants and PCR was carried out using gene-specific primers that amplified the entire length of the respective insert (Fig 3.13.A). DNA from *S. pastorianus* C10-51PGKegl1HIScDNA^{wt} was amplified by primers egl1_F1 and egl1_R6 (Chapter 2, Table 2.2) to yield a fragment of 1466bp (Fig 3.13 A. and B, lane 2). PCR amplification of DNA from *S. pastorianus* C10-51PGK1cbh2HIScDNA^{wt} using primers cbh2_F1 and cbh2_R7 (Chapter 2, Table 2.2) generated a fragment of the expected size 1.5kb (Fig.3.13 A and B, lane 3). *S. pastorianus* C10-51PGK1bglcDNA^{wt} amplified with primers bgl1_F1 and bgl1_R12 (Chapter 2, Table 2.1) yielded a band of the expected

2.3kb size (Fig. 3.13 A and B, lane 4). Hence it appeared that colonies of *S. pastorianus* C10-51 had been engineered to possess cellulase encoding inserts (Table 2.5).

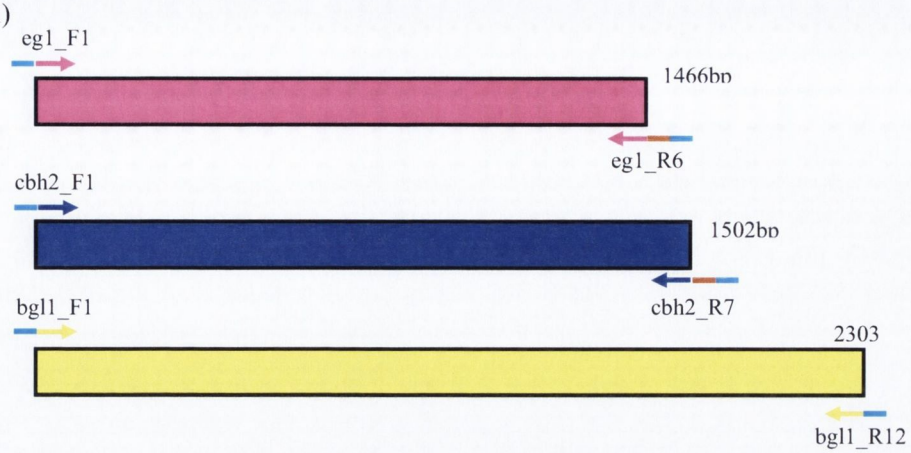
FIG. 3.13. Amplification of cDNA derived cellulase inserts from transformed *S. pastorianus* C10-51.

(A) Schematic diagram of cDNA derived inserts. Oligonucleotide primers colour co-ordinated for respective insert. Turquiose and brown regions represented pGREG sequence overhangs and HIS tag respectively. egl1HIScDNA^{wt} (pink) amplified by primers eg1_F1 and eg1_R6. cbh2cDNAHIS^{wt} (blue) amplified using primers cbh2_F1 and cbh2_R7. bgl1cDNA^{wt} (yellow) amplified using primers bgl1_F1 and bgl1_R12

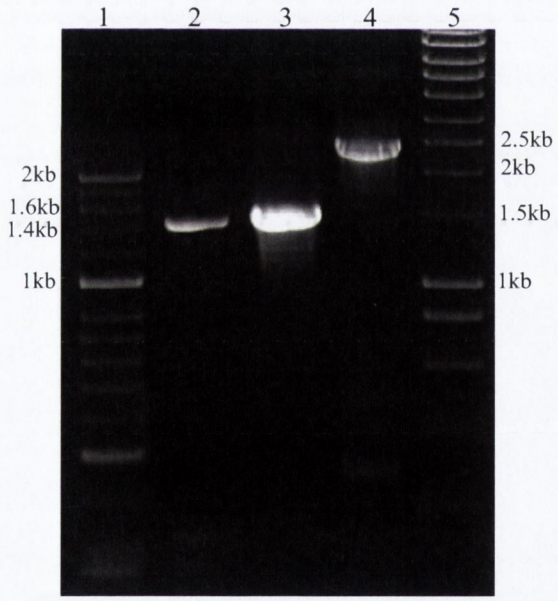
(B) Colony PCR of *S. pastorianus* C10-51 colonies transformed with separate cellulase encoding inserts amplified by corresponding gene-specific primers.

Lane 1; Hyperladder I. Lane 2: C10-51PGK1egl1HIScDNA^{wt} (1466bp). Lane3: C10-51PGK1cbh2HIScDNA^{wt} (1502bp). Lane 4: C10-51PGK1bgl1cDNA^{wt} (2303bp).

(A)


eg1_F1
1466bp
eg1_R6
cbh2_F1
1502bp
cbh2_R7
bgl1_F1
2303
bgl1_R12

(B)


1
2
3
4
5
2kb
1.6kb
1.4kb
1kb
2.5kb
2kb
1.5kb
1kb

3.2.11 Integration of *egl1* into the *S. pastorianus* genome

Plasmids by their nature are mobile genetic elements and unless they are kept under constant selection they can become unstable. Integration of the inserts would insure their stability between generations even under non-selective conditions. An attempt was made to integrate the GAL1egl1cDNA cassette onto the chromosome of *S. pastorianus* (Fig 3.14.A). The target of this integration was the intergenic region between YPR159W-YPR160W of chromosome XVI. This is an area of apparent redundancy that occurs upstream of the non-essential glycogen phosphorylase gene GPH1 (Usher and Bond, 2009).

Plasmid DNA from pGAL1egl1cDNA$^{T1360A::G1361}$ was amplified from the GAL1 promoter to the KanMX (Fig 3.14.A) cassette with sequence specific primers I-GAL1_F and I-Kan_R (Chapter 2, Table 2.3). These primers also had 35bp overhangs with sequence homology to the intergenic region of chromosome XVI (Fig. 3.14.A). The fragment was approximately 3.3kb which corresponded to the expected size (Fig.3.14.B, lane 2). The fragment was gel extracted and transformed into *S. pastorianus* strain C10-51 (Chapter 2, Section 2.8). Transformants were selected for by resistance to G418 antibiotic. To confirm that the integration occurred at the site intended, DNA was extracted from the *S. pastorianus* transformants and PCR amplification was carried out using primers spanning the junction between the intergenic region in *S. pastorianus* and the inserted cellulase encoding insert (Fig 3.15.A). The results of the PCR analysis are displayed in Figure 3.15.B and C. The presence of the *egl1* fragment within the extracted *S. pastorianus* DNA was confirmed by successfully amplifying a product with primers egl1_F1 and egl1_R5 (Fig.3.15.B, lane 2). The band of approximately 1.5kb was observed, which corresponds to the expected size of 1446bp. To test if the integration occurred at the expected site, DNA was amplified using primers that anneal within the cassette and the intergenic region. The primers used were Kan_F and I-R1 (Chapter 2, Table 2.3). The latter primer lies 1004bp downstream of the integration site. The kan_F primer amplifies from the start of the kanMX marker. If the integration cassette inserted at the intended location within the intergenic region a band of 2051bp was expected. The PCR reaction across the integration junction did not yield a DNA fragment (Fig.3.15.B, lane 3).

However, amplification of the same DNA using internal intergenic region primers I-F1 and I-R2 generated a band of approximately 900bp, corresponding to the expected size of 883bp (Fig 3.15.B, lane 4).

Another PCR was carried out on the same extracted DNA using primers from within the *eg1* insert. As shown in Fig. 3.15.C, DNA amplified by a primer upstream of the intergenic region, YPR159_F and eg1_R1 yield a product of approximately 200bp, however a band of 1210bp was expected (lane 2). Amplification across the integration junction using primers eg1_F2 and I-R1 was predicted to generate a band of 3082bp. This reaction failed to yield the expected product (Fig.3.13.C, lane3). As a control the intergenic region was amplified using the oligonucleotide primers I-F2 and I-R1 (Chapter 2, Table 2.3). The observed band of approximately 500bp corresponded to the expected size of 489bp was observed (Fig.3.13.C, lane 4). Finally the DNA was amplified with the original primers used to generate the initial integration cassette, I-GAL1_F and I-Kan_R. The predicted 3.3kb band was not detected (Fig. 3.13.C, lane 5), thus indicating that while G418+colonies were detected, the *egl1* gene cassette did not integrate at the intended location. However, since a DNA fragment corresponding to the *egl1* gene was detected, it is possible that the gene may have integrated at a different chromosomal location.

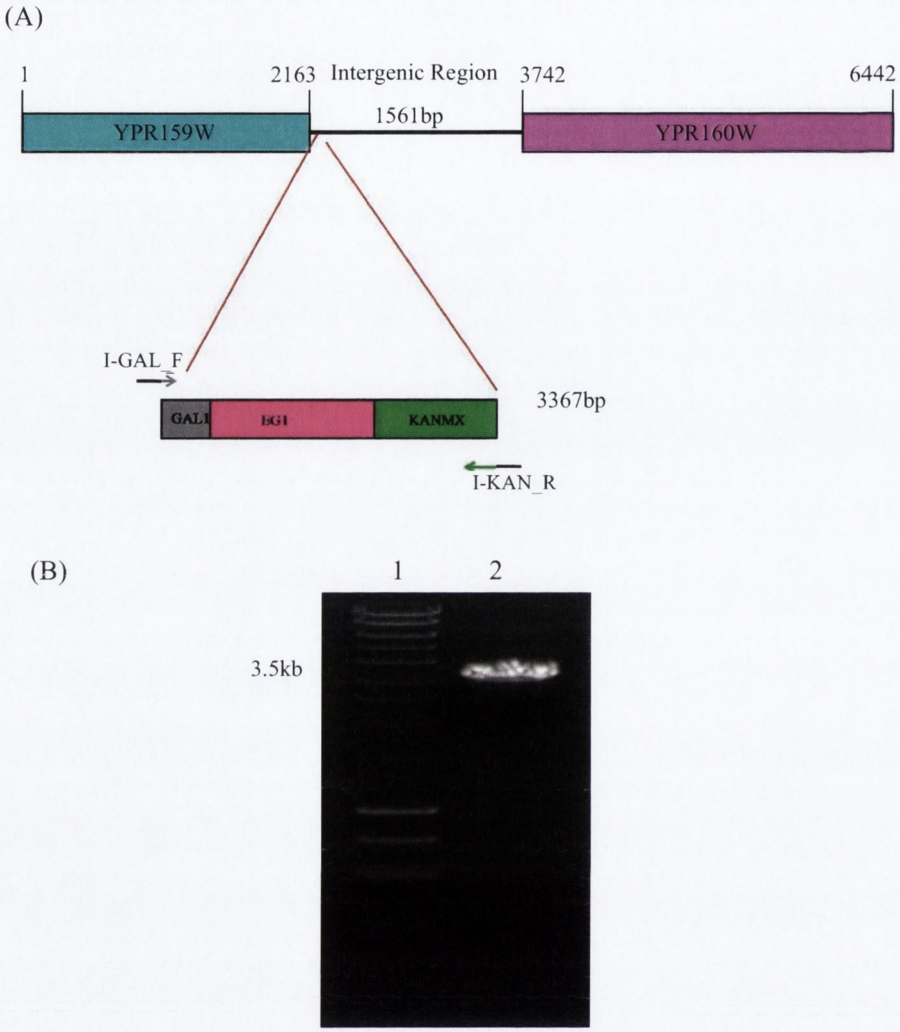


FIG 3.14. Integration of *egl1* cassette in *S. pastorianus* genome.

(A) Schematic diagram of *egl1* cassette to be integrated into genome of *S. pastorianus*. Primers used are highlighted and colour coordinated to the regions they amplify. Red lines highlight target integration site

(B) PCR products of pGALeg1cDNA plasmid amplified with integration primers. Lane 1; Hyperladder 1. Lane 2; 50ng pGAL1eg1cDNA$^{T1360A::G1361}$ plasmid DNA using primers I-GAL1_F and I-KAN_R.

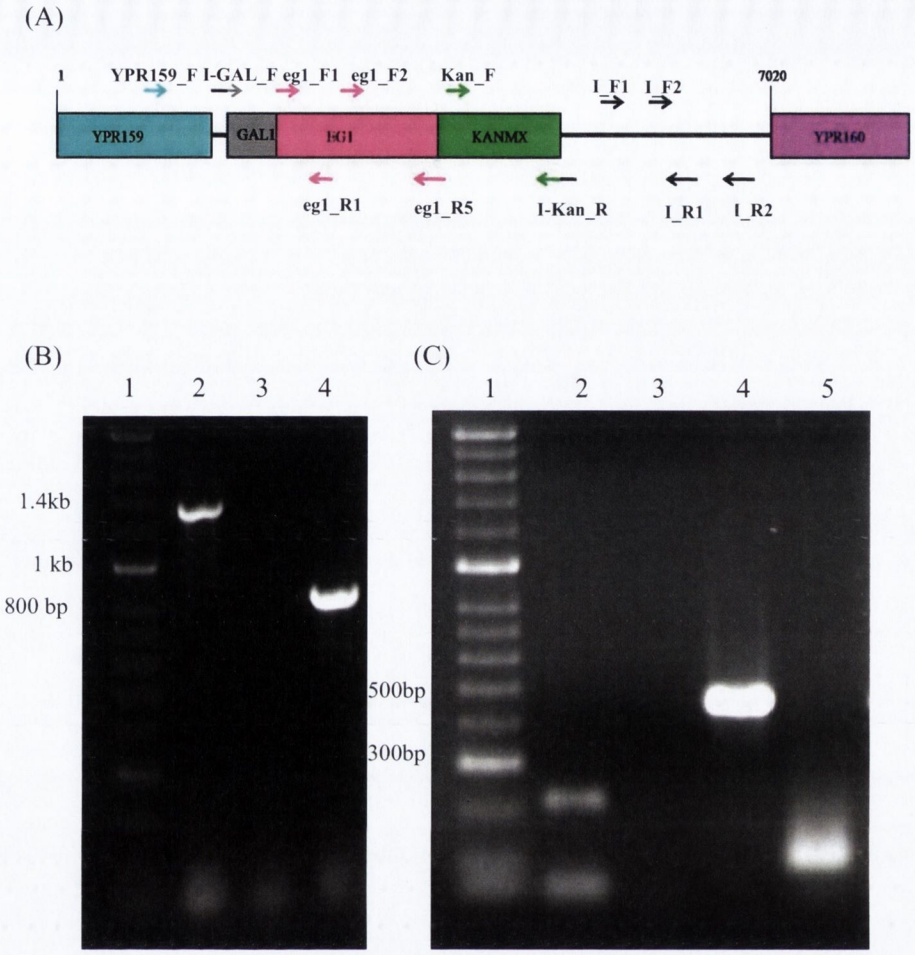


FIG.3.15. PCR of DNA extracted from putative GAL1_egl_kanMX intergrated *S. pastorianus* C10-51.

(A) Schematic diagram of putative integration site.

(A) Lane 1; Hyperladder 2. Lane 2; egl_F1 and egl_R4. Lane 3; Kan_F and I_R1. Lane 4; I_F1 and I_R2

(C) Lane 1; Hyperladder 2. Lane 2; YPR159_F and egl_R1. Lane 3; egl1_F2 and Intergenic_R1. Lane 4; I_F2 and I_R1. Lane 5; I-GAL_F and I-Kan_R.

3.3 Discussion

The aim of the research described in this chapter was to clone copies of the three major classes of *T. reesei* cellulase genes into two different species of *Saccharomyces*, namely *S. cerevisiae* and *S. pastorianus*. The cloning of cellulase encoding genes into *S. cerevisiae* in various combinations has been previously been documented (Chapter 1, Section 1.12). In most cases cellulase genes from different species were combined to recreate cellulase activity *S. cerevisiae*. The genes were controlled by different promoters (PGK1, GAPDH, GAL1, ENO1) and employed either the native signal secretory signals or *S. cerevisiae* encoded signals sequences to direct the protein product to the exterior of the cell or to tether the proteins to the cell wall. Taken together the accumulated data demonstrate that cellulases can be produced in yeast cells and in one case (Den Haan et al., 2007b) the recombinant species were capable of growing on amorphous cellulose as a sole carbon source. In the latter case, co-expression of *T. reesei* EG1 and *Saccharomycopsis fibuligera* (BGL1) was sufficient to allow growth on 1% on amorphous PASC, although very low yields of ethanol were produced.

Despite the many examples of the expression of cellulase genes in *S. cerevisiae*, there has been no attempt to recreate a cellulase system from a single organism in a yeast species. Since cellulases are known to act synergistically to degrade cellulase, the aim of this study was to express *T. reesei* genes *egl1*, *bgl1* and *cbh2* from a common promoter and using the native signal sequences of the *T. reesei* proteins in both *S. cerevisiae* and *S. pastorianus* strains. *S. pastorianus* strains were chosen as this species show high fermentation rates and are generally more robust due to the allopolyploid nature of its genome, therefore, it would be of interest to compare the ability of *S. pastorianus* and *S. cerevisiae* strains to carry out simultaneous saccharification and fermentation of cellulose.

Initially, genomic copies of the three genes were cloned into *S. cerevisiae* S150 strain, however initial analysis of gene expression by reverse transcriptase-PCR indicated that little or no RNA was produced from the cloned gene (data not shown). To improve the likelihood that the intronic sequences contained within the genes would be removed, a branch point consensus sequence, TACTAAC, was introduced into the gene. Ultimately it was decided that a better strategy

would be to clone cDNA copies of the cellulase genes. This was achieved by reverse transcribing RNA from *T. reesei* with gene specific primers. Due to the low expression levels of some of the *T. reesei* cellulase genes and the high GC content, full length cDNAs could not be produced; therefore the cDNAs were prepared in sections and then recombined *in vivo* using the endogenous homologous recombination ability of *Saccharomyces* species. PCR analysis of the resultant cDNA clones confirmed that this strategy was successful, however sequence analysis of both the genomic and cDNA clones revealed that a number of base mutations had been introduced into the genes during the cloning process. The initial genomic DNA cloning was performed using the enzyme Taq Polymerase, which lacks a 3' to 5' exonuclease proof reading activity. Error rates in the range of $2x10^{-5}$ – $1.1x10^{-4}$/bp have been reported for Taq Polymerase. The largest gene cloned in this study was *bgl1* at approximately 2500bp in length. Based in the mutations observed within the cloned *bgl1* gene, an error rate of $4.4x10^{-3}$/bp was observed. This is larger than expected from the use of Taq Polymerase. Similar error rates of $5.2x10^{-3}$/bp and $3.7x10^{-3}$/bp were observed for the *bgl1* genomic copy lacking introns (bgl1gDNA^{recombined}) and for the cbh2gDNA genomic clone respectively.

Due to the high error rate observed within genes cloned following amplification with Taq Polymerase, all subsequent PCR amplifications were performed with the enzyme Verizyme (Pfu) DNA Polymerase (York-Bio), which in addition to possessing 5'→3' synthetic activity, has 3'→5' proofreading and no 5'→3' exonuclease activities. The error rate for Verizyme is reported as being lower than that of Taq Polymerase ($1\text{-}6x10^{-5}$/bp). Amplification of all cDNA copies were carried out with this enzyme except for the initial copy of *bgl1* (bgl1cDNA$^{C64A, A465G, T1061C, T1085C, T2099C, A2212G,}$ Chapter 3, Table 3.2) which was amplified using a Taq:Verizyme ratio of 10:1. Error rates were substantially reduced using Verizyme alone ($6.3x10^{-4}$ – $1.7x10^{-3}$) as compared to the mixture of Verizyme and Taq ($2.8x10^{-3}$) but still remained higher than the expected error rate. It is possible that cloning by homologous recombination into *S. cerevisiae* may have increased the error rate through the introduction of errors during the recombination process. Alternatively the high GC content of *T. reesei* genes may have contributed to the high error rate through impeding the synthetic rate of the Polymerases. Site directed mutagenesis was carried out to correct the missense mutations in the cDNA

clones and eventually wt copies of each cDNA were obtained. However, even during this repair process, additional mutations arose. In one clone bgl1 $^{C64A, A65G, \Delta 1521\text{-}2098}$, a 578bp section of the gene was removed, most likely as a result of erroneous recombination due to the presence of an 11bp common sequence found at nucleotides 1510-1520 and 2088-2098 respectively.

In addition to cloning the cellulase genes into the parental pGREG vector under the control of the GAL1 promoter, each gene was subsequently cloned into a vector in which the PGK1 promoter replaced the GAL1 promoter. This will allow the constitutive expression of the genes as may be required for maximum expression under fermentation conditions. Finally, since plasmids by their nature are mobile and loss can occur if cells are not maintained under selective conditions, an attempt was made to introduce the *egl1* gene into the chromosome of *S. pastorianus*. A DNA fragment encompassing the GAL1 promoter, *egl1* gene and the KANMX cassette was amplified from the pGREG vector. The amplified region contained 35bp overhangs homologous to the site of integration at YPR159 – YPR160 on chromosome XVI. While G418 resistance colonies were obtained, PCR analysis of the recombinants indicated that the fragment had not integrated at the expected location. However, a fragment corresponding to the entire *egl1* region was amplified from the recombinant clones. This suggests that the *egl1* gene may have integrated at a different location in the genome. While successful integration of antibiotic resistance genes or auxotrophic markers in the *S. cerevisiae* genome has been achieved with just 35bps of homologous overhangs (Wendland, 2003), due to the polyploid nature of the *S. pastorianus* genome and the concomitant increase of possible integration sites, longer homologous regions might be needed to ensure successful integration. *S. pastorianus* is a hybrid species resulting from the hybridisation of two yeast species, one of which is almost identical to *S. cerevisiae*. The second parental strain closely resembles *S. bayanus*. Thus *S. pastorianus* possesses *S. cerevisiae*-like genes. The GAL1 promoter (460bp) used in the integration fragment also shares homology to the *S. cerevisiae*-like copy of the GAL1 gene, therefore it is possible that integration at the GAL1 sites is favoured over integration at YPR159-YPR160. Due to lack of time it was not possible to determine the exact location of the integrated cassette, however DNA sequence analysis of the integrant should answer this question

in the future. Successful integration of DNA fragments into the YPR159-YPR160 region of chromosome XVI of *S. pastorianus* has subsequently been achieved with 200-300bp homologous regions in our laboratory.

The serendipitous introduction of errors into the coding regions of *egl1*, *bgl1* and *cbh2* during amplification of the DNA sequences provides us with a series of mutant copies of the genes, which can be used to identify regions of the proteins that may be important for function. Additionally, the cloning of genomic copies of the genes containing native introns or *S. cerevisiae*-optimised introns provides us with the tools to investigate if *T. reesei* genes containing introns can be properly spliced in *S. cerevisiae*. In the following chapter the cloned genomic and cDNAs were analysed for the production of functional EGI, BGLI and CBHII in both *S. cerevisiae* and *S. pastorianus*.

Chapter 4: Demonstration of cellulase enzyme activity in yeast strains expressing cellulase encoding genes

4.1 Introduction

The aim of the experiments described in this chapter was to investigate if the clones produced in Chapter 3 demonstrated any functional cellulase enzyme activity. Endoglucanase activity of EGI enzyme can be demonstrated by its unique ability to hydrolyse carboxymethylcellulose (CMC) (Wood and Bhat, 1988). The CBH and BGL enzymes possess very little activity against CMC (Okada et al., 1998). To demonstrate BGLI activity, the chromogenic reagent X-glucoside (5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside) was used (Cummings and Fowler, 1996). The EGI and CBHII enzymes do not metabolize X- glucoside making it a unique assay for detection of BGLI. While no specific substrate is available for cellobiohydrolase (Wood and Bhat, 1988), activity can be detected

by measuring the hydrolysis of a soluble cellulase substrate such as 4-methylumbelliferyl-β-D-lactoside (Voutilainen et al., 2007) or by monitoring the release of total or reduced sugars following adsorption of the enzyme onto phosphoric acid swollen cellulose (PASC) or crystalline cellulose such as Avicel (Den Haan et al., 2007a).

For the purposes of degrading cellulose into glucose, it is critical that any cellulases produced by the clones are correctly secreted outside the host cell. In the native species *T. reesei* EGI, CBHII and BGLI are extracellular proteins. To ensure secretion of the cellulase enzymes once expressed in the yeast host, the gene insert contained the native *T. reesei* secretory signal. The purpose of retaining the native signal sequence was to investigate if it would be recognised and cleaved correctly by the *S. cerevisiae* secretory mechanism. Thus, enzyme activity of the secreted proteins was measured in the supernatant from the growth cultures of the yeast clones.

In the previous chapter, both wt and mutant cDNA and genomic copies of *egl1*, *bgl1* and *cbh2* were successfully cloned into the pGREG vector under the control of the PGK and GAL1 promoters. The GAL1 promoter is highly inducible by growth of strains in 2% galactose and is fully repressed in the presence of 2% glucose. Therefore, high levels of expressed gene products can be obtained for initial analysis of the cloned genes. However, since our ultimate goal was to achieve growth of yeast strains expressing the cellulase genes on cellulose substrates as a sole carbohydrate source, the presence of galactose in the medium would reduce the utilisation of cellulose at the initial

stages of fermentation as it is easily transported into the cell. In addition, should cellulase activity be achieved, the release of glucose from cellulose would repress the GAL1 promoter thus effectively inhibiting the subsequent production of the cellulases. Consequently, the cellulase genes were placed under the control of the PGK1 promoter. Previous studies have shown that PGK is constitutively expressed under fermentation conditions (James et al., 2003).

As described in Chapter 3, in addition to generating wildtype copies of the cellulase genes, a number of clones containing missense mutations in the coding region of the cellulase genes were also generated (Table 3.2). The generation of mutant copies of the genes affords an opportunity to identify regions of the proteins that may be essential for functionality. The nucleotide sequencing of the *egl1* insert within S150GAL1egl1cDNA $^{T1360\mathbf{A}::\mathbf{G}1361}$ contained mutations at the 3' end of the insert. This mutation (T to A) followed by an insertion of a G resulted in the alteration of a tyrosine to an arginine at amino acid residue 454 (Fig 4.1 A). An inserted guanine at nucleotide position 1361 caused a frame shift and changed the last 6 amino acid residues YYSQCL to LLAMPL in the clone. This occurred in the carbohydrate binding module (CBM). However the frame shift mutation also caused the native stop codon to be knocked out of frame and resulted in the translation of EGI with an additional 6 amino acids, GRPRVM attached. This increased the size of EGI from the expected 459 to 466 amino acids.

The *cbh2* insert within S150PGK1cbh2cDNA$^{G1076\mathbf{C},\ G1345\mathbf{C}}$ had missense mutations at nucleotide positions 1076 and 1345. The errors occurred within the catalytic domain of the enzyme at amino acid residues 359 and 449 respectively (Fig 4.1.B) replacing arginine and alanine residues to prolines respectively. Addition of prolines to a sequence may disrupt the native structural conformation. The generation of this double mutant strain would determine if potential changes to the structure of the catalytic domain impact on enzyme activity. During the correction of the errors in *cbh2*, a number of additional mutants were generated. Strain S150PGK1cbh2cDNA$^{A896\mathbf{C}}$ (Fig 3.11.C) contained a mutation at nucleotide position 896 which altered an asparagine to threonine at amino acid residue 299 (Fig 4.1.C) while strain S150PGK1cbh2cDNA$^{G142\mathbf{C}}$ had a mutation at nucleotide 142 which coded for amino acid residue 48 and altered the wildtype glycine or a mutant arginine.

This mutation occurs within the CBM (Fig 4.1.D) of CBHII. A final round of site-directed mutagenesis resulted in S150PGK1cbh2HIS cDNA ^{wt}; the *cbh2* insert had no missense errors and therefore had the wildtype coding sequence.

The original *bgl1* cDNA clone in strain S150PGK1bgl1cDNA$^{C64A,\ A65G,\ T1061C,\ T1085C,\ T2099C,\ A2212G}$ had six missense mutations. The nucleotide errors at position 64 and 65 altered a histidine to a serine (Table 3.2). These errors occurred within the secretory sequence of *bgl1*at amino acid 22 (Fig 4.1.E). The generation of a mutant with a mutated secretory signal was of benefit because it could determine if altering the signal sequence would prevent secretion. Four other missense errors were located at nucleotide positions 1061, 1085, 2099 and 2212 within the catalytic domain which coded for amino acid residues 354, 362, 700 and 738 respectively. The mutations at the aforementioned nucleotides altered their respective wildtype amino acid valine, isoleucine, phenylalanine and alanine to the mutant alanine, threonine, serine and threonine respectively. A second mutant strain S150PGK1bgl1 cDNA$^{C64A,\ A65G,\ \Delta1521\text{-}2098}$ contained a deletion between nucleotide positions 1521-2098 (Fig 4.1F). This clone also contained the original errors at positions 64 and 65. A final *bgl1* clone was generated that was full length, had the errors within the secretory signal corrected and no missense mutations.S150PGK1bgl1cDNA ^{wt} (Table 2.4).

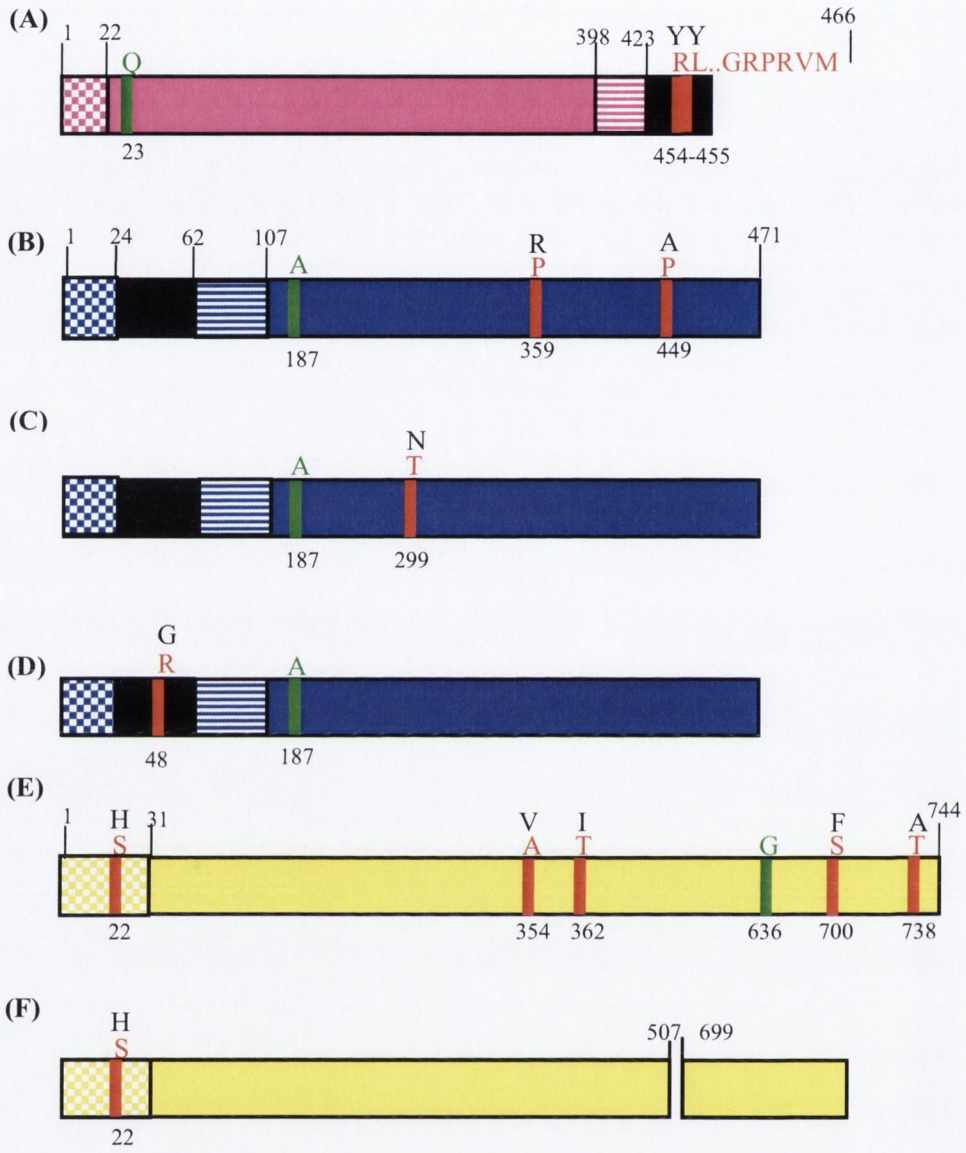
In addition to measuring enzyme activity, attempts were made to tag one of the cellulase proteins, EGI, by introducing a 6x HIS sequence onto the C- terminal of the protein coding region of the gene. If successful, the fusion protein could be possibly purified by nickel column chromatography.

Finally, to determine if the production of any functional cellulase enzymes affected the fitness of the host cell, the growth of yeast strains containing cellulase encoding inserts were compared to the parental host strains.

FIG 4.1. Schematic diagram indicating the location of errors within the cellulase enzymes

Missense amino acids are indicated in red. The wildtype amino acid code is indicated by black lettering. Silent mutations with unaltered amino acids are indicated by green. Chequered pattern represents secretory signal. Solid colour pattern represents catalytic domain (CD). Linker is represented by horizontal lines. Carbohydrate binding module (CBM) is represented by black region.

(A) egl1cDNA$^{T1360A::61361}$. A silent mutation at amino acid residue 23. Missense mutant at residue 454 and an insertion mutation at 455 caused a frameshift which lead to the translation of an addition 6 amino acids at the C-terminus of the protein. This increased the size of EGI to 466 amino acids.
(B) cbh2cDNAcDNA $^{G1076C, G1345C}$. A silent mutation at residue 187 and two missense mutations at amino acids 359 and 449 were detected.
(C) cbh2cDNA ^{A896C}. A new missense mutation was detected at amino acid 299.
(D) cbh2cDNA ^{G142C}. A new missense error occurred at nucleotide 142.
(E) bgl1cDNA $^{C64A, A65G,T1061C, T1085C, T2099C, A2212G}$. A silent mutation occuring at amino acid 636 was observed. The missense mutant at residue 22 was located in the signal peptide. Missense mutations within the CD were detected at 354, 362, 700 and 738.
(F) bgl1cDNA $^{C64A, A65G, \Delta1521-2098}$. The missense mutation in the secretory sequence remained and a 578bp deletion occurred between nucleotide positions 507and 699.


(A)
1
22
Q
23
398
423
YY
RL..GRPRVM
466
454-455
(B)
1
24
62
107
A
187
R
P
359
A
P
449
471
(C)
A
187
N
T
299
(D)
G
R
48
A
187
(E)
1
H
S
22
31
V
A
354
I
T
362
G
636
F
S
700
A
T
738
744
(F)
H
S
22
507
699

4.2 Results

4.2.1 Detecting EGI activity with carboxymethylcellulose (CMC).

The growth culture supernatant of yeast transformed with the *egl1* gene was examined for endoglucanase activity. Detection of the EGI enzyme in the supernatant would confirm that the native *T. reesei* secretory sequence was recognised and correctly processed by the host *Saccharomyces* secretory mechanism.

Carboxymethyl cellulose is a soluble derivate of cellulose and is widely used to detect endoglucanase activity (Wood and Bhat, 1988). CMC consists of contiguous β-1,4-linked methylated D-glucopyranosyl units (Teather and Wood, 1982). These units are hydrolysed by endoglucanases. EGI enzyme is demonstrated by staining CMC with Congo Red which interacts with polysaccharides followed by destaining with NaCl to reveal a zone of hydrolysis (Chapter 2, Section 2.17).

Yeast strains transformed with *egl1* were grown at 30°C for 48h in 50 ml of either YP or SC containing galactose or glucose media depending on whether the plasmid contained the GAL1 or PGK1 promoter respectively. The CMC assay was peformed as described in Chapter 2, Section 2.17 using 20 µl of unconcentrated supernatant.

The rationale for cloning *T. reesei* genomic *egl1* DNA (Chapter 3, Section 3.2.1) was to investigate if the splicing mechanisms of *S. cerevisiae* would recognise and splice the two introns (See Chapter 3, Fig 3.1) within the gene to express and secrete functional EGI enzyme. The introns of *S. cerevisiae* contain a unique branch site consensus sequence, TACTAAC (Thomas et al., 1999). To improve the possibility that the intronic sequences in *T. reesei* cellulase genes would be spliced by yeast, the introns of *egl1* were altered to contain a TACTAAC sequence (Chapter 3, Section 3.2.5). The CMC assay performed on the supernatant of S150GAL1egl1gDNA grown in 50ml SC galactose and S150PGK1egl1^{TACTAAC} grown in SC glucose indicated that no EGI activity was detected and hence *S. cerevisiae* was unable to splice foreign introns (Fig 4.2A, and D). As a positive contol 0.06 mU of commerical cellulase from *T. reesei* strain ATCC 26921 (Sigma-Aldrich, Dublin, Ireland) was spotted onto the plate (Fig 4.2.B and C)

Intron-less cellulase inserts were generated from *T. reesei* cDNA (Chapter 3, Section 3.2.7). The sequence analysis of the insert *egl1*cDNA in strains

S150GAL1egl1cDNA$^{T1360A::G1361}$ and C10-51GALeglcDNA$^{T1360A::G1361}$ indicated that missence mutations occurred at nucleotide positions 1360 and 1361 (Table 3.2) causing a frame shift in the last five amino acids in the CBM coding region. The frame shift knocked out the native stop codon of the gene and resulted in the translation of an additional 6 amino acids. This increased the size of EGI to 466 amino acids. The S150GAL1egl1cDNA$^{T1360A::G1361}$ and *S. pastorianus* C10-51GALegllcDNA$^{T1360A::G1361}$ strains were grown in 50 ml of YPGAL for 48 h. The culture medium was separated from the yeast cell pellet by centrifugation and 20 µl of the supernatant was spotted onto 1% (w/v) CMC plates (Fig 4.2 G and H respectively). The evidence of zones of hydrolysis indicated that functional EGI enzyme was present in the supernatants. This suggested that EGI enzyme was expressed and secreted by the yeast clones. S150GAL1 and C10-51GAL1 containing empty pGREG vector grown in 50 ml YPGAL possessed no innate endoglucanase activity (Fig 4.2 E and F, respectively). Thus mutations at the C-terminus of EGI and the translation of an additional 6 amino acids did not interfere with enzyme activity.

The expression of functional endoglucanase from *egl1* integrated C10-51GAL1egl1cDNA $^{T1360A::G1361integrated}$ was also examined. The CMC assay was perfomed on 20 µl of unconcentrated C10-51GAL1egl1cDNA $^{T1360A::G1361integrated}$ grown in 50 ml of YPGAL and indicated that EGI activity was detected (Fig 4.2.L). This indicated that *egl1* could be integrated and produce functional enzyme. As a positive control 5 mU of commerical *T. reesei* cellulase (Sigma-Aldrich, Dublin, Ireland) was spotted onto the plate (Fig 4.2.I). No activity was detected when either water or supernatants of C10-51GAL1 were spotted onto the CMC plates (Fig 4.2.J and K respectively).

The expression of EGI from the wildtype *egl1* gene was also examined. The strains S150PGK1egl1HIScDNA^{wt} (Table 2.4) and the control S150PGK1 were separately cultured in 2% (w/v) SC glucose at 30°C for 48 h. Supernatants were concentrated 50-fold (Chapter 2, Section 2.14). To determine if zone of hydrolysis reflected the amount of enzyme present, the zone of hydrolysis produced from unconcentrated and concentrated supernatant was compared. Twenty µl of each supernatant was spotted onto a CMC plate assay. No EGI activity was detected in either the unconcentrated or concentrated supernatant of S150PGK1 (Fig 4.2.M and O respectively).

However, a marked difference was observed in the diameter and intensity of the zones of hydrolysis between unconcentrated and 50-fold concentrated supernatant of S150PGK1egl1HIScDNA ^{wt} (Fig 4.2.N and P respectively). This indicated that the zone of hydrolysis reflected the amount of enzyme. It also indicated that the C-terminal HIS tag of strain S150PGK1egl1HIScDNA ^{wt} did not affect enzyme activity.

The concentrated supernatant from S150PGK1egl1HIScDNA ^{wt} was used to semi-quantify the amount of EGI present using commercial *T. reesei* cellulase as a standard. The diameter and intensity of the hydrolysis zone produced by 20 µl of 50-fold concentrated S150PGK1egl1HIScDNA ^{wt} supernatant (Fig 4.3.A) was compared to the zones produced by 30, 6, 3, 0.6 and 0.06 mU of *T. reesei* cellulase (Fig 4.3 B-F). A cellulase unit liberates 1 µmole of glucose from cellulose in 1h at pH 5 at 37°C (www.sigma.com). Twenty µl of 50 fold concentrated supernatant was shown to have a hydrolysis zone equivalent to 0.6 mU of commercial cellulase (Fig 4.3. A and E respectively). Therefore 20 µl of 50-fold concentrated supernatant had approximately 0.6 mU of endoglucanase activity and hence 30 mU ml^{-1} or 0.6 mU ml^{-1} unconcentrated. Thus 50 ml of S150PGK1egl1HIScDNA ^{wt} produced approximately 30mU of EGI.

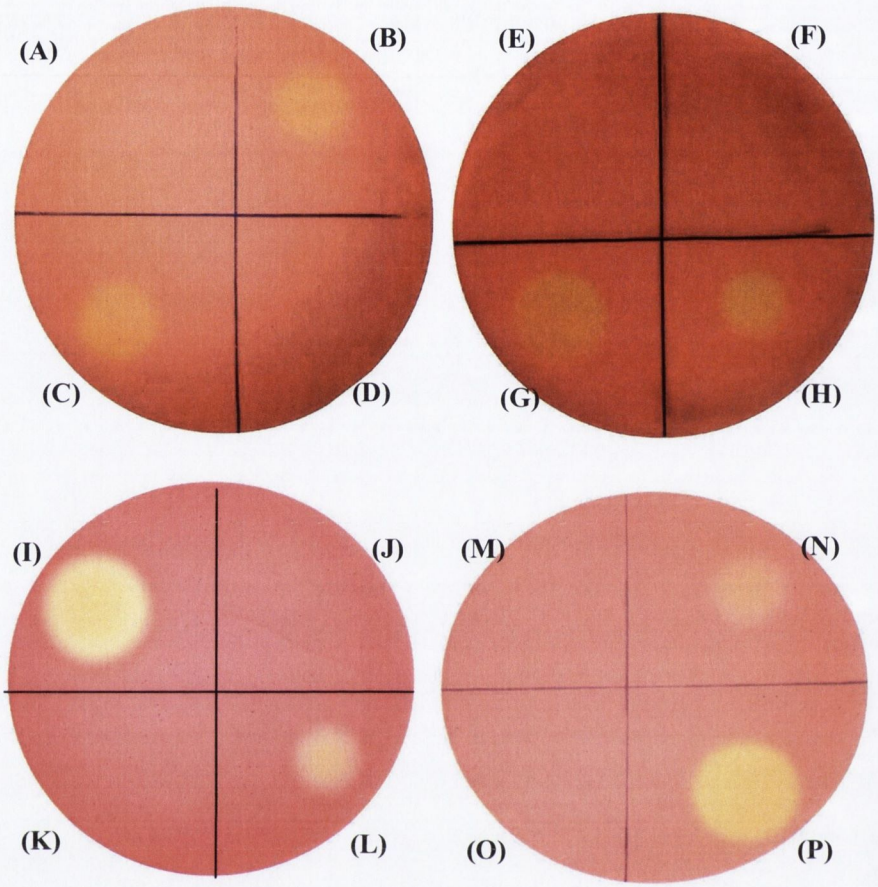


FIG. 4.2. Carboxymethylcellulose (CMC) agar plate assay for endoglucanase activity.

A 20 µl volume of unconcentrated supernatant from yeast growth cultures was tested in each sector unless otherwise stated. (A):S150GAL1egl1gDNA grown on SCgalactose (B):0.06 mU cellulase (Sigma). (C):0.06 mU cellulase. (D): S150PGK1egl1^{TACTAAC} grown on SCglucose. (E):S150GAL1 grown on YPGAL. (F):C10-51GAL1 grown on YPGAL. (G): S150GAL1egl1cDNA$^{T1360A::G1361}$ grown on YPGAL. (H): C10-51GALegl1cDNA$^{T1360A::G1361}$ grown on YPGAL. (I): 5 mU cellulase (Sigma). (J): H_20 control. (K): C10-51GAL1 grown in YPGAL. (L): C10-51GAL1egl1cDNA$^{T1360A::G1361integrated}$ in YEPGAL. (M): S150PGK1 grown on SC glucose. (N): S150PGK1egl1HIScDNA^{wt} grown on SC glucose. (O): 50-fold concentrated S150PGK1. (P): 50-fold concentrated S150PGK1egl1HIScDNA^{wt}.

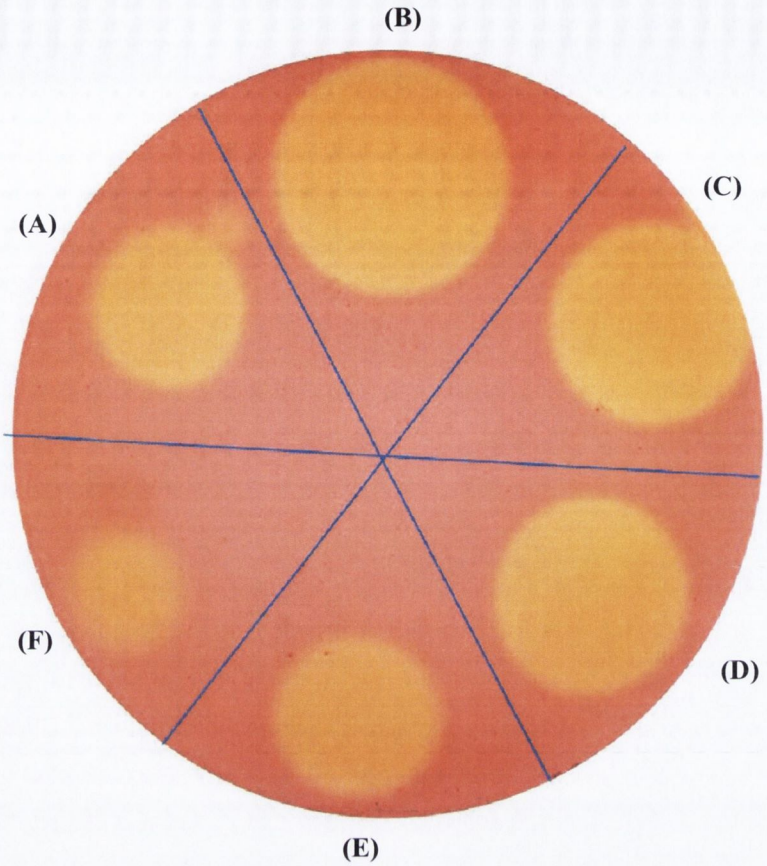


FIG. 4.3 Semi-quantification of EGI present in concentrated supernatant of S150PGK1egl1HIScDNA^{wt}

(A): 50-fold concentrated S150PGK1egl1HIScDNA^{wt} grown in SC glucose. (B): 30 mU cellulase (Sigma). (C): 6 mU cellulase. (D): 3 mU cellulase. (E): 0.6 mU cellulase. (F): 0.06 mU cellulase. A cellulase unit liberates 1 µmole of glucose from cellulose in 1h at pH 5 at 37°C (www.sigma.com).

Each sector had 20 µl of sample spotted.

4.2.2. Detection of C-terminal 6xHIS tagged EG1

To determine if the expressed RNA was effectively translated in the yeast cells, an anti-6xHIS tag antibody was used to monitor EGI protein levels

4.2.2.1 Western Immunoblot

To determine if EG1 protein could be detected in the supernatant of S150PGK1egl1HIScDNA^{wt} cells were grown in 50 ml of SC glucose. The supernatant was concentrated 50-fold as described in Chapter 2, Section 2.14. A Bradford assay was performed using 50 µl of the 50-fold concentrated supernatant. The assay revealed that the total protein concentrations present in the concentrated supernatant were too low to be detected (data not shown). A Western immunoblot was performed on the concentrated sample probing with a poly-HIS antibody as described in Chapter 2, Section 2.15. Each lane had 20 µl (equivalent of 1/50th of original supernatant) of concentrated supernatant sample. As a positive control the bacterial protein FIS containing a C-terminal HIS-tag was included. FIS is approximately 11.2 kDa. Concentrated supernatant of S150 containing empty PGK1 vector acted as a negative control. The *in silco* size of EGI is expected to be 55 kDa, however due to potential hyperglycosylation by *S. cerevisiae* the size could be as much as 80 kDa (Penttïla et al., 1987). A band corresponding to the 11 kDa FIS protein was evident (Fig 4.4.A, lane 3), however no bands were evident in samples of the HIS tagged EGI supernatants nor with the empty vector (Fig 4.4.A, lanes 1 and 2 respectively). This suggested that the level of HIS-tagged EGI present in the 50-fold concentrated supernatant was not sufficent to be detected by the antibody.

4.2.2.2 Enzyme linked immunoadsorbent assay

The presence of EGI in the supernatant was also tested by ELISA. ELISA was performed, using the same samples as the Western blot, as described in Section 2.16. The results are presented on Figure 4.4.B. The results indicated that the HIS tagged FIS (turquoise bar) was clearly detected. The concentrated HIS tagged EGI (pink bar) and the empty vector negative control (black bar) were indistinguishable and not statistically significant. This implied that the level of HIS tagged EGI was below detection. The pattern for the ELISA mirrored the results of the Western immunoblot (Figure 4.4.B).

FIG. 4.4. Detection of HIS tagged EGI protein.

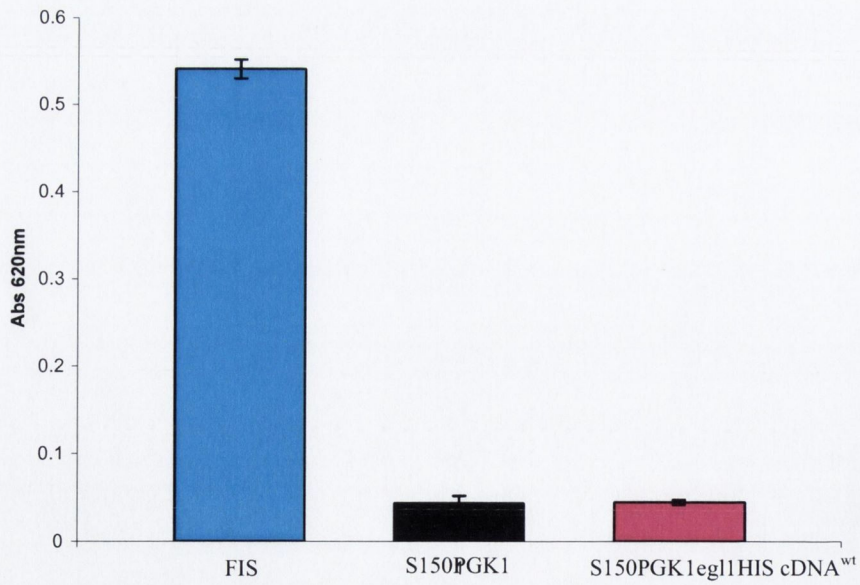
(A) Western immunoblot probed with anti-HIS antibody (1 in 3000). Lane 1: 50-fold concentrated supernatant (equivalent of $1/50^{th}$ of original sample) of S150PGK1egl1HIScDNA^{wt}. Lane 2: 50 fold concentrated supernatant of S150PGK1. Lane 3: 1 µg FIS protein. Blot shows a10 minute exposure. An overnight exposure did not reveal any subsequent bands in lane 1 (Data not shown)

(B) ELISA performed on 50-fold concentrated supernatant. A 1µg quantity of HIS-tagged FIS protein (Turquoise column). S150PGK1 (black column). S150PGK1 egl1HIScDNA^{wt}(Pink column). Absorbance read at 620 nm. Samples were measured in triplicates. Error bars represent standard deviation.

(A)


1
2
3

(B)


0.6
0.5
0.4
0.3
0.2
0.1
0
Abs 620nm
FIS
S150PGK1
S150PGK1egl1HIS cDNA^wt

4.2.3. Detection of CBHII activity using phosphric acid swollen cellulose (PASC)

4.2.3.1. S150 *cbh2* clones

Since no activity was detected in *S. cerevisiae* strains containing genomic copies of the *egl1* gene (Section 4.2.1), it was decided to focus on examing *cbh2* cDNA clones for enzyme activity. CBH enzymes have been observed to have an activity lower than the other cellulase enzymes (Den Haan et al., 2007a). The method used to determine CBHII enzyme activity in the *cbh2* containing clones was developed by Den Haan et al., (2007a). Enzyme activity was investigated in the *cbh2* containing clones by measuring increases in total sugar following adsorption of cellulases present in the supernatants onto phosphoric acid swollen cellulose as described in Section 2.20. Briefly, the supernatant of a *cbh2* containing yeast strain was mixed with PASC to allow the enzyme to bind at 4°C. The PASC with bound enzyme was then incubated at 50°C to allow the CBHII to degrade the cellulose. This releases mainly cellobiose into the reaction mixture and will increase the total level of sugar. Degradation of crystalline cellulose by CBH can also release small quantities of other cellooligosaccharides such as cellotriose. If *cbh2* is not present or active, the sugar content will remain unchanged. The phenol sulfuric acid assay was used to detect total sugar concentration as described in Section 2.21. First, a standard curve was performed to determine the range of sugar concentration that can be detected by the phenol sulfuric acid assay. The absorbance values were plotted against concentrations of cellobiose ranging between 10µg to 200µg (Fig 4.5). The results indicated that the phenol sulfuric acid assay was linear over the range tested.

The yeast strains S150PGK1, S150PGK1cbh2cDNA$^{G1076C,\ G1345C}$, S150PGK1cbh2cDNA ^{A896C} and S150PGK1cbh2HIScDNA ^{wt} were separately cultured in 150 ml of YPD media at 30°C for 72h. The assay was performed on the culture supernatants as described in Chapter 2, Section 2.20. the yeast culture was centrifuged and the supernatant was retained. The assay was performed at 50°C for 72 h to allow the CBHII to hydrolyse the PASC. An aliquot of 250 µl of supernatant of this reaction was then analysed for total sugar concentration using the phenol sulfuric acid assay (Chapter 2, Section 2.21). The results for each strain are shown in Fig 4.6.

Nucleotide sequence analysis of the *cbh2* cDNA insert present in S150PGK1cbh2cDNA^{G1076C,G1345C} was shown to contain two missense mutations at nucleotide positions 1076 and 1345 and a silent mutation at 561 (See Chapter 3, Fig 3.12.B and Table 3.2). These errors occurred within the catalytic domain of the CBHII protein (Fig 4.1C). As described in Chapter 3, Section 3.2.9.2, site-directed mutagenesis PCR was used to correct the missense errors in this strain (See Chapter 3, Fig 3.11.B). The nucleotide sequence analysis of the resultant strain S150PGK1cbh2cDNA^{A896C} indicated that the errors within the *cbh2* insert at 1076 and 1345 were correct; however, a missense mutation at nucleotide position 896 was detected (Fig 4.1D). This new mutation occurs within the catalytic domain of CBHII.

The total sugar released from incubation of the *cbh2* wildtype strain, S150PGK1cbh2HIScDNA^{wt}, was compared to that produced from two mutants strains S150PGK1cbh2cDNA$^{G1076C,\ G1345C}$ and S150PGK1cbh2cDNA^{A896C} (Fig 4.6). The total sugar released following incubation of S150PGK1cbh2HIS cDNA^{wt} with PASC was approximately twice that observed with S150PGK1cbh2cDNA^{G1076C,G1345C}, S150PGK1cbh2cDNA^{A896C} or the control S150PGK1 (Fig 4.6). This suggested that the wt copy of CBHII was functional but no CBHII activity was produced by the mutant strains. Taken together, these findings suggested that any missense mutation within the catalytic domain was sufficient to inactivate enzyme functionality. Additionally the presence of the 6XHIS-tag at the C-terminus of CBHII did not inactivate the protein.

Statistical analysis was performed using Mini-tab statistical software program. A Student-T test was performed to compare the data from S150PGK1cbh2HIS cDNA^{wt} with the control yeast S150PGK1. The threshold for statistical significance was set at the standard p=0.05. Any p values less than this were deemed significant. A T-test comparing the two data sets generated p=0.006. This indicated that the data for S150PGK1cbh2HIScDNA^{wt} was statistically significant. This suggested that S150PGK1cbh2HIScDNA^{wt} had produced functional CBHII enzyme. The levels of total sugar produced by strains S150PGK1cbh2cDNA^{G1076C,G1345C}, S150PGK1cbh2cDNA^{A896C} and S150PGK1 were not statistically significant.

4.2.3.2. Activity in C10-51 *cbh2* clone

The CBHII activity in the supernatant of C10-51PGK1cbh2HIScDNA^{wt} was monitored over a time course of 72 h at 50°C. This strain had the same insert that described for S150PGK1cbh2HIScDNA^{wt} (Chapter 3, Table 3.2). The strain C10-51PGK1 was used as a negative control. The results of the phenol sulphuric acid assay were plotted (Fig 4.7.A). The total sugar released by reaction of C10-51PGK1cbh2HIScDNA^{wt} with PASC (blue line) increased over time, whereas the negative control C10-51PGK1 (red line) remained relatively constant. Statistical analysis indicated that the difference between the strains was statistically significant. At the 72h time point, $p=2x10^{-6}$. The increase in sugar content suggested that active CBHII was present in the supernatant of C10-51PGK1cbh2HIScDNA^{wt}. The enzyme activity reached a plateau after 48 h at 50°C.

To determine if all of the CBHII enzyme present in the supernatant sample was adsorbed to PASC substrate during an overnight incubation at 4°C, the post-adsorption supernatants from the above strains C10-51 PGK1 and C10-51PGK1cbh2HIScDNA^{wt} were reacted with fresh PASC at 50°C for 72 h. The release of sugars from the adsorbed PASC was determined as described above. As a positive control 10mU of commercial *T. reesei* cellulase was also incubated with PASC. It was observed that no detectable enzyme activity was present in the supernatants of C10-51PGK1 and C10-51PGK1cbh2HIScDNA ^{wt} after adsorption (Fig 4.7.B, red column and blue column, respectively). The positive control displayed an increase in total sugar (Fig 4.7.B, brown column). The conclusion from this experiment was that overnight adsorption at 4°C was sufficient to bind all CBHII within C10-51PGK1cbh2HIScDNA^{wt} supernatant to the PASC.

The levels of total sugars released from the S150PGK1cbh2HIScDNA^{wt} and C10-51PGK1cbh2cDNA^{wt} appear to be significantly different. However the assays were carried out independently and with different batches of media; therefore the values should not be compared.

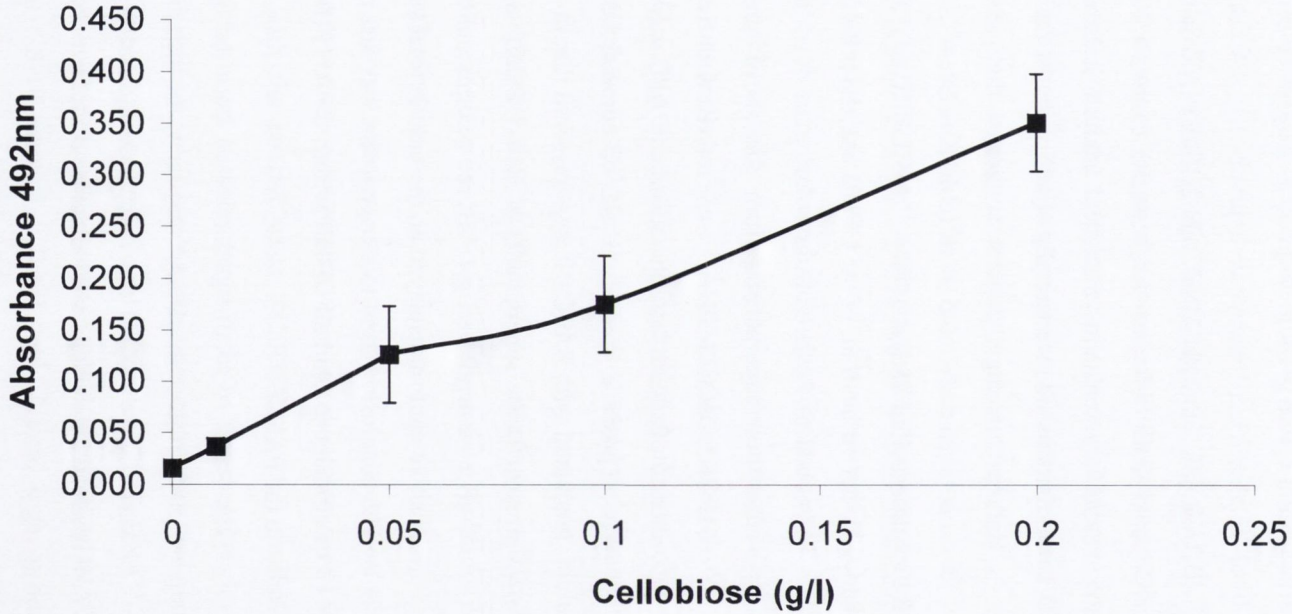


FIG. 4.5. Standard curve for cellobiose as detected by the phenol sulphuric acid assay. Absorbance values read at 492nm. Samples read in triplicate. Errors bars represent standard deviation.

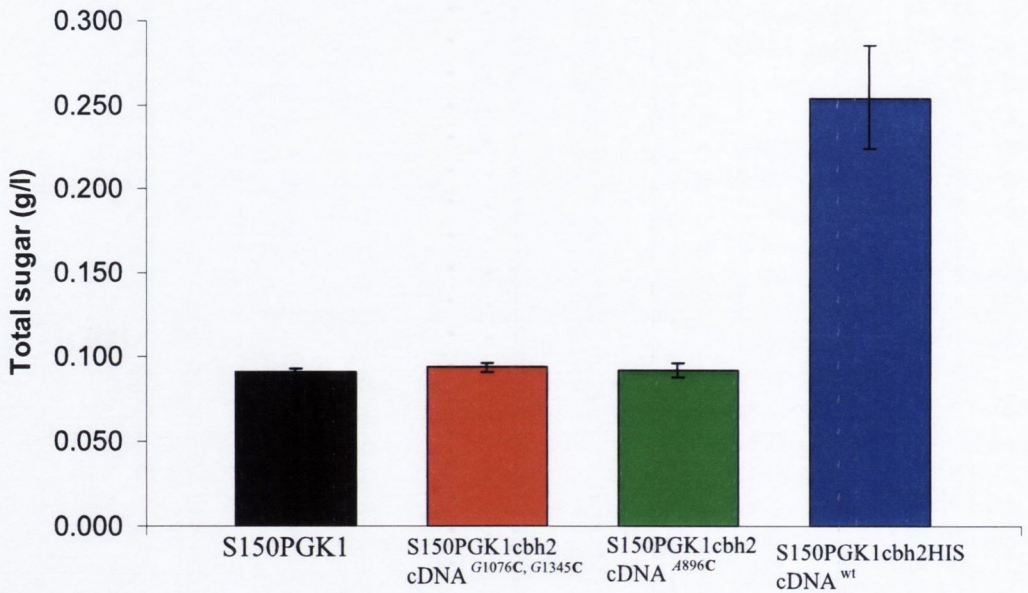
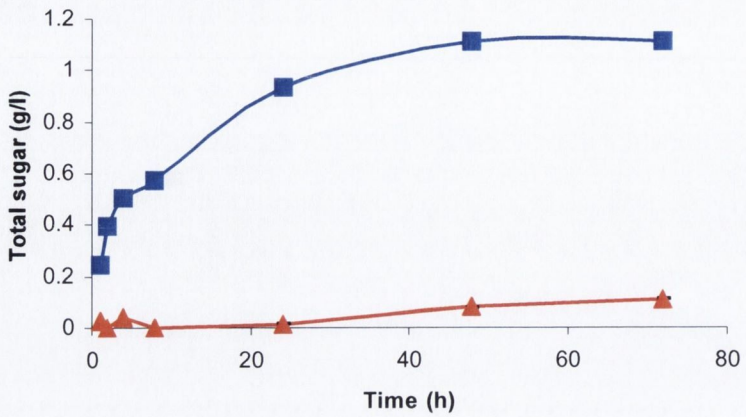


FIG. 4.6. Phenol sulphuric acid assay investigating CBHII activity in S150 clones. Assay used to detect total sugar present per ml of reaction mixtures. S150PGK1 (black column). S150PGK1cbh2 cDNA $^{G1076C,\ G1345C}$ (red column). S150PGK1cbh2cDNA ^{A896C} (green column). S150PGK1cbh2HIScDNA ^{wt} (blue column). Samples tested in triplicate. Error bars represent standard error.

(A)



(B)

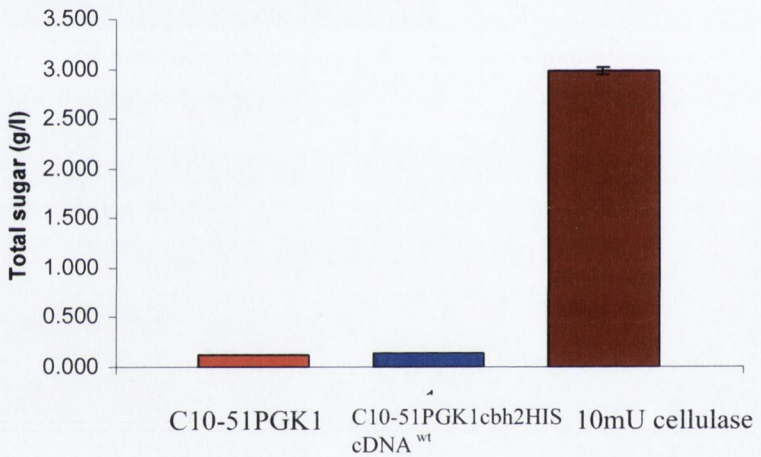


FIG.4.7. Phenol sulphuric acid assay used to detect CBHII activity in C10-51PGK1cbh2HIScDNA ^{wt} strain by measuring total sugar per ml of assay sample.

(A) CBHII activity after hydrolysis of PASC at 50°C. C10-51PGK1 (red line). C10-51PGK1cbh2HIScDNA ^{wt} (blue line). Enzyme activity was monitored over 72 h. Samples measured in triplicate. Error bars are present for each sample but may be too small to observe. Error bars represent standard deviation.

(B) CBHII activity in the post-adsorption supernatant. C10-51PGK1 (red column). C10-51PGK1cbh2HIS cDNA ^{wt} (blue column). 10mU cellulase (brown column) acted as positive control. Bar chart based on readings from the 72h time point. Samples measured in triplicate. Error bars represent standard deviation.

4.2.4. Detection of BGL1 activity using X-glucoside agar

Based on the observation (Section 4.2.1) that no activity was detected in *S. cerevisiae* strains containing genomic cellulase gene, it was decided to focus on examining *bgl1* cDNA clones for enzyme activity. As described for EGI and CBHII (Section 4.2.1 and 4.2.3 respectively) the growth culture supernatant of yeast transformed with the *bgl1* gene was examined for enzyme activity. Detection of the BGLI enzyme in the supernatant would confirm that the native *T. reesei* secretory sequence was recognised and correctly processed by the host *Saccharomyces* secretory mechanism. However, a study by van Rooyen et al., (2005) found that *T. reesei bgl1* expressed in recombinant yeast was not observed extracellularly but was detected within the yeast cell (See Chapter 1, Section 1.12). Therefore it was decided to also examine the cell lysates of yeast transformed with *bgl1* cDNA inserts. The substrate used to detect BGLI enzyme activity was the chromogenic reagent X-glucoside (X-glu; 5-bromo-4-chloro-3-indolylglucopyranoside). BGLI hydrolyses the latter reagent to release 5-bromo-4-chloro-3-indol which generates a blue product. The embedding of X-glu within an agar plate allows for a rapid assay to detect BGLI activity (Bhatia et al., 2002).

Yeast strains transformed with *bgl1* were grown at 30°C for 48h in 50ml of SC glucose media (Chapter 2, Section 2.1). The supernatants were concentrated 50-fold as described in Chapter 2, Section 2.14. The X-glu plate assay was performed as described in Chapter 2, Section 2.18.1 with 20µl (which represented $1/50^{th}$ of total) of either concentrated supernatant or cell lysate.

A positive control of 10 mU of commercial cellulase (Sigma-Aldrich, Dublin, Ireland) produced the expected zone of hydrolysis (Fig 4.8.A). Nucleotide sequence analysis of *bgl1* had detected a number of missense mutations in strain S150PGK1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$ (Chapter 3, Fig 3.12.E, Table 3.2). The errors at nucleotide positions 64-65 occurred within the *bgl1* secretory sequence coded for amino acid residue 22 (Fig 4.1E). Neither the 50-fold concentrated supernatant nor the cell lysate from cells expressing this mutant gene contained BGLI activity (Fig 4.8, B and C respectively). As described in Chapter 3, Section 3.2.9.3, the missense errors of the insert bgl1cDNA$^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$ were corrected by site directed mutagenesis PCR. This generated strain S150PGK1bgl1cDNA $^{C64A, A65G, \Delta 1521\text{-}2098}$. Nucleotide sequence analysis indicated that the errors

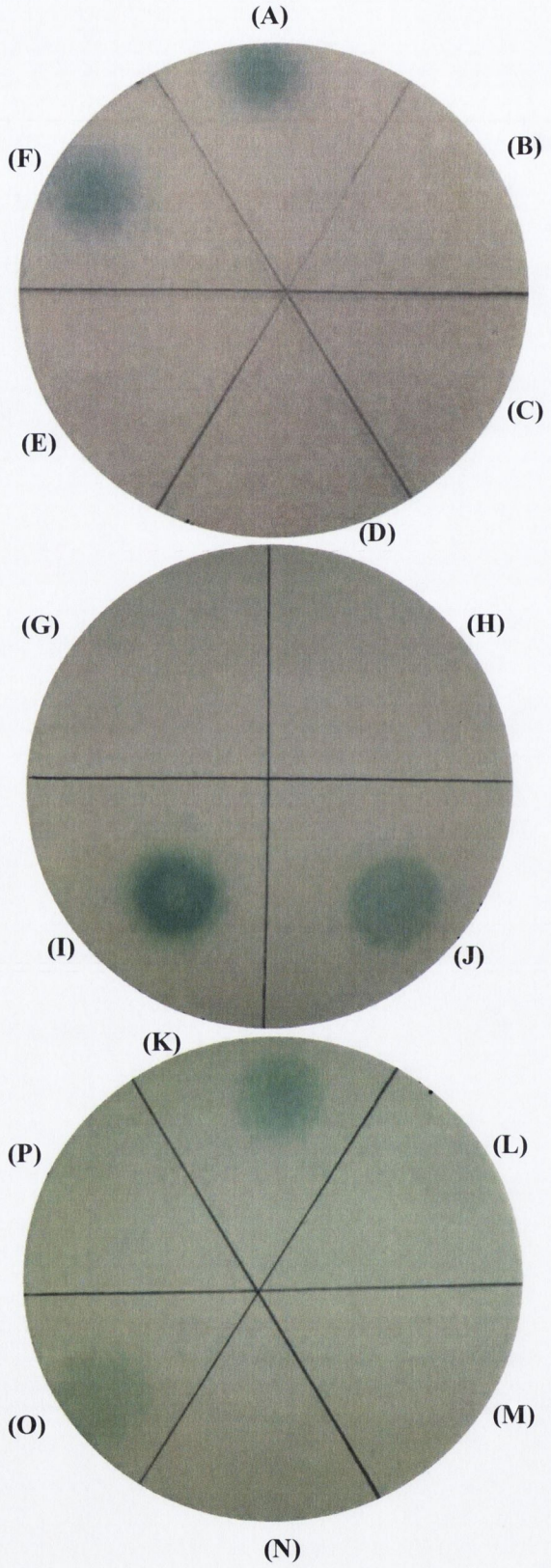
within the secretory signal remained but the mutations at nucleotide positions 1061, 1085, 2099 and 2212 were corrected. Significantly a deletion of 578bp was observed between positions 1521 to 2098 which coded for amino acid residues 507-699 (Fig 4.1F). The 50-fold concentrated supernatant of cells expressing this gene did not contain BGLI activity (Fig 4.8.E). Activity was however observed in the cell lysate (Fig 4.8.F). Taken together, these results indicated that the repairs of mutations in *bgl1* at nucleotide positions 1061, 1085, 2099 and 2212 were required to restore enzyme activity. Interestingly the deletion of a 578bp region of coding region did inactivate the BGLI produced. The observation of Fig 4.8.E suggested that the errors within the *bgl1* secretory sequence at positions 64-65 were sufficient to prevent secretion of recombinant BGLI enzyme. These mutations altered a histidine to a serine (Chapter 3, Table 3.2). A negative control of 0.05M citrate buffer used for lysing cells did not contain any enzyme activity (Fig 4.8.D).

A second attempt to correct the errors of bgl1cDNA$^{C64A,\ A65G,\ T1061C,\ T1085C,\ T2099C,\ A2212G}$ insert eventually resulted in S150 PGK1bgl1cDNA^{wt} (Chapter 3, Section 3.2.10.3). Nucleotide sequencing of the *bgl1* insert within this clone revealed that all missense mutations had been corrected (Chapter 3, Table 3.2). An X-glu assay performed with 50-fold concentrated supernatant and cell lysate of strain S150PGK1bgl1cDNA^{wt} revealed that BGLI activity was detected in both (Fig 4.8, I and J respectively). Interestingly, the BGLI present in 20µl of unconcentrated supernatant was not detectable on the X-glu plate (Fig 4.8.G). This suggested that the plate assay may not be sensitive enough to detect low amounts of BGLI enzyme. A negative control of SC glucose media alone did not produce any activity (Fig 4.8.H). No innate BGLI activity was detected in the 50-fold concentrated supernatant, cell lysate of S150PGK1 or supernatant of C10-51PGK1 (Fig 4.8, M, N and P respectively). The 50-fold concentrated supernatant of C10-51PGK1bgl1cDNA^{wt} was observed to have BGLI activity (Fig 4.8.O). This demonstrated that the secretory mechanisms of *S. pastorianus* could recognise and correctly secreted cellulase enzyme. Concentrated supernatant of S150PGK1bgl1cDNA^{wt} was included as a positive control (Fig 4.8K). Sector L was left blank (Fig 4.8.L).

FIG.4.8. X- glucoside plate assay to detect BGL1 activity in yeast clones.

All strains were grown in SC glucose. Growth culture supernatants were concentrated 50-fold with the exception of G. Cell lysates were prepared by vortexing pellets in 0.05M citrate buffer with zirconia beads. A 20µl volume of each sample was spotted onto the X-glu plate which was divided into sectors.

(A): 10mU of cellulase (Sigma). (B): Supernatant of S150PGK1bgl1cDNA $^{C64A,\ A65G,\ T1061C,\ T1085C,\ T2099C,\ A2212G}$. (C): Lysate of S150PGK1bgl1cDNA $^{C64A,\ A65G,\ T1061C,\ T1085C,\ T2099C,\ A2212G}$. (D): 0.05M citrate buffer. (E): Supernatant of S150PGK1bgl1cDNA$^{C64A,\ A65G,\ \Delta1521\text{-}2098}$. (F): lys S150PGK1bgl1cDNA $^{C64A,\ A65G,\ \Delta1521\text{-}2098}$. (G): unconcentrated supernatant S150 PGK1bgl1cDNA^{wt}. (H): SC glucose media. (I): Supernatant of S150 PGK1bgl1cDNA^{wt}. (J): Lysate of S150PGK1bgl1cDNA^{wt}. (K): Supernatant S150 PGK1bgl1cDNA^{wt}. (L): blank. (M): Lysate of S150PGK1. (N): Supernatant S150PGK1. (O): Supernatant C10-51PGK1bgl1cDNA^{wt} (P): Supernatant C10-51PGK1.


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
(M)
(N)
(O)
(P)

4.2.5. Detection of BGLI using a liquid X-glu assay

The enzymatic profile of the cloned BGLI was determined over a period of 24 h using X-glucoside as a substrate. The liquid X-glucoside assay was performed as described in Chapter 2, Section 2.18.2.

Yeast strains S150PGK1 and S150PGK1bgl1cDNA ^{wt} were grown in 40 ml of SC glucose for 48 h at 30°C. The supernatants were concentrated 40-fold as described in Chapter 2, Section 2.14. The BGLI activity in unconcentrated or a 40-fold concentrated supernatants of S150PGK1 and S150PGK1bgl1cDNA ^{wt} was tested. Two different concentrations of commercial *T. reesei* cellulase (10 mU and 50 mU) were used as positive controls. The values of absorbance at 620nm were plotted against time (Fig 4.9). The 40-fold concentrated supernatant from S150PGK1bgl1cDNA^{wt} (orange line) produced BGLI activity greater than 10 mU (green line) but less than 50 mU of commercial *T. reesei* cellulase (turquoise line). Based on this pattern, the amount of BGLI in 100 µl of 40-fold supernatant was approximately 40 mU. Therefore 40 ml of S150PGK1bgl1cDNA^{wt} secreted approximately 400 mU which is the equivalent of 10 mU ml^{-1} of unconcentrated supernatant. Statistical analysis was performed to compare the data set from the reaction of concentrated supernatant S150PGK1 (purple line) and S150PGK1bgl1cDNA^{wt}. Taking the data from the 24h time point a p value=0.027 was calculated. This indicated that the results for S150PGK1bgl1cDNA^{wt} were statistically significant compared to S150PGK1. Figure 4.9 also indicates that the unconcentrated supernatant of S150PGK1bgl1cDNA^{wt} (yellow line) contained detectable levels of BGLI activity and was statistically different from unconcentrated S150PGK1 (black line), p=0.016. BGLI was most active in the first 4 hrs of the assay at 50°C. Both 50 mU cellulase and S150PGK1bgl1cDNA^{wt} yielded a sharp rise in absorbance values up to the 5 h time point (Fig 4.9). Thereafter, the enzyme activity reached a plateau.

4.2.6 Growth of BGL1 producing clone in cellobiose

To investigate if S150PGK1bgl1cDNA^{wt} could metabolise and grow using cellobiose as the sole carbohydrate source, yeast strains S150PGK1and S150PGK1bgl1cDNA^{wt} were separately pre-cultured in 30 ml of SC glucose media. The culture supernatants were concentrated 40-fold to remove any unmetabolised glucose. Four cultures of equal numbers of yeast cells were set

up in 20 ml of SC cellobiose, two cultures were supplemented with 200µl of concentrated supernatant, the equivalent of 8 ml of unconcentrated supernatant. The other two cultures were not supplemented with supernatant. The yeast was incubated at 30°C and the optical density of the yeast was monitored over 96 h. The growth curve was plotted as optical density (OD_{600nm}) against time (See Fig 4.10). The results indicated that the culture of S150PGK1bgl1cDNA^{wt} supplemented with the concentrated supernatant (orange line) outperformed the other yeast cultures. Cultures of S150PGK1bgl1cDNA^{wt} that were not supplemented with concentrated supernatants did not grow, nor did cultures of S150PGK with or without supplementation. The growth patterns for S150PGK1 and S150PGK1bgl1cDNA^{wt} were indistinguishable (Fig 4.10, black line and yellow line). These findings indicated that the addition of BGLI was required to initiate the metabolism of cellobiose by yeast strain S150PGK1bgl1cDNA^{wt}. The supernatant of S150PGK1 did not possess any BGL1 activity and therefore S150PGK1 could not grow in the cellobiose media.

4.2.7. Growth curve of recombinant yeast strains

To determine if the presence of plasmids expressing cellulase genes would interfere with cell growth, the strains S150PGK1, S150PGK1egl1HIScDNA^{wt}, S150PGK1cbh2HIScDNA^{wt}, S150PGK1bgl1cDNA^{wt}, C10-51PGK1, C10-51PGK1egl1HIScDNA^{wt}, C10-51PGK1cbh2HIScDNA^{wt} and C10-51PGK1bgl1cDNA ^{wt} were cultured in 50 ml of YPD media containing 0.2 mg l^{-1} of G418 for 48 h. The OD_{600} of the cultures was sampled at designated time points. The results of the absorbance values plotted against time for the S150 cultures and the C10-51 cultures are shown in Figures 4.11A and 4.11B respectively. All strains grew exponential during the first 24h followed by a gradually attenuation of the growth rate between 24 and 48h. This indicated that the cultures were entering stationary phase. All of the strains displayed a similar growth curve with no strain dominating. The previous experiments from this chapter indicated that S150PGK1egl1HIScDNA^{wt}, S150PGK1cbh2HIScDNA^{wt} and S150PGK1bgl1cDNA^{wt} produced and secreted active cellulolytic enzymes (Fig 4.11.A, pink line, blue line and yellow line respectively. These strains had the same growth pattern as S150PGK1 (Fig 4.11.A black line). This indicated that expression of cellulase

was not energetically detrimental to the host S150 organism and no loss of fitness occurred.

C10-51PGK1, C10-51PGK1egl1HIS cDNA^{wt}, C10-51PGK1cbh2HIScDNA^{wt} and C10-51PGK1bgl1cDNA^{wt} were observed to produce a similar growth curve (Fig 4.11.B black line, pink line, blue line and yellow line). This indicated that the expression of active cellulase enzymes did not affect growth rates. The observed growth curve in Fig 4.11.B suggested that the C10-51 cultures had not reached stationary phase by the 48h time point. All of the strains appeared to be in exponential growth at 48 h. Comparing Figures 4.11.A to 4.11.B, it was observed that the S150 strains had a faster growth rate than C10-51 in YPD. The average OD for the S150 strains at the 48h point was approximately 7.3, while C10-51 strains only reached a max OD of 4.8. The differential in growth rates may be due to the genomic makeup of the two yeast species; S150 is a haploid whereas C10-51 is polyploid. Haploids cells may divide faster than polyploid because they have less genetic content to copy. Alternatively the growth media may be an important factor, *S. pastorianus* strains have been shown preferred uptake of maltose over glucose (See Chapter 5, Section 5.2.1).

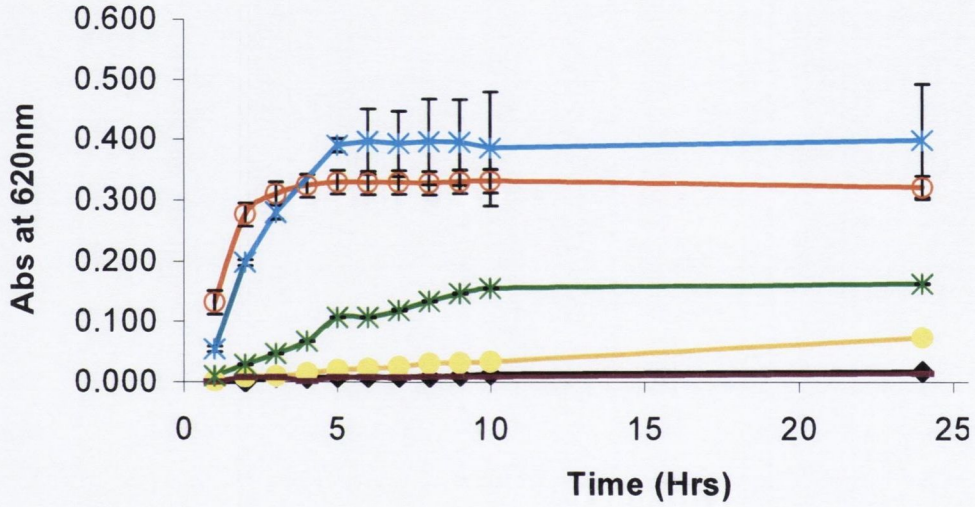


FIG 4.9. Measurement of BGL1 activity using liquid X-glu assay.

Unconcentrated supernatant of S150PGK grown in SC glucose (black line). Unconcentrated supernatant of S150PGK1bgl1cDNA^{wt} (Yellow line). 40-fold concentrated supernatant S150PGK (purple line). 40-fold concentrated supernatant S150PGK1bgl1cDNA^{wt} (orange line). 10 mU cellulase (Green line). 50 mU cellulase (Turquoise line). Absorbance was read at 620 nm. Samples were measured in triplicates. Error bars represent standard deviation. Note the error bars are present but are too small to observe for the unconcentrated supernatant S150PGK, unconcentrated supernatant of S150PGK1bgl1cDNA^{wt}, 40-fold concentrated supernatant S150PGK and 10 mU cellulase.

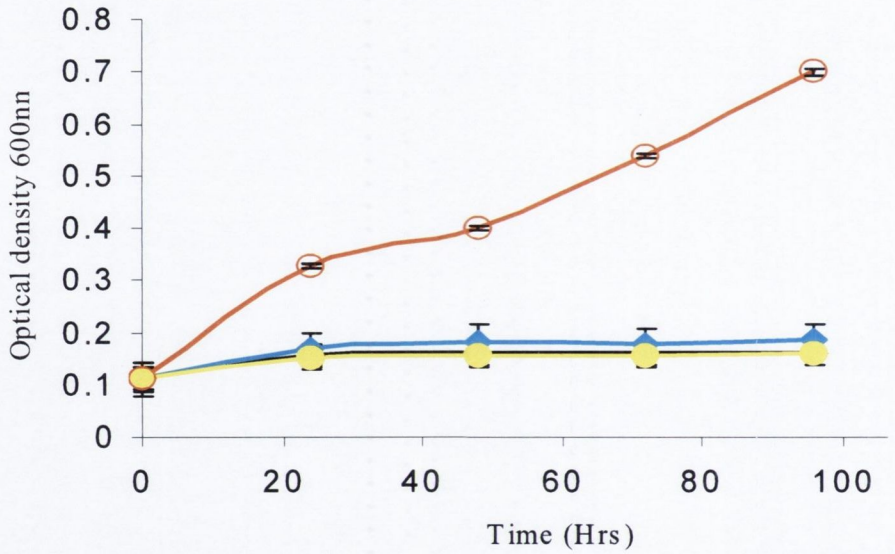


FIG. 4.10. Growth of S150PGK1bgl1cDNA^{wt} in cellobiose media.

S150PGK1 (black line). S150PGK1+200 µl of 40-fold concentrated pre-culture supernatant of same strain (turquoise line). S150PGK1bgl1cDNA^{wt} (Yellow line). S150PGK1bgl1cDNA^{wt} +200 µl of 40-fold concentrated pre-culture supernatant of the same strain (orange). Optical density values were read at 600nm over 96 h. A volume of 200 µl of 40-fold concentrated supernatants is equivalent to 8 ml of unconcentrated supernatants. Samples were measured in duplicates. Error bars represent standard deviation.

FIG 4.11.Growth curves of recombinant yeast versus parental strain

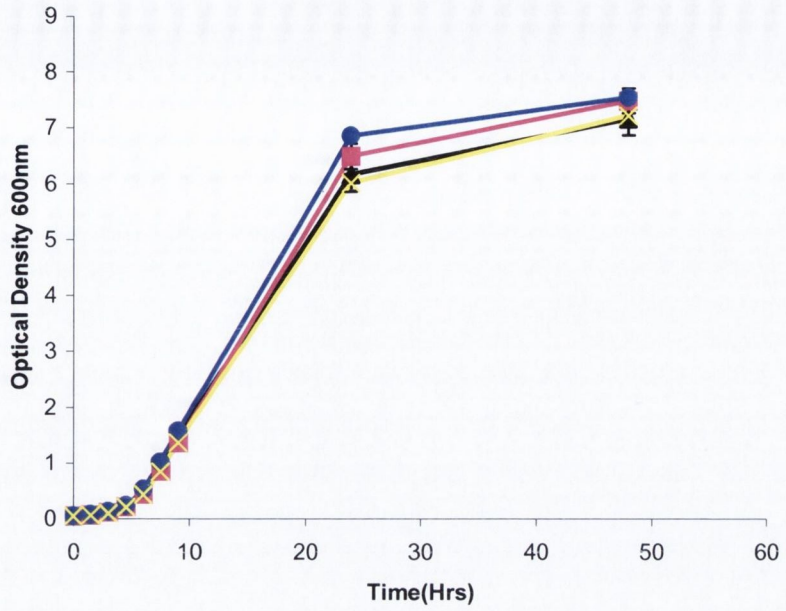
(A). Growth curve of S150 recombinant strains.

The strains S150 PGK1 (black line), S150 PGK1egl1HIS cDNA ^{wt} (pink line), S150 PGK1cbh2HIScDNA ^{wt} (blue line) and S150 PGK1bgl1cDNA ^{wt} (yellow line) were grown at 30°C for 48 h. The OD600 nm of each culture was sampled in triplicate at designated time points. Error bars are present for each sample but may be too small to observe. The error bars represent standard deviation.

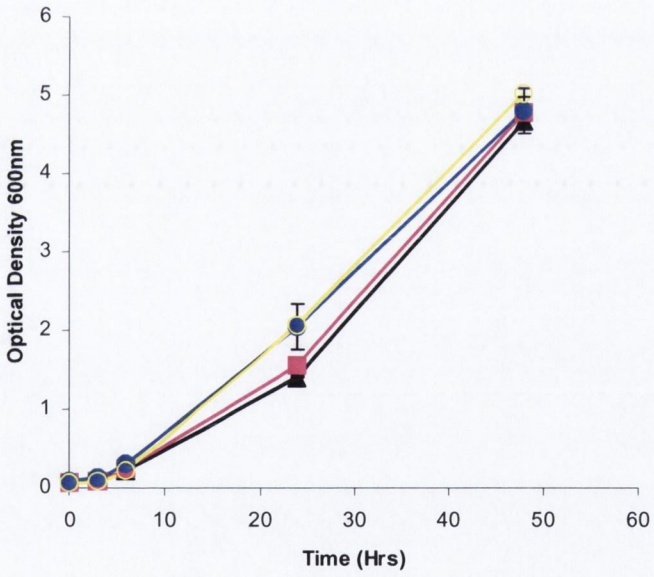
(B). Growth curve of C10-51 recombinant strains.

The strains C10-51PGK1 (black line), C10-51PGK1egl1HIS cDNA ^{wt} (pink line), C10-51PGK1cbh2HIScDNA ^{wt} (blue line) and C10-51PGK1bgl1cDNA ^{wt} (yellow line) were grown at 30°C for 48 h. The OD600 nm of each culture was sampled at designated time points.

(A)


Optical Density 600nm
0
1
2
3
4
5
6
7
8
9
0
10
20
30
40
50
60
Time(Hrs)

(B)


Optical Density 600nm
0
1
2
3
4
5
6
0
10
20
30
40
50
60
Time (Hrs)

4.3 Discussion

The results of the experiments described in this chapter indicated that functional cellulase enzyme was produced by several yeast clones as demonstrated by CMC, X-glu and total sugar content assays (Figures 4.2, 4.6, 4.7 and 4.8). The list of cellulase secreting yeast strains is summarised in Table 4.1. The genomic DNA yeast clones were found not to produce active enzyme. The genomic *egl1* clone S150GAL1egl1gDNA did not hydrolyse CMC (Fig 4.2.A). The lack of EGI activity may have been a result of the continued presence of its introns. Nucleotide sequencing of the *egl1* insert within this clone indicated that no sequence errors occurred (Chapter 3, Table 3.1). This suggested that failure to produce EGI was due to the inability of the *S. cerevisiae* splicesome to recognise and remove the foreign introns within the insert. This finding had previously been reported (Cummings and Fowler, 1996). A study by Langford et al., (1983) inserted an intron containing actin gene from *Acanthamoeba* into *S. cerevisiae* and found that the host splicing machinery did not excise the foreign introns. The only highly conserved sequences in introns are those that are required for removal, the 5'donor site marked by a guanine and thymine and the acceptor site featuring an adenine and guanine (Alberts et al., 1994). These sequences are common to all introns found in fungal species. Studies have indicated that introns within *S. cerevisiae* have a characteristic TACTAAC branch site sequence upstream of the acceptor site that is unique to the organism and is believed to be involved in splicing (Thomas et al., 1999). To investigate if the presence of this motif would enable *S. cerevisiae* to splice foreign introns an *egl1* clone was generated that had mutated introns containing TACTAAC; S150PGK1egl1^{TACTAAC}. Nucleotide sequencing of the clone confirmed that sequence contained no missense mutations. The result of the CMC assay of S150PGK1egl1^{TACTAAC} indicated that no enzyme activity was detected (Fig. 4.2.D). This finding suggested that the host splicing machinery did not remove the introns within *egl1*.

The egl1cDNA clone S150GAL1egl1cDNA$^{T1360A::G1361}$ demonstrated EGI activity (Fig 4.2.G). Nucleotide sequencing of the *egl1* insert within S150GAL1egl1cDNA$^{T1360A::G1361}$ contained mutations at the 3' end of the insert. This mutation was caused by a T to A mutation and insertion of a G resulted in the alteration of a tyrosine to an arginine at amino acid residue 454.

An inserted guanine at nucleotide position 1361 caused a frame shift in the last 6 amino acid residues of the coding sequences. The coding changed from the wildtype *T. reesei* YYSQCL to the mutant LLAMPL in the clone. The frameshift also knocked out the native stop codon, this resulted in the translation of an additional 7 amino acids. This shift and translation of additional amino acids did not result in the inactivity of the enzyme (Fig 4.2,G and H). This may be due to the location of the mutations within the *egl1* sequence. The frameshift occurs in the carbohydrate binding module region. The CBM is required for binding to crystalline cellulose and is not directly involved in enzyme activity. EGI activity was demonstrated using CMC which is a soluble form of cellulose and therefore does not require the binding abilities of the CBM. The data indicated that the final 6 amino acids of EGI are not required for enzyme activity as measured by CMC hydrolysis.

A study by du Plessis et al., (2010) generated a *S. cerevisiae* strain Y294[pLEM1] that contained a randomly mutated *T. reesei egl1* insert. The mutation caused an A to G nucleotide transition at position 1308bp, resulting in a tyrosine to cysteine change at amino acid position 436, which yielded a 1.4 fold increase in EGI activity. The mutation occurred in the CBM domain of the enzyme. This is similar to the mutation detected in the egl1cDNA$^{T1360\mathbf{A}::G1361}$ generated in this study (See Chapter 3, Table 3.2). The binding of cellulose involves the hydrophobic interaction of the sugar rings of cellulose with the aromatic side chains of tyrosines, phenylalainine or tryptophans on a planar face of the CBM (du Plessis et al., 2010). Thus mutation of the tyrosine at amino acid residue 454 of egl1cDNA$^{T1360\mathbf{A}::G1361}$ may have altered the conformation or binding ability of the CBM. However, the EGI enzyme activity in S150GAL1 egl1cDNA$^{T1360\mathbf{A}::G1361}$ or Y294[pLEM1] (du Plessis et al., 2010) was not affected because the catalytic domain was not mutated. The mutated *egl1* in Y294[pLEM1] was observed to have almost a two fold increase in extracellular EGI activity over the wildtype *egl1*.

During the construction of *cbh2* and *bgl1* cDNA clones, a number of mutant genes containing missense mutations were also generated (See Tables 2.4 and 3.2). The *cbh2* insert within S150PGK1cbh2 cDNA$^{G1076\mathbf{C},\ G1345\mathbf{C}}$ had missense mutations at positions 1076 and 1345 (Chapter 3, Fig 3.12.B). The errors occurred within the catalytic domain of the enzyme at amino acid residues 359

and 449 (Fig 4.1.B). These mutations resulted in the replacement of arginine and alanine residues by prolines, respectively. Addition of prolines to a sequence may disrupt the native structural conformation. The errors were corrected by site-directed mutagenesis to yield S150PGK1cbh2cDNA^{A896C} (Chapter 3, Fig 3.11.C). Sequencing of the insert revealed a new mutation at nucleotide position 896 which altered an asparagine to threonine at residue 299 (Fig 4.1C). Another round of corrective mutagenesis was performed. The resultant S150PGK1cbh2cDNA^{G142C} corrected the error at 896, but introduced a mutation at nucleotide 142. The strain S150PGK1cbh2cDNA^{G142C} was not tested however. This error occurred at amino acid residue 48 which is situated in the CBM. A final round of site-directed mutagenesis resulted in S150PGK1cbh2HIScDNA^{wt}; the *cbh2* insert had no missense errors. The investigation for functional CBHII found that only the S150PGK1cbh2HIScDNA^{wt} possessed activity that was statistically significant from the empty vector S150PGK1 strain (Fig 4.6, blue column). This suggested that any mutation within the catalytic domain adversely effected the enzyme functionally. Further experimentation on the mutant strains would be required to reach definitive conclusions.

The mutant strains of *bgl1* cDNA were also examined for activity. The original *bgl1* cDNA clone in strain S150PGK1bgl1cDNA$^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$ had six missense mutations. The nucleotide errors at position 64 and 65 altered a histidine to a serine (Chapter 3, Table 3.2). These errors occurred within the secretory sequence of *bgl1* (Fig 4.1.E). Four other missense errors were located at amino acid positions 354, 362, 700 and 738 within the catalytic domain (Fig 4.1.E). No BGL1 activity was displayed in the supernatant or the cell lysate of this clone (Fig 4.8, B and C respectively). The latter four errors were corrected by site-directed mutagenesis, however the resultant strain S150PGK1bgl1cDNA$^{C64A, A65G, \Delta 1521-2098}$ contained a deletion between nucleotide positions 1521-2098 which encoded amino acids 507-699. The enzyme also contained an error at amino acid 22 (Fig 4.2F). No BGLI activity was detected in the supernatant of this strain (Fig 4.8E), however BGLI activity was detected in the cell lysate indicating that the protein was not secreted most likely due to the error in the signal sequence (Fig 4.8.F). The deletion of 578bp of sequence between positions 1521-2098 did not appear to inactivate enzyme functionality. Further experimentation would be required to

conclusively assess the enzyme activity of S150PGK1bgl1cDNA$^{C64A,\ A65G,\ \Delta1521\text{-}2098}$. This finding indicated that the errors at nucleotide positions 1061, 1085, 2099 and 2212 were sufficient to inactivate functionality in BGL1 and the remedy of these errors restored enzyme activity. A final *bgl1* clone was generated that was full length, had the errors within the secretory signal corrected and no missense mutations. The supernatant and lysate of S150PGK1bgl1cDNA^{wt} was demonstrated to have BGLI activity (4.8. I and J respectively). BGLI activity was also demonstrated in the *S. pastorianus* clone C10-51 PGK1bgl1cDNA^{wt} (Fig 4.8.O).

The PASC assay for C10-51PGK1cbh2HIScDNA^{wt} (Fig 4.7.A, blue line) and the liquid X-glucoside assay on concentrated S150PGK1bgl1cDNA^{wt} supernatant (Fig 4.9, orange line) indicated that the recombinant cellulase enzymes follow a standard enzymatic curve with most activity at 50°C occurring within 24h before the rate of reaction reaches a plateau.

The ability of S150PGK1bgl1cDNA^{wt} to grow on cellobiose as a sole source of carbohydrate was determined. Since *S. cerevisiae* does not possess the necessary transporters to uptake cellobiose into the cell, cultures of S150PGK1bgl1cDNA^{wt} were supplemented with and without concentrated supernatant from the same strain that was grown to saturation in SC glucose and had been shown to secrete functional BGLI. Supplementation of SC cellobiose with the supernatant of S150PGK1bgl1cDNA^{wt} supported growth of S150PGK1bgl1cDNA^{wt} on cellobiose as a sole source of carbon (Fig 4.10). When no supplementation was provided, S150PGK1bgl1cDNA^{wt} did not grow in SC cellobiose. This suggested that addition of BGLI enzyme is required to initiate the metabolism of cellobiose by S150PGK1bgl1cDNA^{wt}. Previously van Rooyen et al., (2005) observed that *S. cerevisiae* transformed with *T. reesei bgl1* with a *T. reesei xyn2* secretion signal under the control of a PGK1 promoter produced recombinant BGL1 activity but failed to grow in cellbiose-containing media. The β-glucosidase genes examined in that study were *bglA* from *A. kawachii*, *bglB* from *C. wickerhamii*, *BGL1* from *S. fibuligera* and *bgl1* from *T. reesei*. The study found that no β-glucosidase activity was observed in the culture supernatant for any of the clones, however active BGL was detected within the periplasmic space of the recombinant yeast strains. The yeast strain Y294[SFI] containing *BGL1* of *S. fibuligera* grew in media containing on cellobiose as the sole carbon source (van Rooyen et al., 2005). The strains

containing β-glucosidase genes from either *T. reesei* or *C. wickerhamii* did not grow in cellobiose-enriched media.

Despite the ability to detect functional enzyme activity of all three classes of cellulase, the levels of protein were too low to be detected by either Western blotting, ELISA or Bradford. The low levels of secreted protein in the supernatant may reflect the choice of culture media. Previously du Plessis et al., (2010) observed that *S. cerevisiae* containing *T. reesei egl1* produced more EGI enzyme activity when grown in rich YPD media than SC ura$^-$ media. The higher enzyme activity in YPD media was probably the result of higher biomass levels. Thus perhaps by growing the strains in rich media, more cellulase enzyme may be secreted.

The glycosylation pattern of protein secreted by *S. cerevisiae* may be an important factor. Yeast is known to hyperglycosylate recombinant secretory protein. Cummings and Fowler (1996) discovered that recombinant *T. reesei* BGL1 secreted from *S. cerevisiae* had a molecular weight of 90kDa. The expected size was 75kDa. Deglycosylation of the recombinant protein reduced its size to the predicted size. The additional mass was a result of hyperglycosylation by *S. cerevisiae* (Cummings and Fowler, 1996). Pentïlla and colleagues observed that recombinant *T. reesei* EGI secreted from *S. cerevisiae* was also hyper N-glycosylated (Pentïlla et al., 1987). In each example enzyme activity was observed with the hyperglycosylated cellulases. Hyperglycosylation can adversely affect the detection of a HIS-tag. A study by Borsig *et al* introduced an N-terminal HIS tag to a galactosyltransferase protein and expressed the enzyme in *S. cerevisiae* (Borsig et al., 1997). The HIS tagged protein was enzymatically active in yeast but could not be purified from cell extract. A bulky N-glycan located 69 amino acid from the HIS-tag caused steric hindrance of the HIS-tag. The removal of this glycosylation site prevented steric hindrance. Nucleotide sequencing of the S150PGK1egl1HIS cDNA^{wt} indicated that there was an N-glycosylation site 65 amino acid residues upstream of the C-terminal HIS-tag. Hyperglycosylation of this site may impair the binding efficiencies of anti-HIS antibodies and perhaps could be an reason why HIS-tagged EGI protein could not be detected (Fig 4.4A and B).

The results of plots 4.11A and B indicated that constitutive expression of functional cellulases did not reduce growth rates of yeast. Penttïlla *et al*

observed similar patterns (Penttïla et al., 1987). The control strain in each graph (black line) produced a similar growth curve to the recombinant clones. Figure 4.11 indicated that S150 strains had a faster growth rate and propagated more than the C10-51 strains. *S. pastorianus* strains are polyploid. Torres et al., (2007) suggested that carrying multiple chromosomes interferes with cell proliferation and results in a lower maximum OD 600nm. The carbohydrate source may also have impacted on the growth of C10-51 strains. *S. pastorianus* strains are known to metabolise maltose better than glucose. Therefore perhaps if the growth curve was repeated in YPmaltose, the C10-51 strains would have grown better.

Table 4.1. Summary of functional strains

Strain name	Cellulase activity detected	Location of activity
S150GAL1	NO	-
S150PGK1	NO	-
S150GAL1egl1gDNA	NO	-
S150PGK1egl1 ^{TACTAAC}	NO	-
S150GAL1egl1cDNA $^{T1360A::G1361}$	YES	supernatant
S150PGK1egl1cDNA $^{T1360A::G1361}$	YES	supernatant
S150PGK1egl1HIS cDNA ^{wt}	YES	supernatant
S150PGK1cbh2 cDNA $^{G1076C, G1345C}$	NO	-
S150PGK1cbh2cDNA ^{A896C}	NO	-
S150PGK1cbh2HIS cDNA ^{wt}	YES	supernatant
S150PGK1bgl1cDNA $^{C64A, A65G, T1061C, T1085C, T2099C, A2212G}$	NO	-
S150PGK1bgl1 cDNA $^{C64A, A65G, \Delta 1521-2098}$	YES	Cell lysate
S150PGK1bgl1cDNA ^{wt}	NO	supernatant
C10-51GAL1	NO	-
C10-51PGK1	NO	-
C10-51GALegl1cDNA $^{T1360A::G1361}$	YES	supernatant
C10-51PGK1egl1cDNA $^{T1360A::G1361}$	YES	supernatant
C10-51PGK1 egl1HIS cDNA ^{wt}	YES	supernatant
C10-51GAL1egl1cDNA $^{T1360A::G1361integrated}$	YES	supernatant
C10-51PGK1cbh2HIScDNA ^{wt}	YES	Supernatant
C10-51PGK1bgl1cDNA ^{wt}	YES	supernatant

Chapter 5: Simultaneous saccharification fermentation of a cellulosic substrate by co-culture of cellulase producing yeast strains

5.1 Introduction

Having clearly shown that functional copies of *T. reesei* EGI, CBHII and BGLI were produced and secreted in both *S. cerevisiae* S150 and *S. pastorianus* C10-51 (Chapter 4, Figs 4.2, 4.6, 4.7 and 4.8), the ultimate goal of this project was to determine if the expressed proteins would allow growth of the yeast strains on cellulose as a sole source of carbohydrate and convert the released sugars into ethanol, i.e. simultaneous saccharification and fermentation (SSF). Generation of an organism capable of SSF is the goal of research into bioethanol production as described in Chapter 1, Section 1.12.

Ethanol is produced by fermentation of simple saccharides. The main metabolic pathway involved in ethanol fermentation is the Embden-Meyerhof pathway of glycolysis (Bai et al., 2008). The pathway results in one glucose molecule being metabolised to produce two pyruvate molecules. Under anaerobic conditions, the pyruvate is reduced to acetaldehyde and further reduced to ethanol with the release of CO_2 and net gain of two ATP molecules (Lodolo et al., 2008).

To produce ethanol from cellulose substrates, the cellulose must first be either hydrolysed *in vitro* completely to its component sugars using mineral acids such as sulphuric acid or with a milder pre-treatment using phosphoric acid to solubilise cellulose. The pre-treated cellulose is then subjected to enzymatic degradation *ex vivo* (Golias et al., 2002). The enzymatic hydrolysis of phosphoric acid swollen cellulose by EGI and CBHII releases cellooligosaccharides such as cellobiose and cellotriose (Fujita et al., 2004); neither of these saccharides can be transported into the cell and converted to ethanol by fermentation pathways. The action of BGLI is required to convert cellobiose and other cellooligosaccharides into glucose, therefore in order to degrade and ferment cellulose all the cellulolytic enzymes must act in concert. This project engineered yeast strains S150 and C10-51 to individually express EGI, CBHII and BGLI. In order to degrade and ferment cellulose, a co-culture of equal numbers of yeast cells that separately contain one of the three cellulase encoding genes was set up. Co-culturing may be advantageous as previous studies have indicated that the co-expression of several cellulase enzymes in the same host can result in lower activity of each enzyme than when expressed individually (du Plessis et al., 2010). (Den Haan et al., (2007b) observed that the maximum β-glucosidase activity in *S. cerevisiae*

strain Y294[SFI] which expressed *S. fibuligera BGLI* was 0.77 U mg^{-1} DCW (dry cell weight) whereas the strain Y294 [CEL5] that co-expressed BGLI and *T. reesei* EGI produced 0.48 U mg^{-1} DCW. Therefore, in order to ensure maximum cellulase activity it was decided to co-culture strains expressing individual enzymes. A study by Golias et al., (2002) found that the co-culture of equal cell numbers of *Klebsiella oxytoca* strain P2 and *S. pastorianus* ATCC 26602 supplemented with GenecorSpezyme CP cellulase fermented 100 g l^{-1} of Sigmacell 50 microcrystalline cellulose into 39.6 g l^{-1} of ethanol. The co-cultures of the cellobiase producing bacteria *Klebsiella oxytoca* mixed with yeast *S. pastorianus* produced more ethanol than either of the constituent strains in pure culture at the same inoculum density.

In industrial settings a number of methods are used for determination of ethanol produced as a result of fermentation. These include distillation followed by physical or chemical characterisation, gas chromatography, infrared or enzymatic methods. The most commonly used method is refractometry. A refractometer measures the extent of refraction caused by the level of dissolved sugars in a brew. The unit of measurement is the degree Brix. A Brix value of 1 is the equivalent of 1g of sucrose/100ml. The rationale is that as the fermentation progresses the sugar concentration will decrease as it is metabolized by the yeast to produce alcohol. However, refractometry is not sufficiently sensitive to determine ethanol concentrations less than 1% (Tonelli, 2009).

Cellulose fermentation is expected to yield very low levels of ethanol (Den Haan et al., 2007b; Fujita et al., 2004). For determination of low levels of alcohol, enzymatic procedures using spectrophotometry are required. The most extensively used enzyme in the determination of ethanol is alcohol dehydrogenase, ADH. This enzyme assay was used to determine ethanol content in the cellulosic fermentation. The basis of the assay is that ADH catalyses the reaction

$$\text{Ethanol} + NAD^+ \leftrightarrow \text{Acetaldehyde} + NADH + H^+$$

where NAD acts as a co-enzyme. The reduced NADH displays a UV absorbance at 340 nm whilst the oxidised NAD^+ has no absorbance (Walker, 1992). The more ethanol present in the sample, the higher the absorbance.

5.2. Results

5.2.1. Comparision of refractometry and enzymatic ADH assay for ethanol determination

Before determing the levels of ethanol produced from the co-culturing of cellulase expressing yeast strains using PASC as the sole carbohydrate source, it was necessary to establish an enzyme assay for ethanol measurement. Firstly, the percentage of alcohol determined by a refractometer and the ADH enzyme assay was compared. Micro-fermentations (10 ml) were set up using either synthetic complete medium supplemented with either 10% (w/v) glucose or 10% (w/v) maltose. The medium was inoculated at a final concentration of $1.5x10^7$ ml^{-1}. The fermentation profiles of strains S150PGK1 and C10-51PGK1 (See Chapter 2, Table 2.4 and 2.5) were compared.

5.2.1.1 Refractometry

The Brix values for each brew were plotted against time (Fig 5.1.A). It was observed that for all samples a general decrease in Brix values occurred as time proceeded (Fig 5.1.A). A physical characteristic of fermentation, the production of CO_2 bubbles was evident in most of the fermentation samples. The strain S150PGK1 strain, fermenting on glucose, exhibited the largest decrease in dissolved sugars from Brix of 10 on day 0 to Brix 3.2 on day 10 (Fig 5.1.A, black line). However, this strain fermented poorly when grown in maltose (Fig 5.1.A, purple line). The next largest decrease was produced by C10-51PGK1 strain fermenting on maltose, which resulted in a decrease from Brix 9.7 to 5.3 by day 10 (Fig 5.1.A, turquoise line). C10-51PGK1 fermented poorly when grown in glucose showing a decrease from Brix 10 to 7 (Fig 5.1.A, blue line). This suggested that *S. pastorianus* strain C10-51 preferentially fermented maltose over glucose. The graph displayed that the majority of the sugar depletion, and by extension fermentation, occurred within the first 48-72hr period. The S150PGK1 strain fermenting glucose and C10-51PGK1 fermenting maltose displayed a sharp decline in sugars during this period (Fig 5.1.A, black line and turquoise line respectively). This was followed by a period where the fermentation rate attenuated and gradually levelled out.

The decrease in sugar, as measured using a refractometer, can be used to calculate alcohol content. Brix values are first converted to specific gravity

using a formula as described by (James et al., 2008). The percentage alcohol produced is then determined using the formula:

% Alcohol = [(Initial specific gravity-final specific gravity)*(0.1281)*1000]

On the basis of this formula, the percentage alcohol of the fermentation was: S150PGK1 glucose (3.45%), C10-51PGK1 glucose (1.53%), S150PGK1 maltose (0.25%) and C10-51PGK1 maltose (2.43%) after 10 days of fermentation.

5.2.1.2. ADH assay:

The ethanol levels produced from the above fermentations were also determined using the ADH enzymatic assay. Due to the sensitivity of the ADH assay, the fermentation supernatant samples were diluted 100-fold to ensure samples were in the linear range for spectrophotometric detection. The assay was carried out as described in Chapter 2, Section 2.23. The Absorbance values at 340 nm for each fermentation supernatant sample were plotted against time (Fig 5.1.B). The overall pattern of alcohol production mirrored the decrease in specific gravity (Brix) in the fermentation as observed in Fig 5.1.A with S150PGK1 fermenting glucose producing the most ethanol, followed by C10-51PGK1 grown on maltose, C10-51PGK1 fermenting glucose and finally S150PGK1 grown in maltose (Fig 5.1.B, black, turquoise, blue and purple lines respectively). The rate of alcohol production in Fig 5.1.B mirrors the rate of sugar utilisation in Fig 5.1.A with the majority of the alcohol produced within the initial 48-72 h period.

To calculate the ethanol levels in each fermentation sample from the ADH assay, a standard curve of absorbance of known amounts of ethanol was used. This was performed using an ethanol standard (Sigma-Aldrich) with the quantity of ethanol ranging from 0- 0.9 g l^{-1}. As shown in Figure 5.2, a proportional relationship between absorbance and ethanol content was evident. The alcohol content in each sample was estimated using the standard curve. The percentage alcohol as determined by ADH assay was S150PGK1 glucose (3.3%), C10-51PGK1 glucose (1.26%), S150PGK1 maltose (0.017%) and C10-51PGK1 maltose (1.6%) following 10 days for fermentation.

5.2.1.3. Comparison of alcohol values

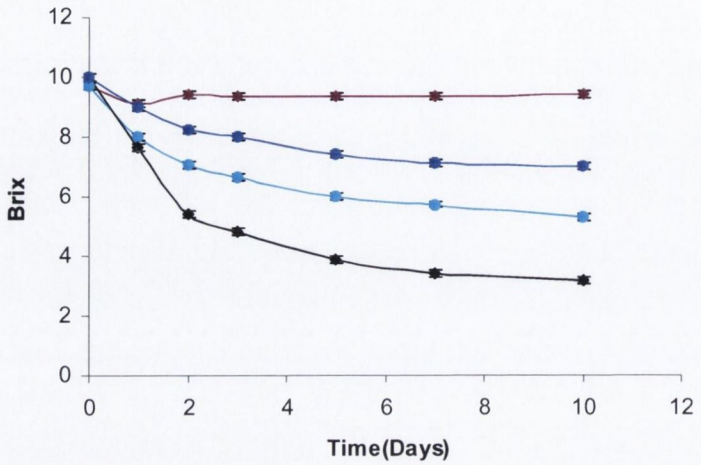
Figure 5.3 summarises the values of percentage alcohol for the brews as calculated by ADH assay (black) or refractometry (orange). The alcohol levels estimated by both methods were generally comparable. The alcohol values calculated by the ADH assay were consistently lower than the refractometer values suggesting that a bias may exist in one method that either over or under estimated alcohol levels. However on the basis of the graph, it was concluded that the ADH assay produced relatively consistent results and was appropriate for use in alcohol determination of alcohol in a cellulosic fermentation.

FIG. 5.1. Fermentation of S150PGK1 and C10-51PGK1 fermenting SC 10% glucose and SC 10% maltose media

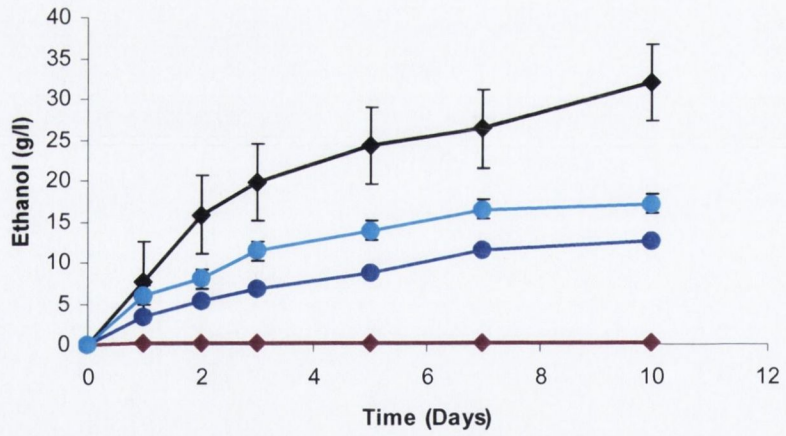
(A) Plot of Brix values against time. Glucose fermentation: S150PGK1 (black line) and C10-51PGK1 (blue line). Maltose fermentation: S150PGK1 (purple line) and C10-51PGK1 (turquoise line). Brix values from fermentation samples supernatant were sampled with a Brixometer. Error bars are present but may be too small to observe.

(B) Plot of ethanol concentration against time for fermentation using the alcohol dehydrogenase (ADH) assay. Glucose fermentation: S150PGK1 (black line) and C10-51PGK1 (blue line). Maltose fermentation: S150PGK1 (purple line) and C10-51PGK1 (turquoise line). Assay performed at 30°C. . Fermentation supernatant samples (15µl) in duplicate were diluted 100-fold to ensure readings were in linear range of the spectrophotometer. Error bars represent standard deviation.

(A)


Brix
Time(Days)
0
2
4
6
8
10
12

(B)


Ethanol (g/l)
Time (Days)
0
5
10
15
20
25
30
35
40
2
4
6
8
10
12

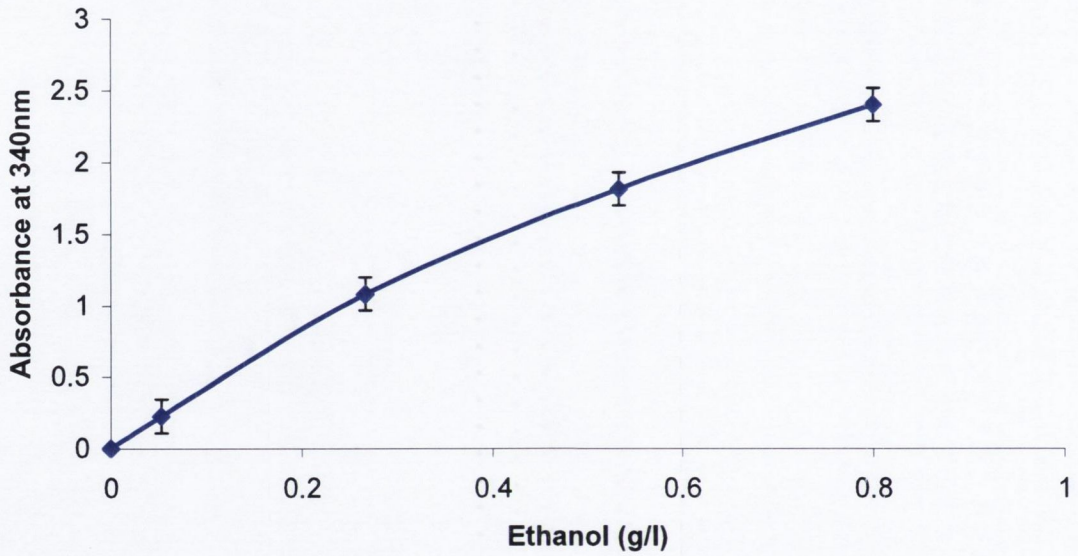


FIG 5.2. Standard curve plot of aborbance at 340nm against alcohol concentration.. A 15µl volume of each sample was assayed in triplicate. Error bars represent standard deviation.

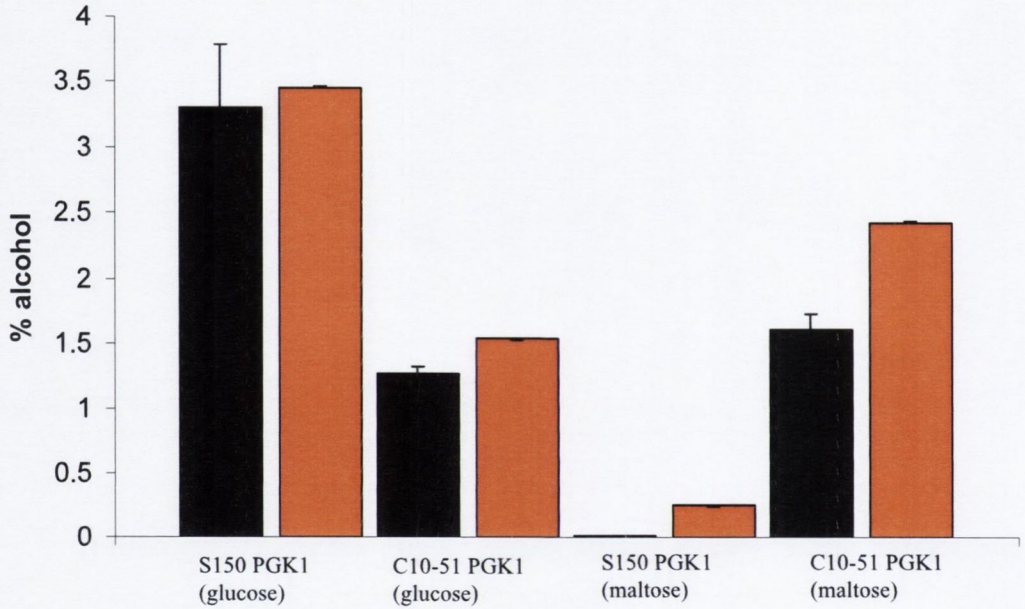


FIG. 5.3. Bar chart comparing percentage ethanol as determined by ADH assay and refractometer. ADH values (black bars). Refractometer (orange bars). Strains and carbohydrate source tested are listed along X axis. Error bars represent standard deviation. Error bars may to too small to observe.

5.2.2. Cellulosic fermentation:

The cloning experiments described in Chapter 3 of this project had generated yeast strains that individually expressed the cellulolytic enzymes EGI, CBHII or BGLI. In order to degrade cellulose all three enzymes must be present and work synergistically. To achieve this, equal numbers of yeast producing each of the cellulase enzymes were co-cultured together in one fermentation reaction (Fig 5.4). The substrate used for fermentation was phosphoric acid swollen cellulose (PASC) as the sole carbohydrate. For yeast to metabolise a sugar, it requires the presence of a functional transporter for uptake into the cell (van Maris et al., 2006). Yeast cells do not possess the transport pathways for cellobiose or any of the cellooligosaccharides. Since PASC was being used as a sole carbohydrate source, this may pose a problem in initiating cell growth and for the initial induction of the PGK1 promoters in order to start cellulase expression within the clones. The experiment described in Chapter 4, Section 4.2.6 indicated that in order for S150PGK1bgl1cDNA ^{wt} to begin metabolising cellobiose, BGL1 must be added to the culture. Therefore, to initiate the fermentation of cellulose, the supernatants of recombinant yeast cultures, which express the cellulases, were added to the fermentation. The purpose of this action was to initiate degradation of the cellulose to release glucose which can then be taken up by the yeast to allow growth and induction of the PGK1 promoter to produce more cellulases. The co-culture of S150 cellulase expressing strains consisted of strains S150PGK1egl1HIScDNA^{wt}, S150PGK1cbh2HIS cDNA ^{wt} and S150PGK1bgl1cDNA ^{wt} (Table 2.4) and is hereafter referred to as S150PGK^{cellulase} co-culture. The C10-51 strains used for the C10-51PGK^{cellulase} co-culture were C10-51PGK1egl1HIScDNA ^{wt}, C10-51PGK1cbh2HIScDNA ^{wt} and C10-51PGK1bgl1cDNA ^{wt} (Table 2.5).

5.2.2.1. Monitoring depletion of sugar in pre-cultures:

Before setting up the cellulosic fermentation, it was first necessary to ensure that the supernatants, which would be used to supplement the PASC fermentation, were fully depleted of all metabolisable sugars so that no sugars were carried over into the PASC fermentation.

The yeast strains S150PGK1, S150PGK1^{cellulase} co-culture, C10-51PGK1 and C10-51PGK1^{cellulase} co-culture were grown in YPD at 25°C. The total sugar concentration in the culture medium was monitored using the phenol sulphuric

acid assay (Chapter 2, Section 2.21) over the course of the experiment. The absorbance values were plotted against time (Fig 5.5).

In the S150 strains glucose levels were almost completely depleted by 30hrs of incubation. No difference was observed in the pattern of sugar usage between the S150PGK1 and the S150PGK1^{cellulase} co-culture (Fig 5.5, black line and green line respectively). The C10-51 strains utilised glucose at a slower rate than observed for the S150. Glucose depletion was not achieved until 50hrs. The pattern of depletion was similar for both C10-51 and C10-51PGK1^{cellulase} co-culture (Fig 5.5, blue line and red line respectively). The slower utilisation of glucose by C10-51 is consistent with the lower levels of ethanol produced by these strains in glucose as observed in Fig 5.1A and B.

5.2.2.2. Simultaneous saccharification and fermentation of PASC

The strategy adopted for experimental SSF is outlined in Fig 5.4. Briefly, yeast strains expressing individually each of the cellulase genes were grown to saturation in 40 ml of SC glucose medium at 25°C for 96 h to ensure glucose depletion. The supernatant of each culture was concentrated 80-fold (See Chapter 2, Section 2.14) and then pooled together to give a final volume of approximately 1.5 ml. Concentration of the sample further ensures that any remaining glucose or any ethanol produced in the pre-culture was removed. All of the concentrated supernatant was added as a cellulase supplement to synthetic medium containing 10% PASC (See Chapter 2, Section 2.1). A standard yeast inoculum for fermentation comprises $1.5x10^7$ cells/ml yeast, therefore $0.5x10^7$ cells/ml each of the three cellulase-expressing strains was mixed together as a co-culture. Thus, the sole source of fermentable carbohydrates can only be generated by the synergistic action of the cellulases on PASC. The fermentations were carried in a volume of 10 ml at 25°C for 10 days. One ml aliquots of the reaction supernatant were taken at Days 1, 5 and 10 and assayed for ethanol content. Control fermentations were also carried out using the strains S150PGK1 and C10-51PGK1 with an inoculum of $1.5x10^7$ cells/ml.

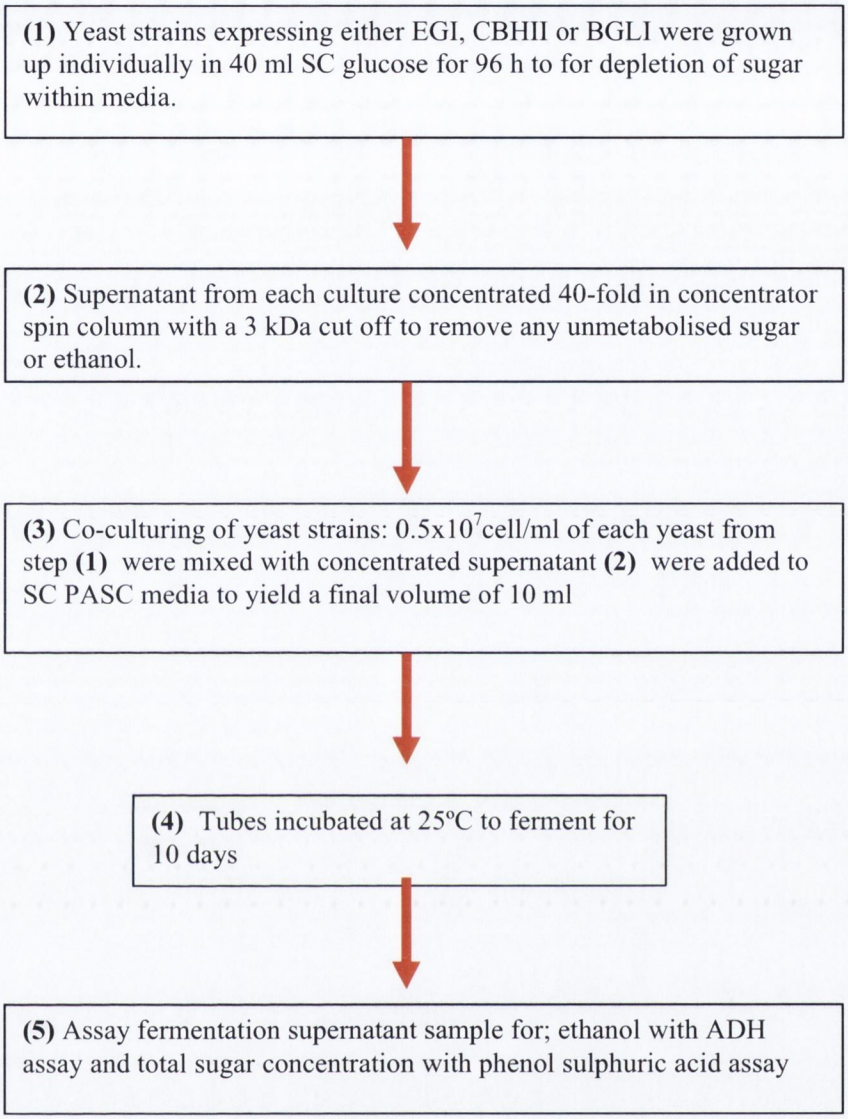


FIG. 5.4. Flow diagram outlining the strategy for SSF of PASC using generated S150 and C10-51 cellulase expressing strains.

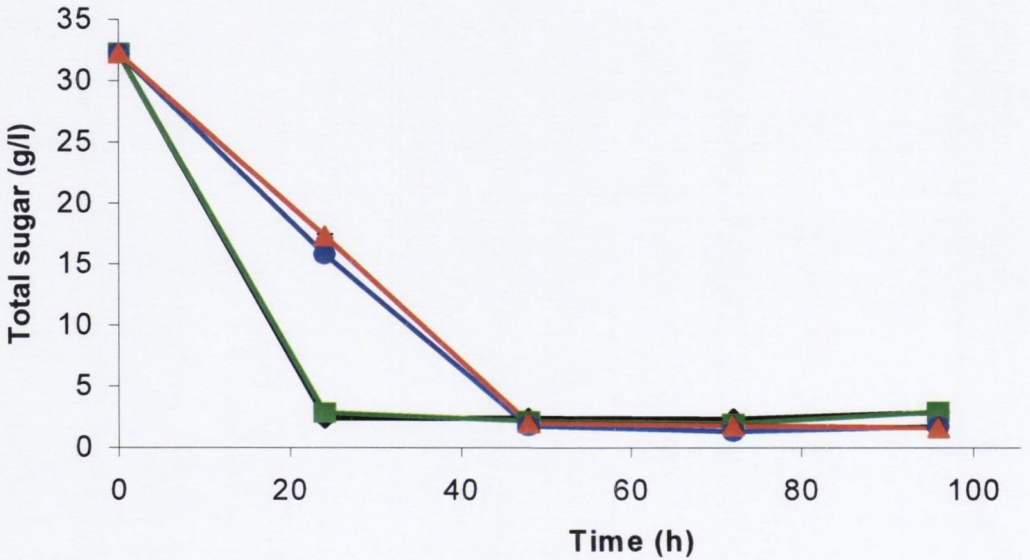


FIG. 5.5. Plot of depletion of sugar within YPD media against time for S150 and C10-51 cultures. S150PGK1 (black line). S150PGK1^{cellulase} co-culture (green line). C10-51PGK1(blue line). C10-51PGK1^{cellulase} co-culture (red line). Data generated from phenol sulphuric acid assay. Samples were assayed in triplicate. Error bars are present but may be too small to observe. Error bars represent standard deviation.

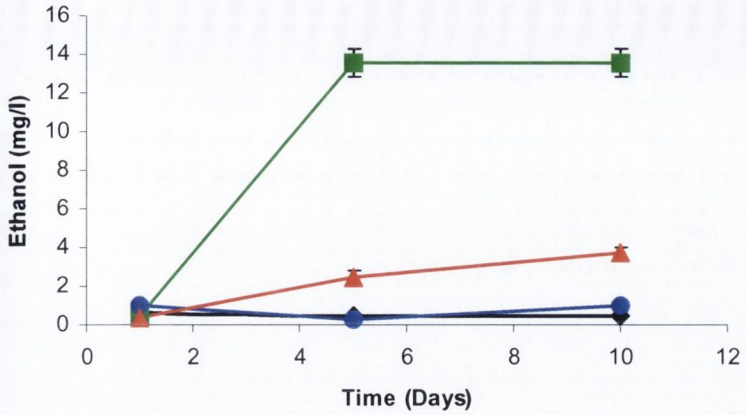
5.2.2.3 ADH assay of cellulosic fermentation samples

Approximately 1 ml aliquots of supernatant from the cellulosic fermentation tubes were taken for ethanol measurement at Days 1, 5 and 10. The ADH assay was performed on 15 µl of the sample aliquot. The Absorbance at 340 nm over 10 days was plotted (Fig 5.6.A). The results indicated that S150PGK1^{cellulase} co-culture (green line) produced the highest absorbance that increased over time (Fig 5.6.A). A slight increase in absorbance over time was also observed for strain C10-51PGK1^{cellulase} co-culture (red line) while no change was evident for S150PGK1 or C10-51PGK1 (black and blue lines respectively). Statistical analysis was performed using Mini-tab statistical software program. A Student-T test was performed to compare the data from the cellulase containing yeast with the control yeast. The threshold for statistical significance was set at the standard p=0.05. Any p values less than this were deemed significant. A T-test comparing S150PGK1^{cellulase} co-culture to S150PGK1 produced a p value = 0.024. Figure 5.6.A demonstrates that C10-51PGK1^{cellulase} co-culture (red line) had a higher absorbance and therefore produced more ethanol than the control C10-51PGK1 (blue line). Statistical analysis comparing these two data sets calculated a p-value of 0.058. This indicated that there was no statistical significance between the C10-51PGK1^{cellulase} co-culture and C10-51PGK1 samples.

Using a 15 µl aliquot of the 10 ml fermentations generated very low absorbance values from the ADH assay (Data not shown). The readings for the 15µl samples were converted into ethanol concentrations using the standard curve (Fig 5.6.A). However, the absorbance values were at the lower end of the ethanol standard curve and therefore the subsequent extrapolation and conversion of these absorbance values to corresponding ethanol concentrations may not be completely accurate. The ADH assay was repeated using a larger aliquot (200 µl) of the medium. The absorbance readings were higher and within the standard curve (data not shown). As shown in Fig 5.6.B the larger sample volume produced a similar pattern as that observed in Fig 5.6.A. The S150PGK1^{cellulase} co-culture produced a demonstrably higher absorbance value than S150PGK1 (Fig 5.6.B, green line and black line respectively). The C10-51PGK1^{cellulase} gave higher values than C10-51PGK1 (Fig 5.6.B, red line and blue line respectively).

By virtue of the fact that the absorbance readings of the 200µl samples occurred within the linear range of the ethanol standard curve, these values were used to generate ethanol concentrations. The alcohol content from the cellulose fermenting brews was determined using the ADH standard curve (Fig 5.2). The alcohol content for S150PGK1, S150PGK1^{cellulase} co-culture, C10-51PGK1 and C10-51PGK1^{cellulase} co-culture was calculated based on the data from Fig 5.6.B to be approximately 1.03, 8.15, 0.83 and 1.77 mg/L respectively (Figure 5.7). The maximum theoretical yield for yeast fermentation is 0.51 g ethanol produced/g of sugar consumed. The 8.15 mg/L ethanol yield of S150PGK1^{cellulase} co-culture was approximately 0.015% of the theoretical yield.

(A)



(B)

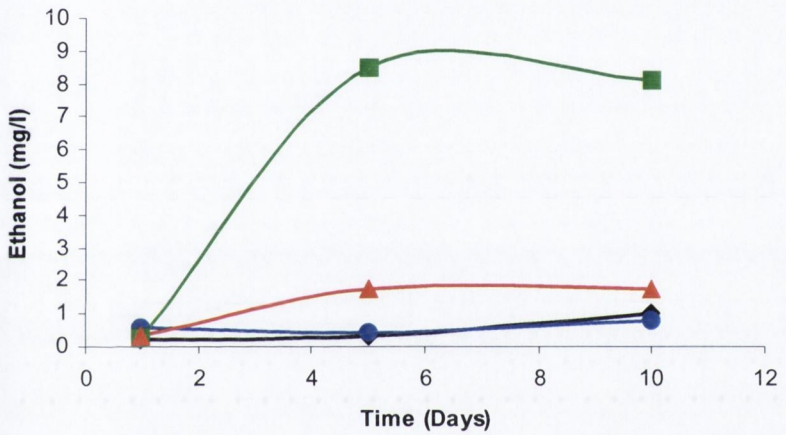


FIG 5.6. Plot of ADH assay results from fermentation of PASC.

(A) 15 µl of each fermentation sample supernatant was used for alcohol determination using the ADH assay; S150PGK1 (black line), S150PGK1^{cellulase} co-culture (green line), C10-51PGK1 (blue line) and C10-51PGK1^{cellulase} co-culture (red line). Duplicate readings taken from each sample. Error bars represent standard deviation.

(B) 200 µl of fermentation sample supernatant was used for alcohol determination using the ADH assay; S150PGK1 (black line), S150PGK1^{cellulase} co-culture (green line), C10-51PGK1 (blue line) and C10-51PGK1^{cellulase} co-culture (red line). Plot based on a single reading for each sample.

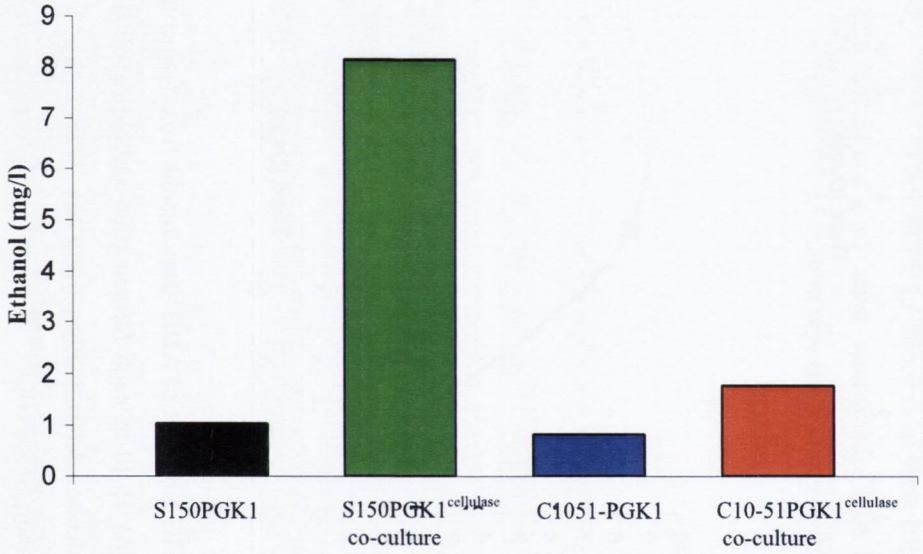


FIG 5.7. Bar chart displaying ethanol content calculated at the end of fermentation of PASC. S150PGK1 (black bar). S150PGK1^{cellulase} co-culture (green bar). C10-51PGK1^{cellulase} (blue bar). C10-51PGK1cellulase co-culture (red bar). Ethanol values displayed as mg l^{-1}.

5.2.2.4 Total sugar assay of cellulosic fermentation samples

The cellulosic fermentation supernatant samples were also tested for total sugar content using the phenol sulphuric acid assay (See Chapter 2, Section 2.21). Aliquots (250 µl) of the supernatant were assayed over the time course of the fermentation. A baseline of absorbance at 492 nm of approximately 0.2 was present in each fermentation media at the start of the reaction (Figure 5.8). The supernatants of S150PGK1^{cellulase} co-culture and C10-51PGK1^{cellulase} co-culture displayed an increase in total sugar content during the fermentation reaction (Fig 5.8, green line and red line respectively). The control yeast cultures S150PGK1 (black line) and C10-51PGK1 (blue line) did not display an increase in the release of additional sugars during fermentation. If cellulose was fully degraded to glucose, and immediately taken in by the yeast, a decrease in the level of sugar might be expected as was observed by Fujita et al., (2004). However, as cellulases degrade cellulose, other cello-oligosaccharides such as cellotriose are released that can not be taken up by the cell. The release of such oligosaccharides will increase the sugar content in the sample supernatant. Based on the data generated, it appears that the cellulases were active but that sugars other than glucose were generated. Analysis of the supernatants on CMC plates confirmed the presence of functional EGI in the pre-fermentation supernatants of C10-51PGK1^{cellulase} co-culture and S150PGK1^{cellulase} co-culture at the start of the fermentation (Fig 5.9 A and C, respectively). EGI activity was also detected in the Day 10 fermentation supernatant sample of C10-51PGK1^{cellulase} co-culture and S150PGK1^{cellulase} co-culture (Fig 5.9 E and F). No activity was evident in the Day 0 fermentation samples of C10-51PGK1 and S150PGK1 (Fig 5.9, B and D).

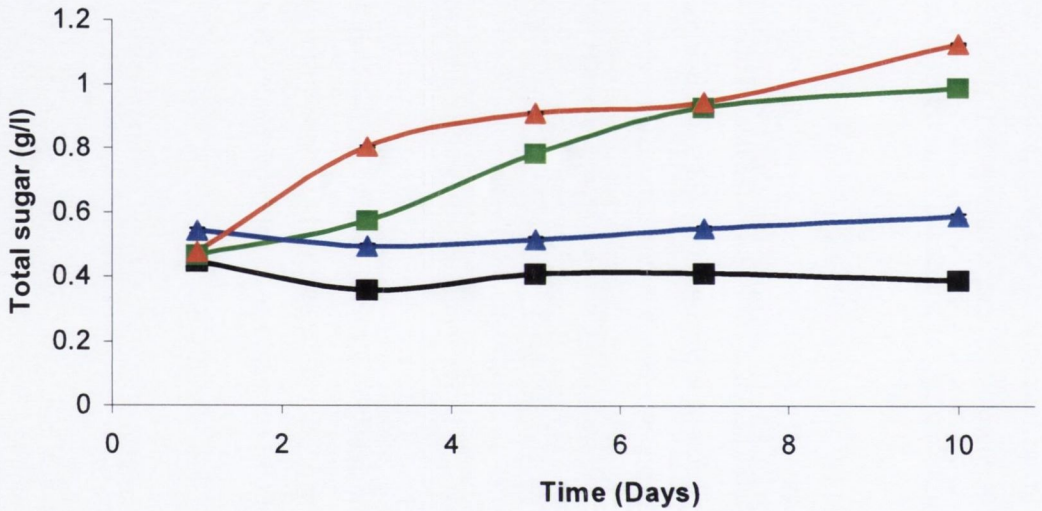


FIG. 5.8. Plot of total sugars present in cellulose fermentation samples supernatant against time. S150PGK1 (black line), S150PGK1^{cellulase} co-culture (green line), C10-51PGK1 (blue line) and C10-51PGK1^{cellulase} co-culture (red line). The data was generated using the phenol sulphuric acid assay. Samples were assayed in triplicate. Error bars are included but may be too small to observe. Error bars represent standard deviation.

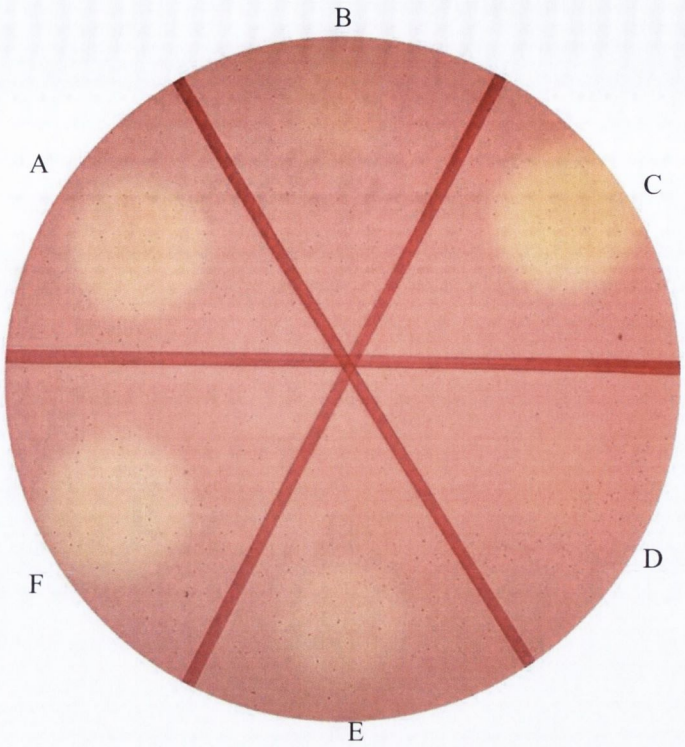


FIG 5.9. CMC plate assay on supernatant of PASC fermentation samples.

(A) C10-51PGK1^{cellulase} co-culture Day 0. (B) C10-51PGK1 Day 0 (C) S150PGK1^{cellulase} co-culture Day 0. (D) S150PGK1 Day 0. (E) C10-51PGK1^{cellulase} co-culture Day 10. (F) S150PGK1^{cellulase} co-culture Day10.

Aliquots (20 µl) of unconcentrated supernatant tested as described in Chapter 2, Section 2.17.

5.3 Discussion:

The goal of this chapter was to attempt SSF of acid- treated cellulose into ethanol using the cellulase encoding yeast strains generated in Chapter 3. The enzymatic ADH assay was used to determine if ethanol was produced by the fermentation cultures. This method is very sensitive and can detect low levels of ethanol present in a sample. The alcohol levels present in a control fermentation as determined by the ADH assay were compared to those determined by refractometry which is the industrial standard for alcohol determinant. The results presented in Figure 5.3 indicated that the methods were comparable. The data shown in Figures 5.1.A and 5.1.B indicated that the choice of carbohydrate has a significant effect on fermentation performance. The results demonstrated that S150PGK1 favoured utilising glucose over maltose. Glucose is energetically favourable to be taken in by the yeast cell because it is transported into yeast cells by diffusion whereas α-glucosides such as maltose requires the expression of a transport system characterised by H^+symporters that are driven by the electrochemical proton gradient across the plasma membrane. Maltose did not appear to be metabolised by S150PGK1. Studies performed by the VTT institute observed that *S. cerevisiae* is a maltose-negative species and does not grow on maltose (Vidgren et al., 2005). The strain does not contain any of the maltose transporter *MAL* genes. *S. cerevisiae* does contain the α-glucosidase transporter (*AGT1*) but it is not known to encode for a function transporter. *S. pastorianus* strains express a full range of maltose transport genes and can grow on maltose (Vidgren et al., 2010). During the fermentation of media with mixed saccharides the uptake of sugars is sequential; monosaccharides are utilised first before disaccharides of increasing complexity such as maltose (Lodolo et al., 2008). Glucose is usually taken in preferentially because of the ease of transport. It may be expected that *S. pastorianus* C10-51PGK1 utilizes and ferments glucose more than maltose. The opposite profile was noted (Fig 5.1.A). C10-51PGK1 preferentially fermented maltose over glucose (Fig 5.1.A, turquoise line and blue line respectively). This may explain why the C10-51 strains growth curve in YPD was lower than S150 (Chapter 4, Figs 4.11 A and B). The preference of maltose over glucose may be beneficial for industrial brewing where wort is the fermentation substrate. Wort is a malted barley extract rich in sugars, the major ones being maltose (60%), maltotriose

(25%) and glucose (15%) (Vidgren et al., 2009). *S. pastorianus* C10-51 ferments maltose more rapidly and therefore increases the rate of fermentation. During a fermentation of cellulose, glucose is the only readily fermentable sugar released from the hydrolysis of the cellulose. Based on the pattern observed in figures 5.1.A and 5.1.B; *S. cerevisiae* is better suited for the production of ethanol from cellulose degradation.

The results of the ADH assay from the PASC fermentation indicated that S150PGK1^{cellulase} co-culture produced more ethanol than S150 PGK1 (Fig 5.6.A). Statistical analysis concluded that the data for S150PGK1^{cellulase} co-culture was statistically significant with p=0.024. The data for C10-51PGK1^{cellulase} co-culture gave a higher absorbance than C10-51PGK1 (Fig 5.6.A), however statistical analysis comparing the two data sets generated a p value of 0.058, indicating that there was no difference between the values.

Analysis of the sugar content of the PASC fermentation demonstrated that the supernatants of the co-cultures of S150PGK1^{cellulase} and C10-51PGK1^{cellulase} had increasing amounts of total sugar as the fermentation proceeded (Fig 5.8). This observation indicated that the cellulases present within these cultures were actively hydrolysing the cellulose to release more sugars. However this pattern was not observed by Fujita et al., (2004). These scientists observed that when fermentation of PASC medium by their *S. cerevisiae* strain co-displaying *T. reesei egl2*, *cbh2* and *A. aculeatus bgl1* (See Chapter 1, Section 1.12) began, the total sugar content decreased over time. Glucose was not detected in the fermentation medium during the reaction. The group concluded that any glucose produced was immediately taken up by the yeast cells. This is in contrast to observation by Hong et al., (2007), who noted that when their cellulase producing recombinant *K .marxianus* yeast fermented on 10% cellobiose medium, a gradual increase of glucose was noted. It was postulated that the low concentration of glucose released by the gradual hydrolysis of cellobiose was not sufficient to activate fermentation and thus the glucose accumulated (Hong et al., 2007). Du Plessis et al., (2010) observed an accumulation of cellobiose during the fermentation of PASC by recombinant *S. cerevisiae* co-expressing *T. reesei egl2*, synthetic *cbh1* and *S. fibuligera bgl1*. The strain was unable to grow on PASC due to the inefficient β-glucosidase activity. Thus β-glucosidase activity is the rate limiting factor. Due to time constraints it was not possible to analyse the sugar content of the PASC

fermentations. Future studies should include HPLC analysis to determine the composition of the medium.

The alcohol content of the PASC fermentation samples was determined using an ADH standard curve. The highest value produced was S150PGK1^{cellulase} co-culture at approximately 8.15mg l^{-1} (Figure 5.7, green column). The C10-51PGK1^{cellulase} co-culture fermented poorly and only produced approximately 1.77 mg l^{-1} (Fig 5.7 red column). This may be due to the temperature at which the fermentation was performed. *S. pastorianus* strains ferment better at low temperatures 6-14°C (Vidgren et al., 2010). Studies in the literature have indicated that the ethanol yield from cellulosic fermentation is expected to be low. Den Haan *et al* fermented PASC with a strain of *S. cerevisiae* expressing *T. reesei* EGI and *Saccharomycopsis fibuligera* BGL1to yield 1 g l^{-1} of ethanol (Den Haan et al., 2007b). Fujita *et al* fermented PASC with a strain of *S. cerevisiae* expressing *T. reesei* EGI, CBHII and *Aspergillus aculeatus* BGL1 to produce 3 g l^{-1} of ethanol. The maximum level of ethanol produced by this project was over 370-fold lower than Fujita. A major difference between the SSF performed in this project and the cellulosic fermentation performed by Fujita and colleagues (2004) and Den Haan et al., (2007b) was the choice of β-glucosidase. The former study chose the *bgl1* gene of *A. aculeatus*, the latter study used *BGL1* from *S. fibuligera*. The *Aspergillus* species are known to produce β-glucosidase with greater specific activity than *T. reesei*. Indeed, many commercial *T. reesei* cellulase preparations are supplemented with *Aspergillus* β-glucosidases. Hence the use of *T. reesei bgl1* may have been disadvantageous. Another difference between this project and the work of Fujita et al., (2004) and Den Haan et al., (2007b) was the choice of expression vector. In this project the pGREG vector with an ARS/CEN origin of replication was used. ARS/CEN plasmids are known to be stably transmitted from mother to daughter cell but with a low copy number. Den Haan and colleagues (2007b) used a multicopy episomal expression vector to express *T. reesei egl1* and *S. fibuligera bgl1*. Previous studies have indicated that *S. cerevisiae* expressing *bgl1* in a 2μ plasmid produced five-fold more β-glucosidase than *S. cerevisiae* expressing an integrated copy of *bgl1* (du Plessis et al., 2010).

The generation of yeast biomass may be crucial. The fermentation of PASC in this project started with a high density yeast inoculum of 1.5 x10^7 cells ml^{-1}.

At such a high concentration of yeast during fermentation, the yeast may be expected double only once or twice before reaching stationary phase. Den Haan et al., (2007b) began fermenting PASC with a low inoculum of yeast, $2x10^5$ cells ml^{-1}. Starting fermentation at a low cell density results in the exponential growth of cells and the generation of significant levels of yeast biomass. This would lead to the secretion of ever increasing amounts of cellulases.

The fermentation technique used in this project could be altered to improve ethanol yield. In this project 10 ml mini-brews were performed in test-tubes whereas Fujita et al., (2004) and Den Haan et al., (2007b) performed 100 ml fermentations in 250 ml conical flasks. The latter groups performed their fermentation with shaking. Shaking was found to greatly enhance enzymatic cellulosic degradation (Kleman-Leyer et al., 1996). In this study, the traditional method of fermentation was employed where the reaction vessels were incubated stationary at an angle. This is performed in industrial brewing so that the only source of yeast agitation is via the production and circulation of CO_2 (Boulton and Quain, 2001). This method may be appropriate when fermenting dissolved sugars but was ineffective when fermenting PASC slurry. Fermentation temperature may also be a key factor. The rates of most enzyme catalysed reactions double for each 10°C increase in temperature (Vidgren et al., 2010). The optimum temperature for cellulase activity is 50°C. *Saccharomyces* yeast cannot survive sustained exposure to this temperature. The optimum growth temperatures for *S. cerevisiae* and *S. pastorianus* are 30°C and 13°C respectively. Cellulase enzymes would not be efficient at the latter temperature. Most cellulosic fermentations using *S. cerevisiae* are performed at 30°C (Den Haan et al., 2007b; Fujita et al., 2004). Studies of C10-51 indicated that at 20°C it undergoes significant chromosomal rearrangements due to stress. Clearly fermenting C10-51 at 30°C would be more stressful so a compromise of 25°C was selected on to minimise stress for C10-51, which may have compromised maximum cellulase activity.

Chapter 6: General Discussion

General discussion:

The aim of the research reported in this thesis was to bioengineer strains of the *Saccharomyces* yeast species *S. cerevisiae* and *S. pastorianus* to individually express and secrete the *T. reesei* cellulase enzymes EGI, CBHII and BGLI (Chapter 4, sections 4.2.1, 4.2.3 and 4.2.4). In previous research, cellulase genes from different species were cloned and expressed in *S. cerevisiae* strains with varying success. In a few cases, sucessful simultaneous saccarification and fermentation was achieved, although alcohol yields remained low. In the majority of reports to date, the *Saccharomyces* species of choice for expression of cellulases has been *S. cerevisiae*. The species *S. pastorianus*, previously referred to as *S. carlsbergensis* was originally isolated as a pure culture in 1883 by Emile Christian Hanson in the Carlsberg brewery. This species is a robust hybrid, adapted to the stressful conditions of the Industrial Brewery and capable of producing high yields of ethanol and thus presents as an alternative species for cellulase expression. The data presented in this thesis reports the expression of *T. reesei* cellulases into a strain of *S. pastorianus* for the first time.

The ultimate goal of this thesis was to examine the capability of both *S. cerevisiae* and *S. pastorianus* cellulase expressing strains to carry out simultaneous saccharify and ferment a cellulosic substrate into ethanol. Since previous reports had shown that co-expressing either endoglucanases (EGI and EGII) with the CBHI, or β-glucosidase, or both, led to lower extracellular endoglucanase activity compared to individual expression, an alternative strategy was employed in this study. Here the cellulase genes, *egl1*, *cbh2* and *bgl1* were individually expressed in either *S. cerevisiae* and *S. pastorianus* strains and then the yeast cells were co-cultured to produce all three enzymes. The reduction in enzyme activity through co-expression of multiple genes is consistent with findings from a previous study where co-expression of secreted enzymes (xylanase and β-xylosidase) led to lower activities than those achieved by individual expression (La Grange et al., 2001).

Simultaneous saccharification and fermentation was carried out using PASC as the substrate and co-culture of yeast clones producing cellulase (Chapter 5, Section 5.2.3). In order to initiate cellulose degradation, the fermentations were supplemented with supernatants of all three yeast strains expressing the

cellulases individually. This novel approach ensures that through the action of the cellulases, glucose can be released to facilitate growth of the yeast on PASC which can then be sustained by subsequent production and secretion of the cellulases from the cells. The cellulase expressing *S. cerevisiae* co-culture was demonstrated to produce more ethanol than the control *S. cerevisiae* culture lacking the genes and was shown to be statistically significant (Chapter 5, Fig 5.7). A maximum of 8.15 mg l^{-1} of ethanol was recorded. The *S. pastorianus* cellulase co-culture did not perform as expected and the level of ethanol was not significant in relation to the *S. pastorianus* control.

There are several key areas to improve the cellulose degradation and bioethanol yield.

For future work it may be desirable to quantify the amount of each cellulase enzyme that was produced by a yeast clone. The degradation of cellulose to glucose depends on achieving an optimal ratio of enzymes (du Plessis et al., 2010). Previous studies have shown that cellobiose has a strong inhibitory affect on the activity of other *T. reesei* cellulases (Gruno et al., 2004). Thus the effective conversion of cellobiose to glucose is an essential step in SSF overproduction of cellobiose. Therefore maintaining the correct ratio of all three enzymes may be critical to ensure efficient production of glucose from cellulose. In a previous study by du Plessis et al., (2010), the incubation in PASC of *S. cerevisiae* strain Y294[BGL1*fur1*::*LEU2*pLEM1C], which secreted BGL1, EGI and CBHI, produced a glucose to cellobiose ratio of 1:6.3, ie 6-fold more cellobiose than glucose. This strain was unable to sustain growth on PASC due to inefficient β-glucosidase activity. In the SSFs carried out in this thesis, low levels of alcohol were obtained with the co-culture of *S. cerevisiae* cellulase expressing strain. Total sugar analysis was carried out over the time period of the fermentation. The results indicated that total sugar content of the supernatant increased as fermentation proceeded. Since any glucose generated from cellulose should be taken up immediately by the cells, this suggests that sugars other than glucose are being produced. The most likely oligosaccharides that might accumulate would be cellobiose and/or cellotroise. Time constraints did not allow further investigation of the carbohydrate content of the medium, however analysis of the medium by HPLC might indicate the rate limiting step in the process.

Expression of a heterologous enzyme can enable growth on a non-native

extracellular substrate if the enzyme converts the non-native substrate into a substrate that can be taken up and metabolised by the cell provided that the rate of supply of the substrate is equal to the rate at which the cells can consume the substrate (McBride et al., 2005). Therefore knowing the concentration of each cellulase is essential. Efforts were performed to semi-quantify the level of EGI secreted by S150PGK1egl1cDNAHIS^{wt} (Chapter 4, Section 4.2.1). The EGI was semi-quantified by comparing diameters of zones of hydrolysis from 50-fold concentrated supernatant of yeast clone with cellulase standards on a CMC agar plate (Chapter 4, Fig 4.3 A-F). A 50ml SC glucose culture of S150PGK1egl1HIS cDNA^{wt} produced approximately 30mU ml^{-1} of EGI enzyme. More accurate measurements could be obtained in future studies using liquid CMC assay to measure the levels of reducing sugars released (Wood and Bhat, 1988). CBHII activity could be quantified by measuring reducing sugars released from reaction with PASC or crystalline cellulose compared to cellulase standards. The level of BGLI produced by S150PGK1bgl1cDNA ^{wt} in 50ml of supernatant of a SC glucose culture was quantified using a liquid X-glucoside assay (Chapter 4, Section 4.2.5). The cellulase standards of 50mU and 10mU were used as positive controls. The recombinant yeast supernatant sample was shown to produce approximately 100mU ml^{-1}. However a standard format to present enzyme activity produced by recombinant yeast is as units per dry cell weight (Den Haan et al., 2007b; Fujita et al., 2004). A more accurate measurement of BGLI activity might be obtained by measuring the release of p-nitrophenol from the degradation of p-nitrophenyl-B-D-glucoside (PNPG). The p-nitrophenol is measured by spectrophotometry at 405nm (Den Haan et al., 2007b). Future work should focus on quantifying BGLI using the standard PNPG assay so that values derived from the BGLI expressing clones from this thesis can be compared to the levels reported elsewhere in the literature.

Future studies could also focus on protein analysis. The cellulase enzymes should be purified from *T. reesei* supernatant and from yeast growth culture supernatant using affinity chromatography with Sepharose columns as described by other groups (Okada et al., 1998; Tomme et al., 1988). The activity of each enzyme produced by recombinant yeast could have been compared to cellulase secreted by native *T. reesei*. This experiment could determine if hyperglycosylation by *S. cerevisiae* affected cellulase activity.

The attachment of a 6xHIS tag to the C-terminal of EGI and CBHII (Chapter 3, Section 3.2.9) allows for the possible purification of the respective enzymes with a nickel column. The immunoblotting of the HIS tagged EGI did not detect the protein in culture supernatant (Chapter 4, Section 4.2.2). The choice of growth culture may be crucial. The yeast clones were grown in synthetic complete minimal media. Heterologous enzyme secretion is usually low in *S. cerevisiae*, however growth in a minimal media will result in an even lower concentration of secreted protein. Du Plessis et al., (2010) reported that growing yeast clones transformed with *egl1* in YPD media resulted in higher enzyme activity than the growth of the same clone in SC media. Van Wyk et al., (2010) grew *S. cerevisiae* strains transformed with cellulase encoding genes in double strength SC media. Thus for future studies, recombinant yeast strains should be cultivated in rich medium to maximise the amount of cellulase enzyme produced.

A fortuitous development during the cloning of the cellulase genes was the isolation of mutant forms of the proteins. This series of mutants provides an opportunity to examine in more detail the regions of the CD and CBM domains that are required for enzyme activity. It would be of interest to compare the EGI activity of S150PGK1egl1HIS cDNA^{wt}, which contained the wildtype *egl1* gene, with S150PGK1egl1cDNA $^{T1360A::G1361}$, which had a mutation that caused a frame shift within the CBM and the addition of 20 extra amino acids to the C-terminus of the protein. Du Plessis and colleagues (2010) generated an *egl1* clone with a single amino acid transition in the CBM and observed an increase in endoglucanase activity compared to wildtype *egl1*. Further analysis is warranted on strain S150PGK1bgl1 cDNA $^{C64A,\ A65G,\ \Delta1521\text{-}2098}$, which despite containing a 578bp deletion in the CD domain, still demonstrated BGLI activity within its cell lysate (Chapter 4, Fig 4.8.F). A mutation in the signal sequence abolished secretion of the protein from the cell. Thus quantitative analysis of the enzyme activity of these mutants may provide insight into the essential regions of the proteins or as in the case of the mutant described in (du Plessis et al., 2010), identify alterations that can increase activity.

Characterising the cellulases at different temperatures is an important factor that should be performed in future. *T. reesei* cellulases are known to be optimum at 50°C however the enzyme half-life would be reduced faster at this elevated temperature. In fact recent experiments in the Bond laboratory

indicate that the half-life of the cellulases is dramatically increased at higher temperatures. Despite having an optimum temperature of 50°C for enzyme activity *in vitro*, maximum enzyme activity was obtained for BGLI at 30°C *in vivo* (W. Kricka and U. Bond, unpublished data). In this project the fermentation of PASC was performed at 25°C, which is half the optimum temperature. For most enzymatic reactions, every increase of 10°C doubles the rate of the reaction (Vidgren et al., 2010). Thus repeating the PASC fermentation at 30°C would ensure that the cellulase enzymes performed with better efficiency and hence may result in an increased production of ethanol. *S. pastorianus* ferments optimally at low temperatures ranging from 6-14°C. Enzyme activity at this low temperature may be reduced compared to 25°C. However, cellulases must be able to perform at lower temperatures because in nature cellulolytic fungi can be found growing on tree barks in woodland habitats at low temperatures. Thus further studies are required to determine the optimum growth temperature to obtain optimum enzyme activity.

The choice of donor cellulolytic organism could be also reviewed. *T. reesei* was chosen because it has been very well characterised and it possess all of the necessary enzymes required to fully degrade cellulose to glucose. However *T. reesei* is known to produce low levels of BGL. *Aspergillus* species are known to produce more BGL. The β-glucosidase produced by *Aspergillus* has higher glucose tolerance than the *T. reesei* enzyme (Lynd et al., 2002). The cloning of a β-glucosidase gene from *Aspergillus* into yeast may improve cellulose saccharification levels and hence increase ethanol yield. Alternatively the cloning of cold adapted cellulase from psychrophilic organisms could be considered. The psychrophilic gram negative bacterium *Pseudoalteromonas haloplankis* produces a cold adapted endoglucanase enzyme Cel5G (Sonan et al., 2007). The Cel5G enzyme is closely related to Cel5A produced by the mesophilic bacteria *Erwina chrysanthemi*. Studies have indicated that the activity of the psychrophilic Cel5G cellulase at 4°C was similar to the values of Cel5A at 35°C (Garsoux et al., 2004). The expression of cold-adapted cellulases in *S. pastorianus* would enable fermentation be performed at 6-14°C without the loss of enzyme activity.

Ultimately a goal of the project was to introduce all three cellulase encoding genes into the genome of the same *S. pastorianus* strain. The aim of consolidated bioprocessing (CBP) is to exploit a single organism that can

degrade cellulose to glucose and the ferment the sugars to ethanol immediately. Chapter 3, Section 3.2.11 described an attempt to integrate the GAL1egl1cDNA$^{T1360A::G1361}$ cassette into the intergenic region found between YPR159 and YPR 160 of chromosome XVI in the *S. pastorianus* genome. Data indicated that the *egl1* gene was present within the DNA of *S. pastorianus* colony but was not present at the expected integration site (Chapter 3, Fig 3.15). For future work genome sequencing of the strain C10-51GAL1egl1cDNA $^{T1360A::G1361integrated}$ (Chapter 2, Table 2.5) should be performed. While generating a more stable yeast strain, the integration of a single copy of each gene in the chromosome of *S. pastorianus* may not produce sufficient amounts of cellulase to convert cellulose to glucose. Du Plessis et al., (2010) observed *S. cerevisiae* strain Y294[BGL1*fur1*::*LEU2*pLEM1C] which had a copy of *S. fibuligera bgl1* integrated into its genome could not grow on PASC due to inefficient β-glucosidase activity. A yeast strain with an integrated copy of *bgl1* was observed to have 5-fold less β-glucosidase activity than yeast expression *bgl1* from a 2μ based plasmid (McBride et al., 2005). Therefore the integration of several copies of cellulase inserts may be required.

The selection of an alternative *S. pastorianus* strain could improve fermentation. The strain C10-51 was selected because it was the 2nd generation of stress-tolerant *S. pastorianus* generated in the laboratory (James et al., 2008). Its genome was similar to the parental strain C10. However recent work in the laboratory comparing growth rates of *S. pastorianus* stress-tolerant strains indicated that C10-51 had the slowest growth rate (T.C. James, U. Bond: unpublished data). The 1st generation stress-tolerant *S. pastorianus* strain C10 may be employed as an alternative. The genome of C10 was found not to differ significantly from its parental CMBS-33. Strain C10 displayed an increased fermentation rate compared to its parent (T.C. James, Bond unpublished data). Therefore a qualitative analysis of *S. pastorianus* strains should have been performed before the final choice of strain was selected.

A future goal of this project is the conversion of a cellulosic biomass into bioethanol. The choice of substrate is important. A key element in the production of bioethanol is sustainability, for example utilizing a locally sourced biomass substrate. A suitable biomass should have high levels of cellulose and low lignin content. The C4 perennial grass species Miscanthus is an ideal substrate. Miscanthus grows readily in Ireland, gives high yields per

hecture, has a long productive life time, low moisture content and low susceptibility to pests (Wang et al., 2010). The lignocellulose content is 46% cellulose, 27.8% hemicellulose and 10.7% lignin (Wang et al., 2010). Hence Miscanthus grass contains a massive source of cellulose that can be potentially converted into bioethanol. However, the conversion of biomass to useable energy will not be economically viable unless hemicellulose is used in addition to cellulose (Jeffries, 2006).

Thus the work presented in this thesis represents a beginning and may be used as the foundations for the eventual development of a yeast strain that is capable of generating bioethanol from biomass.

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1994). Molecular biology of the cell, 3rd edn (New York: Garland Publishing, Inc).

Antoni, D., Zverlov, V.V., and Schwarz, W.H. (2007). Biofuels from microbes. Appl Microbiol Biotechnol *77*, 23-35.

Arantes, V., and Saddler, J.N. (2010). Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnol Biofuels *3*, 4.

Bai, F.W., Anderson, W.A., and Moo-Young, M. (2008). Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnol Adv *26*, 89-105.

Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. Curr Opin Struct Biol *8*, 548-557.

Bhat, M.K., and Bhat, S. (1997). Cellulose degrading enzymes and their potential industrial applications. Biotechnol Adv *15*, 583-620.

Bhatia, Y., Mishra, S., and Bisaria, V.S. (2002). Microbial beta-glucosidases: cloning, properties, and applications. Crit Rev Biotechnol *22*, 375-407.

Bond, U., Neal, C., Donnelly, D., and James, T.C. (2004). Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. Curr Genet *45*, 360-370.

Borsig, L., Berger, E.G., and Malissard, M. (1997). Expression and purification of His-tagged beta-1,4-galactosyltransferase in yeast and in COS cells. Biochem Biophys Res Commun *240*, 586-589.

Boulton, C., and Quain, D. (2001). Fermentation systems. In Brewing yeast & fermentation (Blackwell Science), pp. 260-376.

Bresticgoachet, N., Gunasekaran, P., Cami, B., and Baratti, J.C. (1989). Transfer and expression of an *Erwinia chrysanthemi* cellulase gene in *Zymomonas mobilis*. Journal of General Microbiology *135*, 893-902.

Brown, M.A., Levine, M.D., Romm, J.P., Rosenfeld, A.H., and Koomey, J.G. (1998). Engineering-economic studies of energy technologies to reduce greenhouse gas emissions: Opportunities and challenges. Annu. Rev. Energ. Environ. *23*, 287-385.

Carere, C.R., Sparling, R., Cicek, N., and Levin, D.B. (2008). Third generation biofuels via direct cellulose fermentation. Int J Mol Sci *9*, 1342-1360.

Cassago, A., Panepucci, R., Baiao, A., and Henrique-Silva, F. (2002). Cellophane based mini-prep method for DNA extraction from the filamentous fungus Trichoderma reesei. BMC Microbiol *2*, 14.

Cho, K.M., Yoo, Y.J., and Kang, H.S. (1999). δ-Integration of endo/exo-glucanase and beta-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. Enzyme and Microbial Technology *25*, 23-30.

Conesa, A., Punt, P.J., van Luijk, N., and van den Hondel, C.A. (2001). The secretion pathway in filamentous fungi: a biotechnological view. Fungal Genet Biol *33*, 155-171.

Cornell, M.J., Alam, I., Soanes, D.M., Wong, H.M., Hedeler, C., Paton, N.W., Rattray, M., Hubbard, S.J., Talbot, N.J., and Oliver, S.G. (2007). Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. Genome Res *17*, 1809-1822.

Cullen, D., and Kersten, P. (1992). Fungal enzymes for lignocellulose degradation. In Applied molecular genetics of filamentous fungi, J.R. Kinghorn, and G. Turner, eds. (New York: Chapman & Hall), pp. 100-131.

Cummings, C., and Fowler, T. (1996). Secretion of *Trichoderma reesei* beta-glucosidase by Saccharomyces cerevisiae. Curr Genet *29*, 227-233.

Dashtban, M., Schraft, H., and Qin, W. (2009). Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. Int J Biol Sci *5*, 578-595.

Den Haan, R., McBride, J.E., La Grange, D.C., Lynd, L.R., and Van Zyl, W.H. (2007a). Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. Enzyme and Microbial Technology *40*, 1291-1299.

Den Haan, R., Rose, S.H., Lynd, L.R., and van Zyl, W.H. (2007b). Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng *9*, 87-94.

du Plessis, L., Rose, S.H., and van Zyl, W.H. (2010). Exploring improved endoglucanase expression in *Saccharomyces cerevisiae* strains. Appl Microbiol Biotechnol *86*, 1503-1511.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental

evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *99*, 16144-16149.

Dunn, B., and Sherlock, G. (2008). Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res *18*, 1610-1623.

Erdeniz, N., Mortensen, U.H., and Rothstein, R. (1997). Cloning-free PCR-based allele replacement methods. Genome Res *7*, 1174-1183.

Fernandes, S., Tuohy, M.G., and Murray, P.G. (2010). Cloning, heterologous expression, and characterization of the xylitol and L-arabitol dehydrogenase genes, Texdh and Telad, from the thermophilic fungus *Talaromyces emersonii*. Biochem Genet *48*, 480-495.

Fonseca, G.G., Heinzle, E., Wittmann, C., and Gombert, A.K. (2008). The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl Microbiol Biotechnol *79*, 339-354.

Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., *et al.* (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J Biol Chem *278*, 31988-31997.

Fujita, Y., Ito, J., Ueda, M., Fukuda, H., and Kondo, A. (2004). Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. Appl Environ Microbiol *70*, 1207-1212.

Fujita, Y., Takahashi, S., Ueda, M., Tanaka, A., Okada, H., Morikawa, Y., Kawaguchi, T., Arai, M., Fukuda, H., and Kondo, A. (2002). Direct and efficient production of ethanol from cellulosic material with a yeast strain displaying cellulolytic enzymes. Appl Environ Microbiol *68*, 5136-5141.

Garsoux, G., Lamotte, J., Gerday, C., and Feller, G. (2004). Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*. Biochem J *384*, 247-253.

Gruno, M., Valjamae, P., Pettersson, G., and 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

Golias, H., Dumsday, G.J., Stanley, G.A., and Pamment, N.B. (2002). Evaluation of a recombinant *Klebsiella oxytoca* strain for ethanol production from cellulose by simultaneous saccharification and fermentation: comparison with native cellobiose-utilising yeast strains and performance in co-culture with thermotolerant yeast and *Zymomonas mobilis*. J Biotechnol *96*, 155-168.

Gray, K.A., Zhao, L., and Emptage, M. (2006). Bioethanol. Curr Opin Chem Biol *10*, 141-146.

Gruno, M., Valjamae, P., Pettersson, G., and 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.

Hamacher, T., Becker, J., Gardonyi, M., Hahn-Hagerdal, B., and Boles, E. (2002). Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology *148*, 2783-2788.

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

Henrissat, B., Teeri, T.T., and Warren, R.A. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett *425*, 352-354.

Ho, N.W., Chen, Z., and Brainard, A.P. (1998). Genetically engineered Saccharomyces yeast capable of effective cofermentation of glucose and xylose. Appl Environ Microbiol *64*, 1852-1859.

Hong, J., Wang, Y., Kumagai, H., and Tamaki, H. (2007). Construction of thermotolerant yeast expressing thermostable cellulase genes. J Biotechnol *130*, 114-123.

Idiris, A., Tohda, H., Kumagai, H., and Takegawa, K. (2010). Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol *86*, 403-417.

Ilmen, M., Saloheimo, A., Onnela, M.L., and Penttila, M.E. (1997). Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei.* Appl Environ Microbiol *63*, 1298-1306.

Iwashita, K., Nagahara, T., Kimura, H., Takano, M., Shimoi, H., and Ito, K. (1999). The bglA gene of *Aspergillus kawachii* encodes both extracellular and cell wall-bound beta-glucosidases. Appl Environ Microbiol *65*, 5546-5553.

James, T.C., Campbell, S., Donnelly, D., and Bond, U. (2003). Transcription profile of brewery yeast under fermentation conditions. J Appl Microbiol *94*, 432-448.

James, T.C., Usher, J., Campbell, S., and Bond, U. (2008). Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. Curr Genet *53*, 139-152.

Jansen, G., Wu, C., Schade, B., Thomas, D.Y., and Whiteway, M. (2005). Drag&Drop cloning in yeast. Gene *344*, 43-51.

Jeffries, T.W. (2006). Engineering yeasts for xylose metabolism. Curr Opin Biotechnol *17*, 320-326.

Jeffries, T.W., Grigoriev, I.V., Grimwood, J., Laplaza, J.M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., *et al.* (2007). Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat Biotechnol *25*, 319-326.

Jeffries, T.W., and Shi, N.Q. (1999). Genetic engineering for improved xylose fermentation by yeasts. Adv Biochem Eng Biotechnol *65*, 117-161.

Julius, D., Brake, A., Blair, L., Kunisawa, R., and Thorner, J. (1984). Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. Cell *37*, 1075-1089.

Keranen, S., and Penttila, M. (1995). Production of recombinant proteins in the filamentous fungus Trichoderma reesei. Curr Opin Biotechnol *6*, 534-537.

Kleman-Leyer, K.M., Siika-Aho, M., Teeri, T.T., and Kirk, T.K. (1996). The cellulases endoglucanase I and cellobiohydrolase II of *Trichoderma reesei* act synergistically to solubilize native cotton cellulose but not to decrease its molecular size. Appl Environ Microbiol *62*, 2883-2887.

Kleywegt, G.J., Zou, J.Y., Divne, C., Davies, G.J., Sinning, I., Stahlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T.T., and Jones, T.A. (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.

Knowles, J., Lehtovaara, P., Teeri, T., Penttila, M., Salovuori, I., and Andre, L. (1987). The application of recombinant-DNA technology to cellulases and lignocellulosic wastes. Philosophical Transactions of the Royal Society of London. Series A *321*, 449-454.

Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T.T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., *et al.* (2002). The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. J Am Chem Soc *124*, 10015-10024.

Kubicek, C.P., Mikus, M., Schuster, A., Schmoll, M., and Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. Biotechnol Biofuels *2*, 19.

La Grange, D.C., Pretorius, I.S., Claeyssens, M., and van Zyl, W.H. (2001). Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* beta-xylosidase (xlnD) and the *Trichoderma reesei* xylanase II (xyn2) genes. Appl Environ Microbiol *67*, 5512-5519.

Langford, C., Nellen, W., Niessing, J., and Gallwitz, D. (1983). Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. Proc Natl Acad Sci U S A *80*, 1496-1500.

Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J.K., Culley, D.E., Collett, J.R., Martin, J., Druzhinina, I.S., Mathis, H., Monot, F., *et al.* (2009). Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. Proc Natl Acad Sci U S A *106*, 16151-16156.

Lee, S., Oh, Y., Kim, D., Kwon, D., Lee, C., and Lee, J. (2011). Converting carbohydrates extracted from marine algae into ethanol using various ethanolic *Escherichia coli* strains. Appl Biochem Biotechnol (Epub ahead of print)

Lin, Y., and Tanaka, S. (2006). Ethanol fermentation from biomass resources: current state and prospects. Appl Microbiol Biotechnol *69*, 627-642.

Linder, M., and Teeri, T.T. (1997). The roles and function of cellulose-binding domains. Journal of Biotechnology *57*, 15-28.

Lodolo, E.J., Kock, J.L., Axcell, B.C., and Brooks, M. (2008). The yeast *Saccharomyces cerevisiae*- the main character in beer brewing. FEMS Yeast Res *8*, 1018-1036.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev *66*, 506-577, table of contents.

Mach, R.L., and Zeilinger, S. (2003). Regulation of gene expression in industrial fungi: *Trichoderma*. Appl Microbiol Biotechnol *60*, 515-522.

Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., *et al.* (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol *26*, 553-560.

McBride, J.E., Zietsmann, J.J., Van Zyl, W.H., and Lynd, L.R. (2005). Utilization of cellobiose by recombinant β-glucosidase-expressing strains of *Saccharomyces cerevisiae*: characterization and evaluation of the sufficiency of expression Enzyme and Microbial Technology *37*, 93-103.

Miettinen-Oinonen, A. (2004). *Trichoderma reesei* strains for production of cellulases for the textile industry. In Biological and Environmental Sciences (Helsinki, University of Helsinki), p. 100.

Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., Shimonaga, T., Hattori, M., and Ashikari, T. (2009). Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res *16*, 115-129.

Okada, H., Sekiya, T., Yokoyama, K., Tohda, H., Kumagai, H., and Morikawa, Y. (1998). Efficient secretion of *Trichoderma reesei* cellobiohydrolase II in *Schizosaccharomyces pombe* and characterization of its products. Appl Microbiol Biotechnol *49*, 301-308.

Palamarczyk, C., Maras, M., Contreras, R., and Kruszewska, J. (1998). Protein secretion and glycosylation in *Trichoderma*. In Trichoderma & Glicoladium, C. Kubicek, and G. Harman, eds., pp. 121-138.

Penttila, M.E., Andre, L., Lehtovaara, P., Bailey, M., Teeri, T.T., and Knowles, J.K. (1988). Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. Gene *63*, 103-112.

Penttila, M.E., Andre, L., Saloheimo, M., Lehtovaara, P., and Knowles, J.K. (1987). Expression of two *Trichoderma reesei* endoglucanases in the yeast *Saccharomyces cerevisiae*. Yeast *3*, 175-185.

Rabinovich, M.L., Melnick, M.S., and Bolobova, A.V. (2002). The structure and mechanism of action of cellulolytic enzymes. Biochemistry (Mosc) *67*, 850-871.

Rainieri, S., Zambonelli, C., and Kaneko, Y. (2003). *Saccharomyces sensu stricto*: systematics, genetic diversity and evolution. J Biosci Bioeng *96*, 1-9.

Romanos, M.A., Scorer, C.A., and Clare, J.J. (1992). Foreign gene expression in yeast: a review. Yeast *8*, 423-488.

Runquist, D., Hahn-Hagerdal, B., and Radstrom, P. (2010). Comparison of heterologous xylose transporters in recombinant Saccharomyces cerevisiae. Biotechnol Biofuels *3*, 5.

Ryabova, O.B., Chmil, O.M., and Sibirny, A.A. (2003). Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast *Hansenula polymorpha*. FEMS Yeast Res *4*, 157-164.

Saerens, S.M., Duong, C.T., and Nevoigt, E. (2010). Genetic improvement of brewer's yeast: current state, perspectives and limits. Appl Microbiol Biotechnol *86*, 1195-1212.

Saloheimo, M., Nakari-Setala, T., Tenkanen, M., and Penttila, M. (1997). cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. Eur J Biochem *249*, 584-591.

Saloheimo, A., Rauta, J., Stasyk, O.V., Sibirny, A.A., Penttila, M., and Ruohonen, L. (2007). Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases. Appl Microbiol Biotechnol *74*, 1041-1052.

Saloheimo, M., Kuja-Panula, J., Ylosmaki, E., Ward, M., and Penttila, M. (2002). Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* beta-glucosidase BGLII (cel1A). Appl Environ Microbiol *68*, 4546-4553.

Sanchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. Biotechnol Adv *27*, 185-194.

Sedlak, M., and Ho, N.W. (2004). Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant *Saccharomyces* yeast. Yeast *21*, 671-684.

Sonan, G.K., Receveur-Brechot, V., Duez, C., Aghajari, N., Czjzek, M., Haser, R., and Gerday, C. (2007). The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. Biochem J *407*, 293-302.

Stahlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T.T., and Jones, T.A. (1996). Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from Trichoderma reesei. J Mol Biol *264*, 337-349.

Takashima, S., Iikura, H., Nakamura, A., Hidaka, M., Masaki, H., and Uozumi, T. (1998). Overproduction of recombinant *Trichoderma reesei* cellulases by Aspergillus oryzae and their enzymatic properties. J Biotechnol *65*, 163-171.

Teather, R.M., and Wood, P.J. (1982). Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol *43*, 777-780.

Thomas, C.F., Jr., Leof, E.B., and Limper, A.H. (1999). Analysis of *Pneumocystis carinii* introns. Infect Immun *67*, 6157-6160.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and 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.

Tonelli, D. (2009). Methods for determining ethanol in beer. In Beer in health and disease prevention, V.R. Preedy, ed. (Elsevier), pp. 1055-1066.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916-924.

Tsai, S.L., Oh, J., Singh, S., Chen, R., and Chen, W. (2009). Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. Appl Environ Microbiol *75*, 6087-6093.

Tuohy, M.G., Walsh, D.J., Murray, P.G., Claeyssens, M., Cuffe, M.M., Savage, A.V., and Coughlan, M.P. (2002). Kinetic parameters and mode of action of the

cellobiohydrolases produced by *Talaromyces emersonii*. Biochim Biophys Acta *1596*, 366-380.

Usher, J., and Bond, U. (2009). Recombination between homoeologous chromosomes of lager yeasts leads to loss of function of the hybrid GPH1 gene. Appl Environ Microbiol *75*, 4573-4579.

van Maris, A.J., Abbott, D.A., Bellissimi, E., van den Brink, J., Kuyper, M., Luttik, M.A., Wisselink, H.W., Scheffers, W.A., van Dijken, J.P., and Pronk, J.T. (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. Antonie Van Leeuwenhoek *90*, 391-418.

Van Rensburg, P., Van Zyl, W.H., and Pretorius, I.S. (1998). Engineering yeast for efficient cellulose degradation. Yeast *14*, 67-76.

van Rooyen, R., Hahn-Hagerdal, B., La Grange, D.C., and van Zyl, W.H. (2005). Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. J Biotechnol *120*, 284-295.

van Wyk, N., den Haan, R., and van Zyl, W.H. (2010). Heterologous co-production of *Thermobifida fusca* Cel9A with other cellulases in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol *87*, 1813-1820.

Vidgren, V., Huuskonen, A., Virtanen, H., Ruohonen, L., and Londesborough, J. (2009). Improved fermentation performance of a lager yeast after repair of its AGT1 maltose and maltotriose transporter genes. Appl Environ Microbiol *75*, 2333-2345.

Vidgren, V., Multanen, J.P., Ruohonen, L., and Londesborough, J. (2010). The temperature dependence of maltose transport in ale and lager strains of brewer's yeast. FEMS Yeast Res *10*, 402-411.

Vidgren, V., Ruohonen, L., and Londesborough, J. (2005). Characterization and functional analysis of the MAL and MPH Loci for maltose utilization in some ale and lager yeast strains. Appl Environ Microbiol *71*, 7846-7857.

Voutilainen, S.P., Boer, H., Linder, M.B., Puranen, T., Rouvinen, J., Vehmaanpera, J., and Koivula, A. (2007). Heterologous expression of Melanocarpus albomyces cellobiohydrolase Cel7B, and random mutagenesis to improve its thermostability. Enzyme and Microbial Technology *41*, 234-243.

Voutilainen, S.P., Murray, P.G., Tuohy, M.G., and Koivula, A. (2010). Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity. Protein Eng Des Sel *23*, 69-79.

Walfridsson, M., Hallborn, J., Penttila, M., Keranen, S., and Hahn-Hagerdal, B. (1995). Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase. Appl Environ Microbiol *61*, 4184-4190.

Walker, J.R.L. (1992). Sprectophotometric determination of enzyme activity-alcohol-dehydrogenase (ADH). Biochemical Education *20*, 42-43.

Wang, B., Wang, X., and Feng, H. (2010). Deconstructing recalcitrant Miscanthus with alkaline peroxide and electrolyzed water. Bioresour Technol *101*, 752-760.

Waters, D.M., Murray, P.G., Ryan, L.A., Arendt, E.K., and Tuohy, M.G. (2010). *Talaromyces emersonii* thermostable enzyme systems and their applications in wheat baking systems. J Agric Food Chem *58*, 7415-7422.

Weimer, P.J., French, A.D., and Calamari, T.A. (1991). Differential fermentation of cellulose allomorphs by ruminal cellulolytic bacteria. Appl Environ Microbiol *57*, 3101-3106.

Wen, F., Sun, J., and Zhao, H. (2010). Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. Appl Environ Microbiol *76*, 1251-1260.

Wendland, J. (2003). PCR-based methods facilitate targeted gene manipulations and cloning procedures. Curr Genet *44*, 115-123.

Wood, T.M., and Bhat, K.M. (1988). Methods for measuring cellulase activities. Methods in Enzymology *160*, 87-112.

Zaldivar, J., Nielsen, J., and Olsson, L. (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. Appl Microbiol Biotechnol *56*, 17-34.

Zhang, Y.H., and Lynd, L.R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng *88*, 797-824.

Zhang, Y.H.P., Cui, J.B., Lynd, L.R., and Kuang, L.R. (2006). A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: Evidence from enzymatic hydrolysis and supramolecular structure. Biomacromolecules 7, 644-648.