Co-expression of cellulase genes in *Saccharomyces cerevisiae* for cellulose degradation

Lisa du Plessis

Thesis presented in partial fulfillment of the
requirements for the degree of
Masters of Science
at Stellenbosch University

December 2008

Supervisor: Prof WH van Zyl
Co-supervisor: Dr SH Rose

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2008

SUMMARY

Complete degradation of cellulose produces mainly glucose, which can be fermented to ethanol. Therefore cellulose presents an abundant renewable energy resource for the production of an alternative, environmentally friendly, transportation fuel. Enzymatic degradation of cellulose is achieved by the synergistic action of three cellulase enzyme groups: endoglucanases, exoglucanases and β-glucosidases. However, cellulolytic organisms do not produce significant amounts of ethanol. Therefore, a need has arisen to develop a recombinant microorganism with the ability to produce cellulolytic enzymes, hydrolyze cellulose and ferment the resulting sugars to ethanol in a single process step, referred to as "Consolidated Bioprocessing" (CBP). This would provide a cost-effective, economically feasible strategy for the production of bioethanol.

The naturally fermentative yeast, *Saccharomyces cerevisiae,* is often used as host for the expression of recombinant proteins due to several characteristics, including its robustness in industrial processes, the well developed genetic tools available for manipulation and its proven safety status. A number of cellulase genes have previously been successfully expressed by recombinant *S. cerevisiae* strains. In this study, all three components of the cellulase system were co-expressed in *S. cerevisiae* to test the ability of the yeast to effectively produce the heterologous proteins, and consequently produce enough glucose for growth on an amorphous cellulosic substrate.

The *Trichoderma reesei* endoglucanase gene *egII* (*Cel5A*) was successfully expressed by a *S. cerevisiae* Y294 strain. Recombinant EGII displayed activities of 19.6 nkat.ml^{-1} and 22.3 nkat.ml^{-1} towards CMC and barley β-glucan, respectively. The major endoglucanase gene, *egI* (*Cel7B*) from *T. reesei* was subjected to random mutagenesis by propagating the *egI*-containing plasmid in an *E. coli* mismatch repair deficient strain. Screening of *S. cerevisiae* transformants revealed a strain, *S. cerevisiae* Y294[pLEM1], with improved levels of endoglucanase activity (21.8 nkat.ml^{-1}), compared to *S. cerevisiae* Y294[pAZ40], expressing the wild type gene (10.3 nkat.ml^{-1}). Through subcloning of the mutated *ENO1* promoter region and the mutated *egI* gene fragment, it was established that the mutations located in both the promoter- and gene sequences were responsible for the improved levels of activity displayed by *S. cerevisiae* Y294[pLEM1].

The *egII* gene and the altered *egI* gene were co-expressed with a codon optimised *T. reesei* cellobiohydrolase (sCBHI) and a β-glucosidase from *Saccharomycopsis fibuligera*. This resulted in a reduction in endoglucanase levels, possibly due to the metabolic burden placed on the yeast by co-expressing the different cellulases. The hydrolysis products produced by cellulase co-expressing strains were cellotriose, cellobiose and glucose, although the glucose yield was insufficient to enable growth on cellulose as sole carbon source. As the major hydrolysis product was cellobiose, it is likely that a bottleneck exists at its conversion to glucose, suggesting inadequate β-glucosidase activity.

This study has provided insight into co-expression of cellulase enzymes by the yeast *S. cerevisiae*. The knowledge obtained could be applied in optimizing cellulase cocktails for efficient cellulose degradation and eventual production of ethanol by recombinant yeast. It has also demonstrated the applicability of random mutagenesis for improving the activity of cellulases.

OPSOMMING

Die afbraak van sellulose produseer hoofsaaklik glukose, wat na etanol gefermenteer kan word. Dus bied sellulose ‘n oorvloedige hernubare energiebron vir die produksie van ‘n alternatiewe, omgewingsvriendelike, vervoerbrandstof. Ensiematiese afbraak van sellulose verg die sinergistiese werking van drie sellulase ensiemgroepe: endoglukanases, eksoglukanases en β-glukosidases. Sellulolitiese organismes produseer egter nie betekenisvolle hoeveelhede etanol nie. Dus het ‘n behoefte ontstaan om ‘n rekombinante mikroorganisme te ontwikkel met die vermoë om sellulolitiese ensieme te produseer, sellulose te hidroliseer en die resulterende suikers na etanol te fermenteer in een stap, bekend as “Gekonsolideerde Bioprosessering” (GBP). Dit bied ‘n koste-effektiewe en ekonomies uitvoerbare strategie vir die produksie van bioetanol.

Die fermenterende gis, *Saccharomyces cerevisiae,* word gereeld as gasheer vir die uitdrukking van rekombinante proteïne gebruik as gevolg van verskeie kenmerke, insluitend sy robuustheid in industriële prosesse, die goed ontwikkelde genetiese hulpmiddels beskikbaar vir manipulering en sy beproefde veiligheidsstatus. ‘n Aantal sellulase gene is reeds suksesvol uitgedruk deur rekombinante *S. cerevisiae* rasse. In hierdie studie is al drie komponente van die sellulase-sisteem gesamentlik in *S. cerevisiae* uitgedruk om die vermoë van die gis te toets om heteroloë proteïne te produseer, en genoegsame glukose te vervaardig vir groei op ‘n amorfe sellulolitiese substraat.

Die *Trichoderma reesei* endoglukanase geen *egII* (*Cel5A*) is suksesvol deur *S. cerevisiae* Y294 uitgedruk. Die rekombinante EGII ensiem het aktiwiteitsvlakke van 19.6 nkat.ml^{-1} en 22.3 nkat.ml^{-1} getoon teenoor CMC en gars β-glukaan, onderskeidelik. Die endoglukanase geen, *egI* (*Cel7B*) van *T. reesei* is blootgestel aan lukrake mutagenese tydens vermeerdering van die *egI*-bevattende plasmied in ‘n *E. coli* ras wat nie repliseringsfoute kan herstel nie. Na sifting van *S. cerevisiae* transformante is ‘n ras, *S. cerevisiae* Y294[pLEM1], geïdentifiseer wat beskik oor verbeterde vlakke van endoglukanase aktiwiteit (21.8 nkat.ml^{-1}), in vergelyking met die ras wat die wildetipe geen uitdruk, *S. cerevisiae* Y294[pAZ40] (10.3 nkat.ml^{-1}). Deur subklonering van die gemuteerde *ENO1* promoter area en die gemuteerde *egI* geenfragment, is vasgestel dat mutasies teenwoordig in beide die promoter- en geen

volgordes verantwoordelik was vir die hoër vlakke van aktiwiteit geproduseer deur *S. cerevisiae* Y294[pLEM1].

Die *egII* geen en gemuteerde *egI* geen is gesamentlik met ‘n kodon geoptimiseerde *T. reesei* sellobiohidrolase (sCBHI) en ‘n β-glukosidase van *Saccharomycopsis fibuligera* uitgedruk. Dit ‘n verlaging in endoglukanase vlakke teweeg gebring, waarskynlik as gevolg van ‘n metaboliese las wat op die gis geplaas is vanweë ko-uitdrukking van die verskillende sellulases. Die afbraakprodukte geproduseer deur gesamentlike uitdrukking van die sellulases was sellotriose, sellobiose and glukose, alhoewel die glukose opbrengs onvoldoende was om groei op sellulose as enigste koolstofbron, te onderhou. Aangesien die hoof afbraakproduk sellobiose was, is dit moontlik dat ‘n bottelnek by die omskakeling na glukose bestaan. Dit dui daarop dat die β-glukosidase aktiwiteit onvoldoende was.

Hierdie studie het insig gebied tot die gesamentlike uitdrukking van sellulase ensieme deur die gis *S. cerevisiae*. Die kennis kan toegepas word in die optimisering van sellulase mengsels vir effektiewe sellulose afbraak en uiteindelike produksie van etanol deur ‘n rekombinante gis. Dit het ook die toepasbaarheid van lukrake mutagenese vir die verbetering van aktiwiteit van sellulases gedemonstreer.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions for their invaluable contributions to the successful completion of this study:

Prof. W.H. van Zyl, Department of Microbiology, University of Stellenbosch, who acted as my supervisor, for accepting me as a student and for his constant enthusiasm and encouragement;

Dr. Shaunita Rose, Department of Microbiology, University of Stellenbosch, who acted as my co-supervisor, for her guidance, patience, and for believing in me;

My co-workers in Lab A335, for their valuable suggestions, encouragement and for creating a pleasant work environment;

The Staff of the Department of Microbiology, in particular Annetjie Hugo, for their assistance;

The National Research Foundation, Harry Crossley Foundation and Ernst and Ethel Eriksen Trust, for financial support;

My friends, especially Laurianne van der Merwe, Nicolette Fouché, Niël van Wyk, Colin Ohlhoff and Megan Kotzee for their constant encouragement and understanding, and above all, their friendship;

Gerrit Botha, for his emotional support, love and pride in me;

My parents, Cas and Maryna du Plessis, to whom this thesis is dedicated. Thank you for financial support, encouragement and love, and for believing in my abilities.

INDEX

CHAPTER 3:

Construction of *Saccharomyces cerevisiae* strains co-expressing cellulase genes for efficient hydrolysis of amorphous cellulose

CHAPTER 4:

GENERAL DISCUSSION AND CONCLUSIONS

CHAPTER 1

GENERAL INTRODUCTION AND PROJECT AIMS

1. INTRODUCTION

Cellulose is the most abundant polymer on earth, with the majority thereof found in the cell walls of plants (Saloheimo *et al.* 1994). Cellulose is a linear homopolymer consisting of repetitive D-glucose molecules linked by β-1,4-glycosidic bonds. The cellulose chains are tightly packed together to form crystalline areas, with interspersed amorphous regions (Lynd *et al.* 2002). Together with hemicellulose and lignin, cellulose maintains structural integrity in the plant cell walls (Klemm *et al.* 2005). This association, as well as the crystalline nature of cellulose, renders it inaccessible and recalcitrant to hydrolysis by cellulolytic enzymes (Van Rensburg *et al.* 1998). Cellulose has enormous potential as a renewable energy source for bioethanol production, hence the focus on its cost-effective conversion.

Trichoderma reesei is one of the most studied cellulolytic organisms. This filamentous fungus produces all of the cellulase enzymes required for the hydrolysis of cellulose to glucose (Saloheimo *et al.* 1994). Endoglucanases randomly hydrolyze glycosidic bonds in the amorphous regions of the cellulose chains. Free chain ends are generated, which are hydrolyzed in a processive manner by cellobiohydrolases from the reducing or non-reducing ends. Finally, the resulting cellobiose and soluble cellooligosaccharides are cleaved by β-glucosidases to generate glucose as main end product (Fujita *et al.* 2004). *T. reesei* is an excellent cellulase producer, but does not have the fermentation capability to produce sufficient amounts of ethanol. This problem could be overcome by expressing its cellulase genes in a naturally fermentative organism.

Saccharomyces cerevisiae is a popular host for such heterologous expression studies mainly due to its safety of use (GRAS status) and ease with which it can be genetically manipulated (Ostergaard *et al.* 2000). The development of a *S. cerevisiae* strain with the ability to utilize and ferment abundant, inexpensive, natural substrates to ethanol in one step (referred to as consolidated bioprocessing, CBP), could provide an economically attractive alternative to the use of environmentally harmful energy sources such as fossil fuels (Den Haan *et al.* 2007).

The synergistic activity of endoglucanases, cellobiohydrolases and β-glucosidases is necessary for effective conversion of cellulose to its constituent glucose molecules (Den Haan *et al.* 2007). Recently, it has been demonstrated that concurrent expression of these enzymes resulted in efficient degradation of cellulose and production of ethanol. This was achieved by pre-cultivation of the cellulase-expressing

yeast cells prior to contact with cellulose (Fujita *et al.* 2004, 2002). However, growth on cellulose and subsequent production of ethanol, by a yeast strain co-expressing an endoglucanase and β-glucosidase, has only been demonstrated in 2007 (Den Haan *et al.* 2007). Further studies are being conducted to optimize co-synergistic expression, using different cellulase combinations or expression systems, ultimately aiming to achieve large-scale, cost-effective bioethanol production via CBP.

2. AIMS OF THE STUDY

The objective of this study was the co-expression of the three cellulase enzymes (endoglucanase, cellobiohydrolase and β-glucosidase) required for efficient degradation of cellulose using *S. cerevisiae* as host. This study was performed in order to optimize cellulase combination ratios and gain more insight into synergistic expression to aid in future development of CBP technology.

The specific aims of the present study were as follows:

(i) The functional expression of the *T. reesei* endoglucanases (*egI* and *egII*) in the yeast *S. cerevisiae*;

(ii) The improvement of secreted EGI activity, by means of random mutagenesis of the plasmid bearing the *egI* gene;

(iii) The characterization of the improved EGI regarding pH and temperature optimum profiles;

(iv) Establishing the location of putative mutations (on the plasmid) contributing to the improved levels of EGI activity;

(v) The identification of point mutations responsible for improved activity via automated sequencing;

(vi) Co-expression of the *egI* and *egII* genes with the *sCBHI* (of *T. reesei*, synthetic, codon optimized cellobiohydrolase) and *bgl1* (β-glucosidase of *Saccharomycopsis fibuligera*) in different combinations;

(vii) Determination of hydrolysis products formed from phosphoric acid swollen cellulose (PASC) by yeast strains co-expressing cellulase genes;

(viii) Test the cellulase co-expressing strains for growth on amorphous cellulose as sole carbon source.

3. REFERENCES

Den Haan R, Rose SH, Lynd LR, Van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae.* Metabolic Engineering 9: 87-94

Fujita Y, Ito J, Ueda M, Fukuda H, 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. Applied and Environmental Microbiology 70: 1207-1212

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

Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. Angewandte Chemie International Edition 44: 3358-3393

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002*)* Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews 66: 506-577

Saloheimo A, Henrissat B, Hoffrén A-M, Teleman O, Penttilä M (1994) A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. Molecular Microbiology 13: 219-228

Van Rensburg P, Van Zyl WH, Pretorius IS (1998) Engineering yeast for efficient cellulose degradation. Yeast 14: 67-76

Ostergaard S, Olsson L, Nielsen J (2000) Metabolic Engineering of *Saccharomyces cerevisiae.* Microbiology and Molecular Biology Reviews 64: 34-50

CHAPTER 2
REVIEW OF LITERATURE

1. INTRODUCTION

Fossil fuels are currently used as the world's main energy source with consumption increasing due to the growing population demand (Sun and Cheng 2002). The rate at which fossil resources are being exploited has ceased to match the global rate of fuel consumption. This scenario has led to increased fuel costs and inevitably will lead to the depletion of the world's energy supply (Lin and Tanaka 2006, Van Maris *et al*. 2006). Furthermore, the combustion of fossil fuels causes the emission of large quantities of carbon dioxide into the atmosphere (Van Wyk 2001). This, as well as the release of other greenhouse gases results in global climate changes. Therefore, from an economical and environmental point of view, a transition needs to be made from non-renewable to renewable energy sources (Ragauskas *et al*. 2006). Research is currently focusing on exploring the feasibility of alternative energy resources. The criteria required for such an alternative fuel resource includes its availability in large quantities, environmental benefits and above all, lower production cost compared to fossil fuels (Hill *et al.* 2006). Plant biomass seems to offer an attractive low-cost, abundant energy source which will be discussed briefly (Lynd *et al*. 2005).

1.1 Ethanol as alternative energy source

Energy used for transportation is estimated at 27% of primary energy used worldwide, with the majority attributed to road transport (Antoni *et al.* 2007). Fuel derived from biomass contributes to about 10% of the world's energy supply. Bioethanol and biodiesel are produced on an industrial scale from sucrose, starches and plant oils, all which are readily processable (Antoni *et al.* 2007, Pu *et al.* 2007). Other biofuels include biomethane, biobutanol, biohydrogen and biomethanol. Ethanol (as a fuel for transportation) offers high octane and high heat of vaporization, resulting in greater energy output and improved net performance (Zaldivar *et al*. 2001, Aristidou and Penttilä 2000). Its combustion generates a low net emission of carbon dioxide, non-combusted hydrocarbons, carbon monoxide, nitrogen oxides and volatile organic compounds (Galbe and Zacchi 2002). However, transportation from production centers to destinations via pipeline poses problems such as the tendency of ethanol to absorb water (www.agmrc.org>/NR/rdonlyres/4EE0E81C-C607-4C3F-BBCF-B75B7395C881/0/ksupipelineethl.pdf). The presence of water in gasoline/ethanol blends reduces engine performance. All catalyst manufactured vehicles are able to use 10% (E10) to 20% (E20) ethanol blends with gasoline, whereas flexible fuel vehicles are able to run on mixtures of up to 85% (E85) ethanol. Brazil and the

USA are the largest bioethanol producers, together contributing to 72.6% of the total bioethanol produced worldwide (48.7 x $10^6 m^3$/annum) in 2006 (Table 1) (Antoni *et al.* 2007).

Table 1: Countries producing the highest levels of bioethanol, ranked according to production levels in 2004 (Antoni *et al.* 2007).

	2004 ($10^6 m^3$)	2005 ($10^6 m^3$)	2006 ($10^6 m^3$)
Brazil	15.09	15.99	16.99
USA	13.37	16.13	18.37
China	3.65	3.80	3.85
India	1.75	1.70	1.90
France	0.83	0.91	0.95
Russia	0.75	0.75	0.65
South Africa	0.42	0.39	0.39
UK	0.40	0.35	0.28
Saudi Arabia	0.30	0.12	0.20
Spain	0.30	0.35	0.46
Thailand	0.28	0.30	0.35
Germany	0.27	0.43	0.76
Others	3.34	4.75	3.55
Total	40.75	45.97	48.70

1.2 Consolidated bioprocessing (CBP)

Microbial conversion of biomass to ethanol is achieved in four steps: the production of enzymes, the hydrolysis of plant material by these enzymes, the fermentation of hexose sugars (glucose, mannose, galactose) and the fermentation of pentose sugars (xylose, arabinose) (Lynd *et al.* 2002, Lynd 1996). The degree to which these steps are integrated determine the four different processing strategies as indicated in Figure 1. No integration occurs in separate hydrolysis and fermentation (SHF), implying the use of four bioreactors. Cellulose hydrolysis products are concurrently fermented to ethanol (upon their release from the cellulose chain) in a second process called simultaneous saccharification and fermentation (SSF) (Sun and Cheng 2002). This prevents accumulation of simple sugars and prevents end product inhibition. Since hydrolysis and hexose (C_6) fermentation are consolidated in SSF, three reactors are needed, whereas only two bioreactors are used when pentose (C_5) sugars, resulting from

hemicellulose hydrolysis, are also fermented in the same reactor (Lynd 1996). In the latter case, the process is referred to as simultaneous saccharification and cofermentation (SSCF). Ultimately, time is reduced and less reactor volume is required (Sun and Cheng 2002). An ideal solution to the development of a cost-effective process would be the use of a cellulolytic microorganism with the ability to produce cellulolytic enzymes, hydrolyze biomass and ferment the resulting sugars to ethanol in a single process step, referred to as "Consolidated bioprocessing" (CBP) (Lynd *et al.* 2002, Lynd 1996). Although many microorganisms possess some of the abilities, no single microorganism exist naturally with all of the above properties (Van Zyl *et al.* 2007).

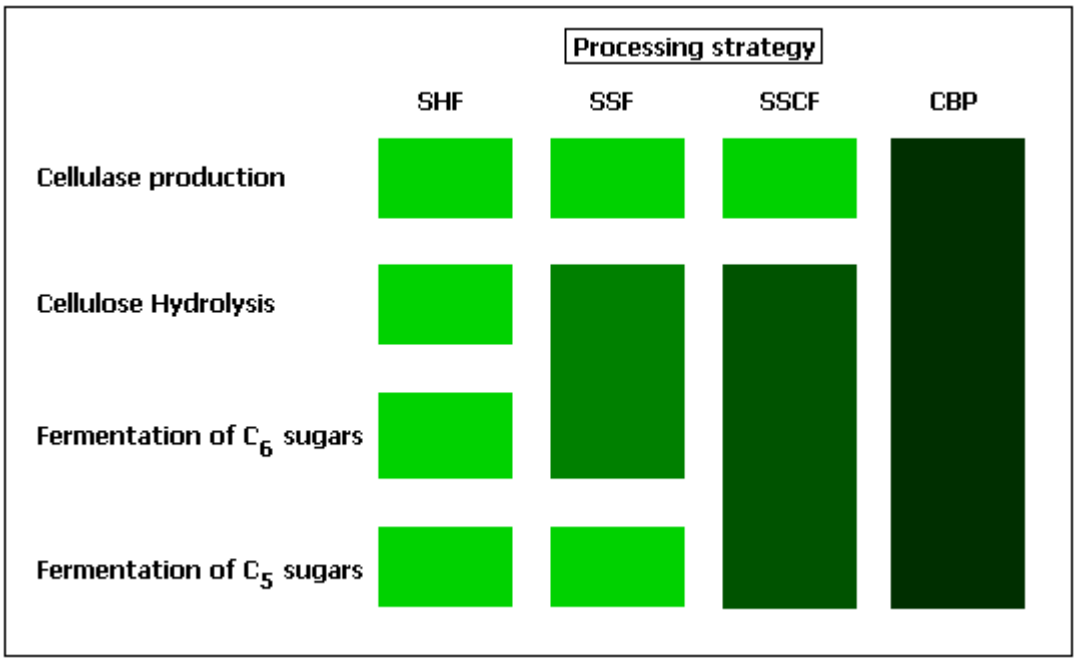

Figure 1: The four process strategies and their degree of consolidation in converting cellulosic biomass to ethanol: separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), simultaneous saccharification and cofermentation (SSCF) and consolidated bioprocessing (CBP) (Adapted from Lynd 1996).

The development of a CBP-enabling microorganism can be achieved by either one of two strategies; the native strategy or the recombinant strategy (Lynd *et al*. 2005). According to the native strategy, naturally cellulolytic microorganisms can be engineered to improve product yields. The feasibility of this strategy depends on the tolerance of the engineered organism to high levels of the end product, in this case, ethanol. Alternatively, using the recombinant strategy, non-cellulolytic organisms exhibiting high product yields can be engineered to utilize and ferment cellulosic biomass by means of a heterologous cellulase system.

2. LIGNOCELLULOSE

Plant biomass is the most abundant renewable organic compound on earth. Lignocellulose can be derived from wood, grass, agricultural residues, forestry waste and municipal solid wastes (Galbe and Zacchi 2002). In nature, this material is biologically degraded to hummus, water and carbon dioxide. Due to its abundance and low cost, it can be utilized on a large scale to inexpensively produce other products (Eriksson *et al.*1990). Lignocellulosic biomass is therefore an economically attractive source for the production of renewable energy such as ethanol (Pérez *et al.* 2002). The carbon dioxide released during fermentation of these substrates and the burning of ethanol as fuel is reutilized through photosynthesis to produce new biomass, thereby completing the carbon cycle (Van Wyk 2001). Issues raised regarding starch production for biofuel as opposed to food, is circumvented by rather targeting non-food substrates such as cellulose-containing waste material (Antoni *et al.* 2007).

Lignocellulose is the major constituent of plant biomass and acts mainly as a structural component (Howard *et al.* 2003, Pérez *et al.* 2002). It is an intermeshed complex consisting of lignin (15-20%), hemicellulose (25-35%) and cellulose (40-50%), with the composition of the three polymers varying depending on the source (Gray *et al.* 2006, Howard *et al.* 2003, Pérez *et al.* 2002). The typical composition in agricultural residues and wastes are illustrated in Table 2.

Table 2: Content of various lignocellulosic materials (Adapted from Howard *et al.* 2003).

Lignocellulosic materials	Cellulose (%)	Hemicellulose (%)	Lignin (%)
Hardwood stems	40-55	24-40	18-25
Softwood stems	45-50	25-35	25-35
Nut shells	25-30	25-30	30-40
Corn cobs	45	35	15
Paper	85-99	0	0-15
Wheat straw	30	50	15
Rice straw	32.1	24	18
Sorted refuse	60	20	20
Leaves	15-20	80-85	0

Table 2: Content of various lignocellulosic materials (Adapted from Howard *et al.* 2003) (continue).

Cotton seed hairs	80-95	5-20	0
Newspaper	40-55	25-40	18-30
Waste from chemical pulps	60-70	10-20	5-10
Fresh bagasse	33.4	30	18.9
Solid cattle manure	1.6-4.7	1.3-3.3	2.7-5.7
Switchgrass	45	31.4	12.0
Grasses (average for all grasses)	25-40	25-50	10-30

2.1 Cellulose

In 1838 the French scientist Anselme Payen recognized the major component of wood to be a fibrous structure with a molecular formula of $C_6H_{10}O_5$; currently known as cellulose. Cellulose is the most commonly produced organic polysaccharide with approximately 1.5×10^{12} tons being generated annually through photosynthesis (Klemm *et al.* 2005). This structural polymer is found almost exclusively in the cell walls of plants, but is also produced by certain bacteria, algae, fungi and some animals (e.g. tunicates) (Klemm *et al.* 2005, Lynd *et al.* 2002).

Cellulose is a homopolysaccharide, consisting of insoluble, linear chains of β-1,4-linked β-D-glucopyranose units (Howard *et al.* 2003, Klemm *at al.* 2002, Sjöström 1993). The chain length of cellulose is expressed as the number of constituent glucose molecules, referred to as the degree of polymerization (DP) (Klemm *et al.* 2005). The purest form of cellulose is found in cotton, where it has a DP of approximately 15 000 units (O'Sullivan 1997). In wood, the DP is about 10 000 glucose units per cellulose chain. The cellulose chains undergo self-assembly into larger units known as protofibrils with a lateral dimension of 1.5 to 3.5 nm (Figure 2) (Klemm *et al.* 2005, Lynd *et al.* 2002). Protofibrils are packed into microfibrils (10-30 nm), which are linked by hydrogen bonds and Van der Waals forces to constitute the cellulose fiber (about 100 nm) (Klemm *et al.* 2005, Lynd *et al.* 2002, Pérez *et al.* 2002).

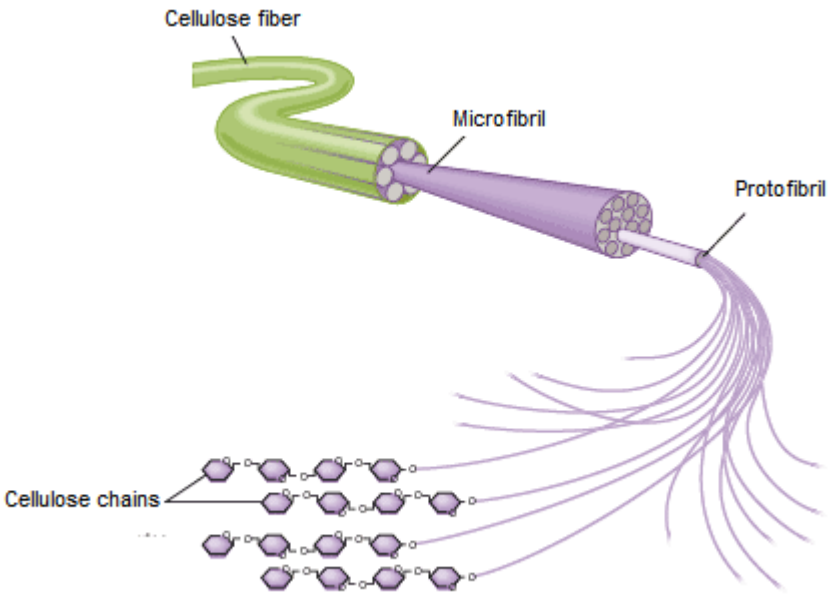

Figure 2: The structure of cellulose (www.nutrition.jbpub.com/resources/chemistryreview9.cfm)

The secondary cell wall of plants contains most of the cellulose mass. Here the microfibrils run parallel, giving it a densely packed, ordered arrangement. In the three layers of the secondary wall (S1, S2, and S3) the orientation of these groups of microfibrils differs with respect to the axis of the cell (Figure 3) (Kirk and Cullen 1998). In the primary cell wall, however, the chains run in all directions within the plane of the cell wall (Klemm *et al*. 2002, O'Sullivan 1997). Here non-cellulosic components such as hemicellulose, proteins and pectins dominate, decreasing the mechanical stability of the cell.

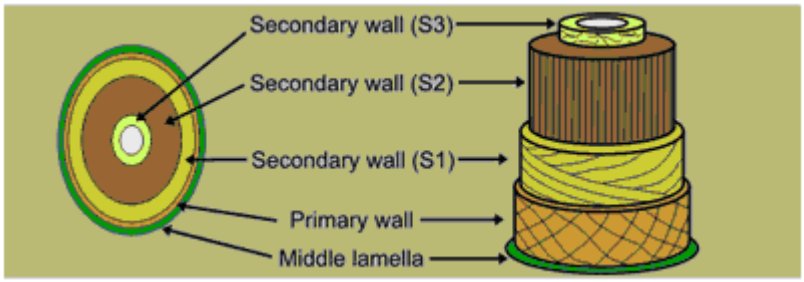

Figure 3: The cell wall of woody plants consist of a primary layer (P) and a secondary layer (S), which in turn consists of three sublayers (S1, S2 and S3). The cells are separated by a middle lamella (ML) (Kirk and Cullen 1998, www.ccrc.uga.edu/~mao/intro/ouline.htm).

The cellulose chain has a D-glucose molecule with a C_4-OH group at the non-reducing end, whereas at the reducing end it has a C_1-OH group (Klemm *et al.* 2005). In the $^4C_1$ conformation, the D-glucose molecule has three equatorial positioned hydroxyl groups and three axial hydrogen atoms (Klemm *et al.* 2005, 2002). Every second glucose ring is rotated at 180° in the plane, defining structural repeating units known as cellobiose (Figure 4). These glucose rings form sheets lying in the plane, which are stacked on top of each other to form a cellulose molecule (Zhang and Lynd 2004). The atoms in the cellulose chain are fixed in discreet positions with respect to one another by hydrogen bonding and Van der Waals interactions. This results in a highly crystalline structure which renders it recalcitrant to enzymatic degradation (Zhang and Lynd 2004, Lynd *et al.* 2002, Van Rensberg *et al.* 1998). However, a lateral order of distribution exists, where fibers with varying degrees of crystallinity are separated by amorphous regions (Lynd *et al.* 2002). This, as well as various types of irregularities in the microfibrils, makes it partially accessible to large molecules such as cellulolytic enzymes.

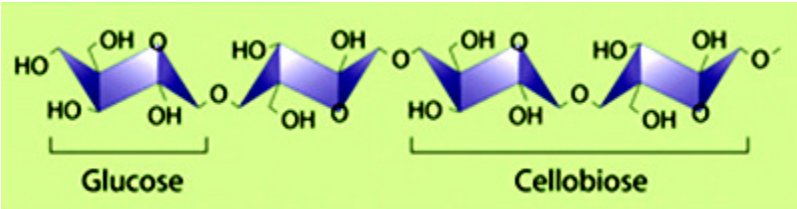

Figure 4: The cellulose chain consists of glucose units linked by β-1,4-glycosidic bonds. Two adjacent units define a cellobiose molecule (Pérez *et al.* 2002).

Six polymorphs of cellulose exist (I, II, III_1, III_{11}, IV_1, IV_{11}), although only cellulose I is found in nature (O'Sullivan 1997). Cellulose I has been found to exist in two crystalline forms, cellulose I_α and cellulose I_β (Klemm *et al*. 2002, O'Sullivan 1997). The I_β structure takes the form of a two chain unit cell, with the chains running parallel, whereas the I_α form consists of only one cellulose chain per unit cell. A unit cell is generally defined as the smallest group of atoms or molecules whose repetition at regular intervals in three dimensions produces the lattices of a crystal (www.die.net). The I_α/I_β ratio varies depending on the cellulose source, with the I_β form dominating in cotton, wood and ramie fibers and the I_α form, in bacterial cellulose and the cell walls of some algae (Pu *et al*. 2007, Klemm *et al.* 2005, 2002, Sjöström 1993).

Cellulose II, the second most investigated form of cellulose, may be formed from cellulose I by regeneration (solubilization and recrystallization) or mercerization (swelling of fibers in an alkali treatment) (O'Sullivan 1997). The arrangement of the cellulose chains in the unit cell of cellulose II differs from that of cellulose I, running antiparallel to one another. Celluloses III_1 and III_{11} are formed reversibly from celluloses I and II, respectively, using liquid ammonia treatment. The polymorphs IV_1 and IV_{11} originate from heating of celluloses III_1 and III_{11}, respectively, in glycerol.

2.2 Hemicellulose

Hemicellulose, the second most abundant polymer after cellulose, is located in the spaces between cellulose microfibrils in the primary and secondary walls (Sjöström 1993). Initially, this polysaccharide was wrongfully labeled as an intermediate in cellulose biosynthesis, hence the nomenclature: hemicellulose (Eriksson *et al.* 1990). It is however now acknowledged as a separate group of plant polysaccharides. Hemicelluloses differ from celluloses in that it has side groups linked to the backbones, making this heteropolysaccharide less crystalline. It also has a lower molecular weight than cellulose, with 100 to 150 sugar residues per chain (Pérez *et al.* 2002, Kirk and Cullen 1998).

The side groups on the backbone of hemicelluloses consist of sugars (such as D-xylose, L-arabinose, D-mannose, D-glucose and D-galactose) as well as sugar acids (including D-glucuronic acid and D-galacturonic acid) and acetyl esters (Howard *et al.* 2003, Kirk and Cullen 1998, Eriksson *et al.* 1990). The composition of hemicelluloses differs in softwood and hardwood. Glucuronoxylan dominate in hardwood hemicellulose, and as the name implies, is xylose based. The backbone consists of 1,4-linked β-D-xylopyranose units which form twisted ribbons (Warren 1996). Every ten xylose residues contain about seven *O*-acetyl groups at the C-2 or C-3 position, while one 4-*O*-methyl-α-D-glucuronic acid residue per ten xylose units is 1,2-linked to the backbone (Sjöström 1993, Eriksson *et al.* 1990). Glucomannan, present in lower quantities in hardwood, is composed of 1,4-linked β-D-xylopyranose and β-D-mannopyranose units (Warren 1996). These molecules form flat extended ribbons (Figure 5).

Galactoglucomannan, the major polysaccharide found in softwood hemicellulose, is composed of 1,4-linked β-D-glucopyranose and β-D-mannopyranose units (Sjöström 1993, Eriksson *et al.* 1990). Approximately one *O*-acetyl group substitution per 3-4 hexose units are found at the C-2 and C-3

positions. Arabinoglucuronoxylan is another hemicellulose found in softwoods. The 1,4-linked β-D-xylopyranose units in the backbone have 4-*O*-methyl-α-D-glucuronic acid groups at the C-2 position, about two groups per ten xylose units. On average, 1.3 α-L-arabinopyranose substitutions are found per ten xylose units (Figure 5).

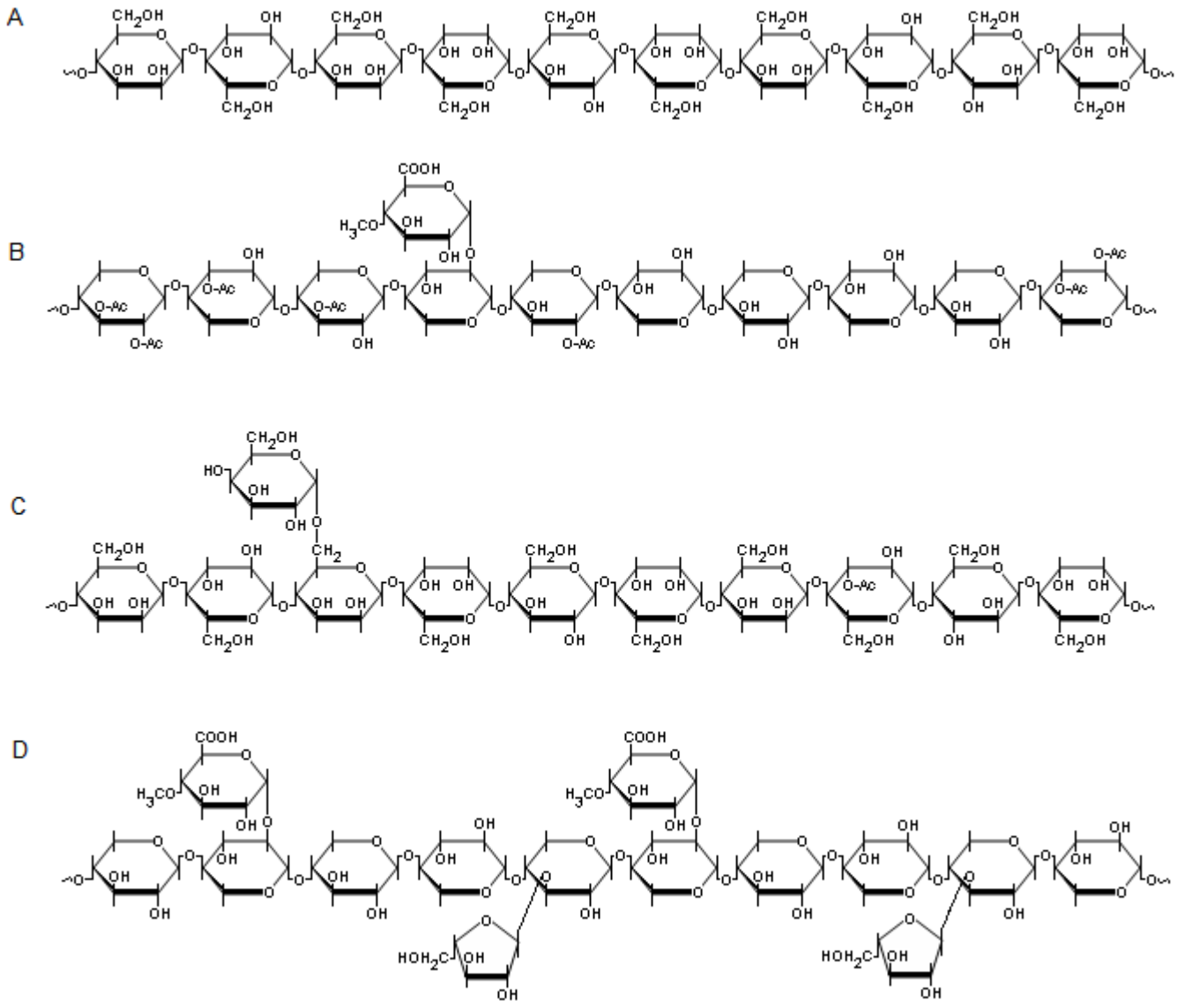

Figure 5: Structure of (A) glucuronoxylan and (B) glucomannan from hardwood and (C) galactoglucomannan and (D) arabinoglucuronoxylan from softwood (Coughlan and Hazlewood 1993).

Arabinogalactan is a minor hemicellulose component found in both hardwoods and softwoods. Unlike the other hemicellulose components, this polysaccharide has a 1,3 linked backbone, consisting of β-D-galactopyranose units. Substitutions are found at the C-6 position of almost every unit, and consist mainly of galactopyranose residues and some L-arabinose units.

2.3 Lignin

The complex chemical structure of lignin was resolved in the late 1960's and its name was derived from the Latin word *lignum*, meaning wood (Sjöström *et al.* 1993, Kirk *et al.* 1980). Lignin is a non-carbohydrate component of plant biomass (Sandgren *et al.* 2005). The covalent linkages (such as benzyl ester, benzyl ether and glycosidic linkages) between lignin and other carbohydrates results in the formation of lignin-carbohydrate complexes (LCC) (Pu *et al.* 2007). Lignin thus binds the other wood polysaccharides together providing mechanical strength as well as resistance to microbial attack (Sandgren *et al.* 2005). Lignin represents between 15% and 36% of the lignocellulosic material in wood (Eriksson *et al*. 1990). It is mainly concentrated in the middle lamella of wood cells, but is also located in the secondary walls (Sjöström *et al.* 1993).

Lignins are complex aromatic heteropolymers consisting of phenylpropane units with different side chains (Hon 1995). The lignin structure is formed by the polymerization of three types of monomer units: coniferyl alcohol, sinapyl alcohol and *p*-coumaryl alcohol (Figure 6) (Kirk *et al.* 1980). These units form functional groups such as phenolics, methoxyls, alcoholic hydroxyl groups and carbonyl groups (Sjöström *et al.* 1993). Most of the phenolic hydroxyl groups are linked to adjacent phenylpropane units. Softwoods lignins consist of mostly polymerized coniferyl alcohol (guaiacyl lignin) and some *p*-coumaryl alcohol (Eriksson *et al.* 1990). Hardwoods contain both coniferyl and sinapyl- (syringyl lignin) alcohol in a ratio varying from 4:1 to 1:2, as well as smaller amounts of *p*-coumaryl alcohol (Hon 1995, Sjöström *et al.* 1993). Some alcoholic hydroxyl groups are linked with *p*-hydroxybenzoic acid or *p*-hydroxycinnamic acid, depending on the wood source. In addition to carbon-carbon (C-C) linkages, units are also held together by ether (C-O-C) linkages, which make lignin resistant to chemical degradation. These linkages are especially plentiful between syringyl units in hardwood lignin.

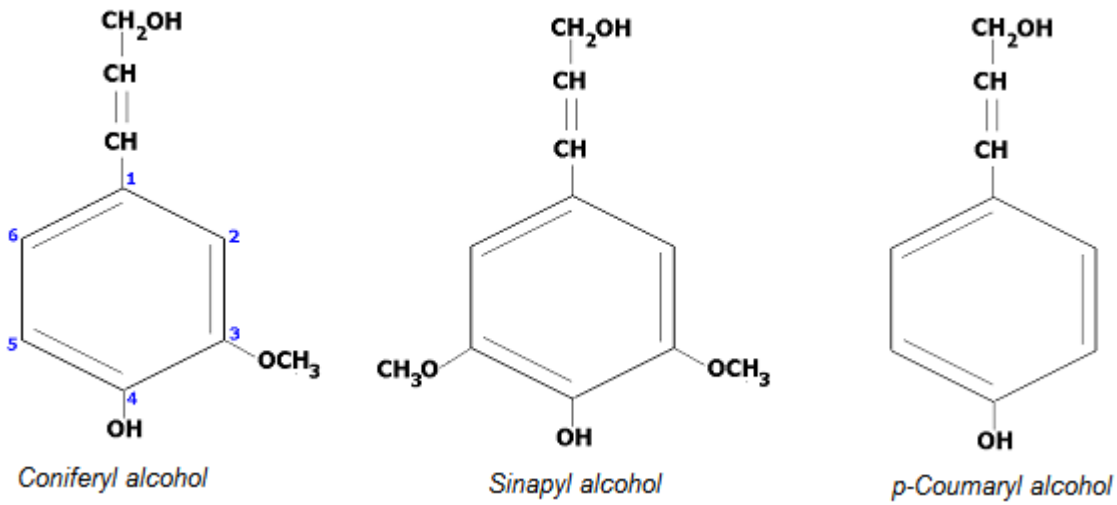

Figure 6: The three precursor monomeric units of lignin (Kirk *et al.* 1980).

3. MICROBIAL CELLULOSE DEGADATION

Cellulose is degraded by enzymes belonging to the glycosyl hydrolase families, which hydrolyze oligosaccharides and polysaccharides (Bayer *et al.* 1998). Microorganisms degrade polysaccharides by the concerted action of multiple enzymes in a system (Warren 1996). The more polysaccharides in the substrate, such as the plant cell wall, the more components the system requires.

Cellulase enzyme systems consist of three main components: endoglucanases (endo-1,4-β-D-glucanases; EC 3.2.1.4), exoglucanases (exo-1,4-β-D-glucanases; EC 3.2.1.91) and β-glucosidases (1,4-β-D-glucosidase; EC 3.2.1.21) (Sandgren *et al.* 2005, Takashima *et al.* 1999). Cellulases cleave β-1,4-glycosidic bonds between glucosyl residues by acid-catalyzed hydrolysis (Teeri 1997). Endoglucanases cut the cellulose chain randomly within internal amorphous sites, generating oligosaccharides of various lengths (Figure 7). New chain ends are generated which act as substrate for exoglucanases. Exoglucanases degrade the oligosaccharides from the reducing- or non-reducing ends leaving cellobiose as the end product. β-glucosidases hydrolyze soluble cellodextrins and cellobiose, generating glucose as the main end product (Lynd *et al.* 2002).

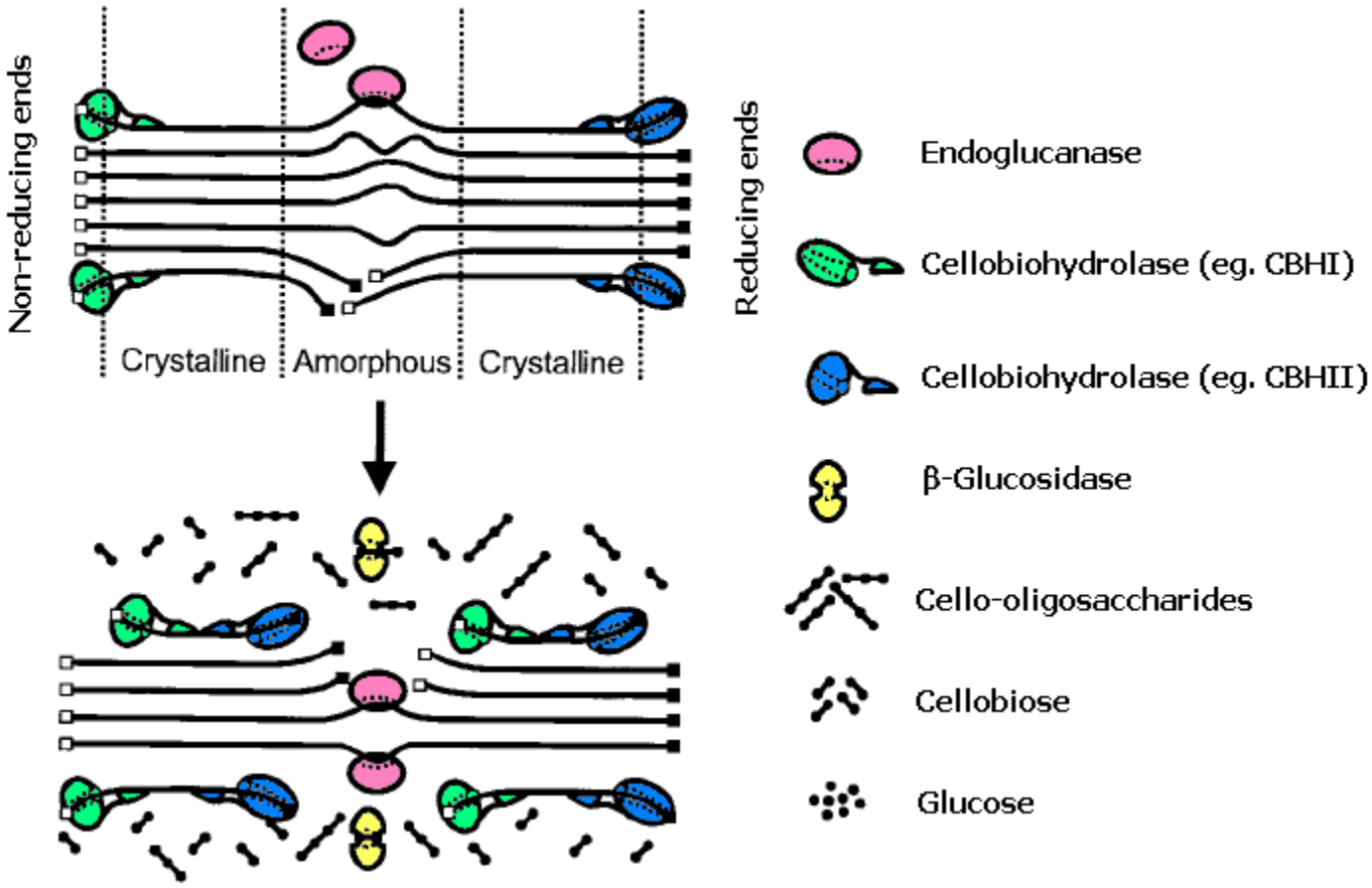

Figure 7: Schematic representation indicating the mode of action of the cellulase enzymes of non-complexed cellulase systems in the hydrolysis of amorphous and crystalline cellulose (Lynd *et al.* 2002).

Most cellulases are modular structures containing catalytic and carbohydrate-binding modules (CBMs) (previously referred to as cellulose binding domains, CBDs), connected by flexible serine-rich linker regions. However, some cellulases exist which only contain a catalytic domain (Lynd *et al.* 2002, Warren 1996). The CBM of cellulases is important for binding to the cellulose substrate while the catalytic domain cleaves the glycosidic bonds of the cellulose chain. Therefore, the CBM is responsible for bringing the catalytic core domain in close proximity of the cellulose, as well as to weaken the hydrogen bonds of the crystalline structure through the binding of amino acids (such as tyrosine and tryptophan) to the cellulose. The importance of cellulose binding for hydrolis has been demonstrated by replacing the CBM of endoglucanase II (EGII) by three tandemly aligned CBMs (CBM_{CBHII}-CBM_{CBHI}-CBM_{EGII}–EGIIcat). This resulted in increased affinity and consequently higher hydrolytic activity towards phosphoric acid swollen Avicel (PASC) (Ito *et al.* 2004).

Fungi and bacteria are the most important cellulose degraders in nature (Gruno *et al.* 2003). Anaerobic cellulolytic bacteria and fungi are abundant in the rumen and tend to produce cellulases which form part of a multi-enzyme complex (Gray *et al.* 2006, Warren 1996). The functional structure is called a cellulosome which consists of cellulolytic enzymes such as endoglucanases, cellobiohydrolases, xylanases and lichenases attached to a common scaffold. In these systems the enzymes are brought in close proximity to the substrate for hydrolysis, as well as for sufficient uptake of the products produced (Lynd *et al.* 2002). Anaerobic microorganisms, however, have a slow growth rate and produce low enzyme titers, which makes it unattractive for use in commercial cellulase production (Sun and Cheng 2002).

Numerous aerobic microorganisms have also been isolated with cell-wall degrading abilities (Warren 1996). They produce enzymes that do not form complexes, but are either freely secreted into the surrounding media, or are occasionally attached to the cell surface (Gray *et al.* 2006, Lynd *et al.* 2002). The non-complexed cellulolytic system of the aerobic fungus *Trichoderma reesei* has been the most intensively studied and is therefore probably the best understood (Warren 1996).

The soft rot fungus *T. reesei* (teleomorph: *Hypocrea jecorina*) was one of the first cellulolytic organisms isolated in the 1950s and is typically found in plant litter and soil (Howard *et al.* 2003, Sandgren *et al.* 2001). This fungus has received extensive interest, due to its ability to produce large amounts of extracellular cellulases (Mach and Zeilinger 2003). Its applications include the pulp and paper industry, the textile industry and food and feed production. The enzymes in the cellulase system of *T. reesei* degrade both hemicellulose and cellulose. Although cellulose is highly crystalline and insoluble, this filamentous fungus can penetrate the cavities between cellulose fibers to deliver its cellulase enzymes (Lynd *et al.* 2002). *T. reesei* produces five endoglucanases (EGI, EGII, EGIII, EGIV, and EGV), two exoglucanases (cellobiohydrolases CBHI and CBHII) and two β-glucosidases (BGLI and BGLII) (Lynd *et al.* 2002), which will be discussed briefly.

3.1 *T. reesei* endoglucanases

Endoglucanases cleave the more amorphous regions of cellulose, thereby facilitating cellulose hydrolysis by decreasing the degree of polymerization (Lynd *et al.* 2002). Endoglucanase I (EGI) (50 kDa) is one of the major endoglucanases secreted by *T. reesei* (Rabinovich *et al.* 2002). EGI is also

known as Cel7B because it is grouped into family 7 of the glycosyl hydrolases (www.cazy.org). The C-terminal CBM of the EGI enzyme belongs to family I and is linked to a cleft-like open active site (Figure 8) (Rabinovich *et al*. 2002). EGI is known to have broad substrate specificity, displaying (apart from cellulose derivatives) activity towards birchwood glucuronoxylan and locust bean gum galactomannan, and especially high activity towards unsubstituted beech xylan (Rabinovich *et al*. 2002, Bailey *et al.* 1993). EGI has shown to have higher specificity for the internal bonds of cellotetra- to cellohexaose, whereas in xylan hydrolysis, it cleaves the bond second from the reducing end molecule in xylotriose to xylohexaose. The main products formed from the hydrolysis of cellulosic substrates are cellobiose and glucose (Karlsson *et al.* 2002).

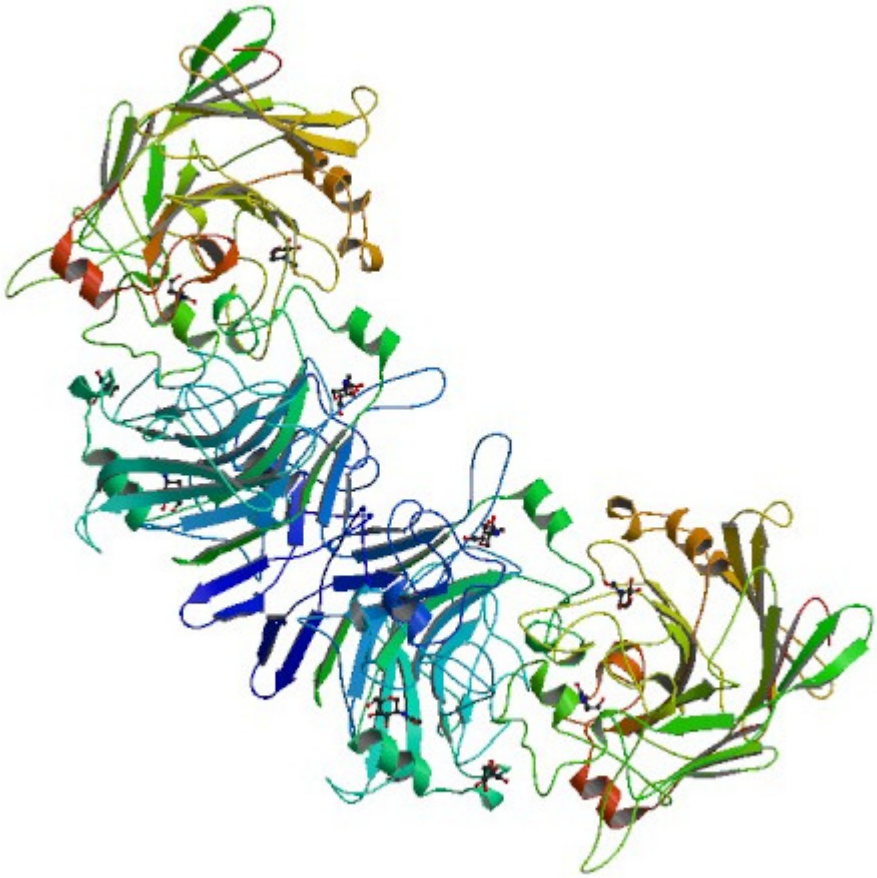

Figure 8: Ribbon diagram of *T. reesei* endoglucanase I (EGI or Cel7B), displaying the β-jellyroll fold typical of family 7 glycosyl hyrolases (www.cazy.org).

Endoglucanase II (EGII or Cel5A) is the other major endoglucanase produced by *T. reesei* and has a molecular weight of 48 kDa. However, unlike EGI, EGII displays strict specificity towards cellulose and its derivatives. It also hydrolyses cellotetra- to cellohexaose, but preferably cleaves bonds near the reducing ends of the oligosaccharides, the main end products being glucose, cellobiose and cellotriose. EGII has an N-terminal CBM belonging to family I, and similarly to EGI, an active site in the form of an open cleft (Karlsson *et al.* 2002, Rabinovich *et al.* 2002). Although the crystal structure for EGII has not yet been determined, it is known to represent an α-β-barrel (www.cazy.org).

The low molecular mass (25 kDa) endoglucanase, EGIII (Cel12A), represents less than 1% of the total cellulases expressed by *T. reesei* and differs structurally from the other endoglucanases (Sandgren *et al.* 2005). Although the active site is similar to that of EGI and EGII, a carbohydrate binding domain is absent (Karlsson *et al.* 2002). A large substrate binding groove is present in the protein, formed by the concave surface of the β-sandwich formed by 15 antiparallel β-sheets (Figure 9) (Sandgren *et al.* 2001). A single α-helix completes the enzyme's compact structure. EGIII has displayed activity towards not only cellulose, but also β-glucan and konjac glucomannan (Karlsson *et al*. 2002). Amorphous cellulose degradation by EGIII results in the release of glucose, cellobiose, cellotriose and cellotetraose, with cellobiose being the main end product.

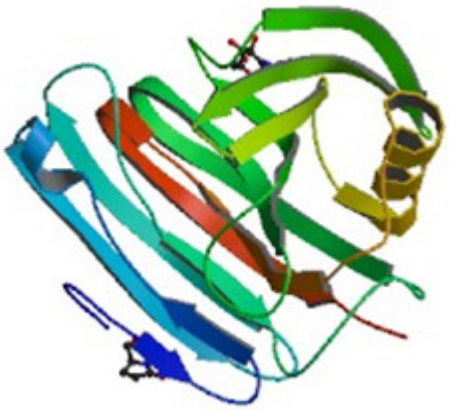

Figure 9: Ribbon diagram of the low molecular weight endoglucanase III (EGIII or Cel12A) from *T. reesei* (www.cazy.org).

Hydrolysis of amorphous cellulose substrates by endoglucanase IV (EGIV or Cel61A) (55 kDa) yields mainly cellobiose, as well as lesser amounts of glucose and cellotriose (Karlsson *et al.* 2001). EGIV displays activity towards shorter oligosaccharides such as cellotetraose and cellopentaose, although cellotriose proved to be a poor substrate. A unique feature of this endoglucanase is its exceptionally long *O*-glycosylated linker region connecting the catalytic and C-terminal cellulose binding module (Saloheimo *et al.* 1997). The crystal structure of EGIV has not yet been determined (www.cazy.org). Endoglucanase V (Cel45A) is another low molecular mass (23 kDa) cellulase secreted by *T. reesei* (Karlsson *et al*. 2002). EGV displays a completely different product formation pattern in the hydrolysis of amorphous cellulose, with cellotetraose produced as the main product, with lower levels of cellopentaose and cellotriose. This enzyme has an unusually small catalytic core, compared to other cellulases, linked to a C-terminal CBM (Saloheimo *et al.* 1994). The protein fold of *T. reesei* EGV is still unresolved (www.cazy.org).

3.2 *T. reesei* cellobiohydrolases

Cellobiohydrolases are classified as processive enzymes due to their manner of action; hydrolyzing the cellulose chains from the ends (Teeri 1997). It is mostly these enzymes that have the ability to hydrolyze crystalline cellulose efficiently (Schülein 2000). Cellobiohydrolase I (CBHI) and cellobiohydrolase II (CBHII) represent the majority of the total cellulase protein produced by *T. reesei* (60% and 20% respectively). The hydrolysis of microcrystalline cellulose by CBHs results in cellobiose as the main end product, with lesser amounts of cellotriose and glucose (Medve *et al.* 1998). CBHI has sequence similarity to EGI and is categorized into the same family of glycosyl hydrolases; therefore also known as Cel7A (Saloheimo *et al.* 2002, 1997). This enzyme has a C-terminal CBM and hydrolyzes the cellulose chain from the reducing end (Lynd *et al.* 2002, Saloheimo *et al.* 2002, 1997, 1994).

CBHII, also referred to as Cel6A, differs from CBHI in its preference for the non-reducing chain ends in the hydrolysis of cellulose (Lynd *et al.* 2002, Rabinovich *et al.* 2002). Similarly to EGIV, CBHII (Cel6A) has an unusually long *O*-glycosylated linker region connecting the N-terminal CBM to the catalytic domain (Saloheimo *et al.* 2002, 1997, 1994). The active site of both the cellobiohydrolases, CBHI and CBHII, is located in a tunnel formed by extended loops (Figure 10) (Sandgren *et al.* 2005, Karlsson *et al.* 2002, Davies and Henrissat 1995). The tunnel surrounds the cellulose chain, enabling it

to be threaded through. The cellulose chain is therefore hydrolyzed processively, releasing cellobiose, while remaining attached to the enzyme. CBHI has four surface loops and ten glucose binding sites in the tunnel, while CBHII has only two loops and six glucose binding sites (Sandgren *et al.* 2005, Teeri 1997).

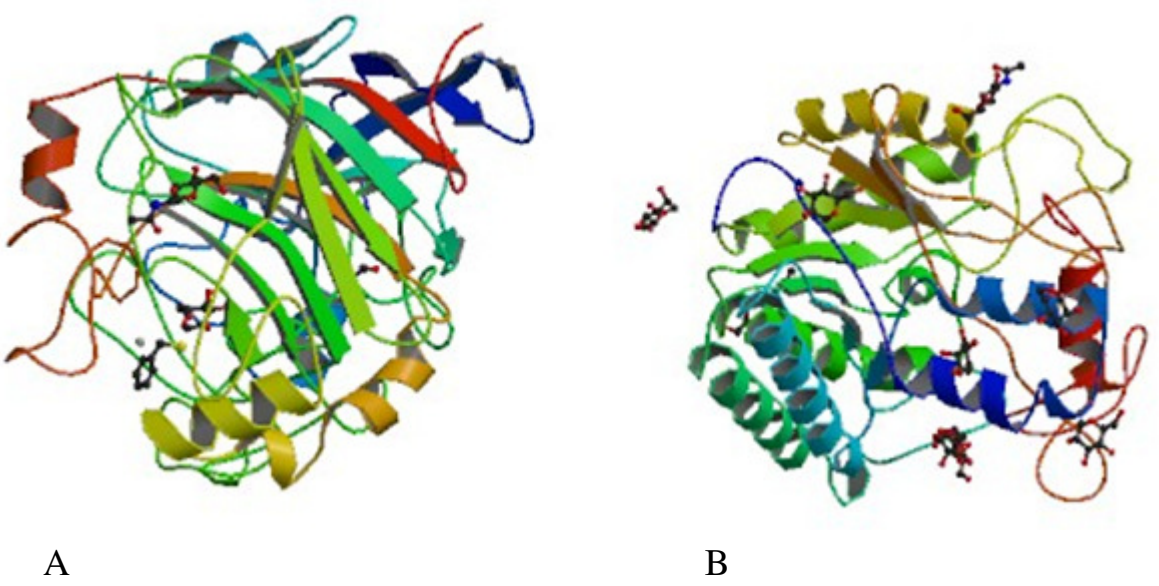

Figure 10: Ribbon diagrams of *T. reesei* (A) CBHI (or Cel7A) and (B) CBHII (or Cel6A), representing β-jelly roll and α-β-barrel structures, respectively (Saloheimo *et al.* 2002, 1997, www.cazy.org).

3.3 *T. reesei* β-glucosidases

The β-glucosidases produced by *T. reesei* fall into the glycosyl hydrolase families 1 (BGLII) and 3 (BGLI), hence its alternative nomenclature Cel1A and Cel3A, respectively (www.cazy.org). The BGLI enzyme is secreted by *T. reesei*, while data suggest BGLII to be cell wall bound (Saloheimo *et al.* 2002). Together with EGIII, β-glucosidase I is the only other cellulase lacking a CBM (Saloheimo *et al.* 1997). β-glucosidases hydrolyze mainly cellobiose, an inhibitor of cellulase enzymes (Sun and Cheng 2002). This enzyme's broad substrate specificity also allows hydrolysis of longer cello-oligosaccharides, as well as different aryl- and alkyl-β-D-glucosides (Van Rooyen *et al.* 2005). In a study by Saloheimo *et al.* (2002) BGLII displayed activity towards cellotriose and cellotetraose. Since glucose and cellotriose were produced from cellotetraose, it has been suggested that this enzyme hydrolyzes only the bond next to a terminal glucose unit. However, it has also been demonstrated that BGLII has transglycosylation activity at high concentrations of glucose; producing

sophorose and cellobiose. β-glucosidases prefer substrates with a large number of available chain ends, with the pocket-like active site favouring the β-*O*-glycosidic bond at the non-reducing ends of oligosaccharides (Opassiri *et al.* 2006, Davies and Henrissat 1995).

The level of β-glucosidase production by *T. reesei* is restricted and account for only 1% of the total secreted protein (Takashima *et al.* 1999). This limits the hydrolysis of cellulose by the entire cellulase system. Therefore β-glucosidases from other origins are being investigated. In a study where four β-glucosidase genes of fungal origins were expressed in *Saccharomyces cerevisiae*, the BGL1 from *Saccharomycopsis fibuligera* produced the highest level of activity (Van Rooyen *et al.* 2005). This enzyme is therefore more attractive for use in future studies regarding synergistic cellulose hydrolysis.

3.4 Synergism

The phenomenon known as synergism takes place when the collective activity of the components of a cellulase system is higher than the sum of the activities of the individual enzymes (Lynd *et al*. 2002). Cellulose can be degraded to cellobiose and cellooligosaccharides by the simultaneous action of an endoglucanase and a cellobiohydrolase by endo-exo-synergism. The endoglucanases provide free chain ends for cellobiohydrolase attack (Teeri 1997). Medve *et al.* (1998) has shown that the synergistic action of EGII and CBHI is much higher than the theoretically determined sum of the individual conversion of the two enzymes. Synergism between EGIII and CBHI of *T. reesei* has also been reported in the hydrolysis of crystalline cellulose (Okada *et al.* 1998). In a study on the synergistic activity of CBHII and EGI, a greater degree of hydrolysis was observed using a combination of the two enzymes, compared to their individual activities (Bailey *et al.* 1993). However, the strongest synergism between purified cellulase enzymes from *T. reesei* has been found for EGI in conjunction with CBHI (Henrissat *et al.* 1985).

Exo-exo-synergism can also exist between two cellobiohydrolases, when one enzyme act on the reducing end of the cellulose chain, while the other hydrolyzes from the non-reducing end, as has been observed for CBHI and CBHII (Sandgren *et al.* 2005, Fujita *et al.* 2004, Lynd *et al.* 2002). Synergy also exists between exo- or endoglucanases and β-glucosidases, as well as between the catalytic domains and CBMs within the cellulase enzyme itself (Zhang and Lynd 2004, Lynd *et al.* 2002).

Synergism is dependent on the ratio of individual enzymes, the substrate saturation and the properties of the substrate (Medve *et al.* 1998).

3.5 Heterologous protein production for CBP by *Saccharomyces cerevisiae*

Escherichia coli was originally used as a host for the production of heterologous proteins mainly due to its rapid growth rate and cost effective cultivation (Yin *et al.* 2007, Domínguez *et al.* 1998). *E. coli* has limited applications due to its inability to process eukaryotic introns and multimeric proteins. Other setbacks include the lack of phosphorylation and the lack of post-translational modifications which is required for proper folding and function of proteins (Yin *et al.* 2007). Therefore, a wider variety of hosts are constantly being investigated and evaluated (Domínguez *et al.* 1998). Eukaryotic hosts have the ability to perform posttranslational modifications required for effective production of many functional proteins which are heterologously expressed. As mentioned previously, the recombinant cellulolytic strategy for organism development requires the host organism to produce high product titers and display high product tolerance (Lynd *et al.* 2002). In this regard, the yeast *Saccharomyces cerevisiae* has received the most attention for the production of heterologous proteins for CBP (Domínguez *et al.* 1998).

Saccharomyces cerevisiae is commonly used in the food and beverage industry due to its GRAS status (Generally Regarded As Safe) (Domínguez *et al.* 1998). It is often the yeast of choice for heterologous protein production, owing to the ease with which it can be genetically manipulated, its high ethanol tolerance, its robustness in industrial processes and the great amount of information available regarding its molecular biology and fermentation technology (Den Haan *et al.* 2007). The first recombinant protein produced by this yeast was the human interferon in the early 1980's. *S. cerevisiae* has ever since played a major role in contributing to biotechnological research (Ostergaard *et al.* 2000).

S. cerevisiae has the ability to convert hexose sugars (glucose, fructose, galactose and mannose) and some disaccharides (sucrose and maltose) to ethanol by means of fermentation, but is unable to utilize cellobiose and longer cellooligosaccharides (Van Zyl *et al.* 2007). The yeast is therefore unable to grow on polysaccharides such as cellulose and hemicellulose resulting in a major limitation in the versatility of the yeast. Research is currently directed towards constructing a recombinant *S. cerevisiae* strain which can produce ethanol directly through utilizing cellulose by means of a heterologous cellulase

system. All five endo-β-1,4-glucanases (encoded by *egI, egII, egIII, egIV* and *egV*) and two cellobiohydrolases (encoded by *cbhI* and *cbhII*) from *T. reesei* have already been efficiently secreted into the culture medium by recombinant *S. cerevisiae* strains (Lynd *et al.* 2002). All entered the secretory pathway of the yeast, but were highly glycosylated and heterogeneous in size. *S. cerevisiae* elongates the mannose chain in heterologous proteins by adding even more mannose residues. An outer chain is formed which results in "hyperglycosylation" (Romanos *et al.* 1992).

The first direct and efficient fermentation of cellulosic material by a recombinant *S. cerevisiae* strain was reported by Fujita *et al*. (2002). Cellulose degrading yeast cells were constructed which co-displayed *Aspergillus aculeatus* β-glucosidase 1 (BGL1) and *T. reesei* endoglucanase II (EGII) on the cell surface. The yeast strain displaying EGII and the yeast strain codisplaying BGL1 and EGII showed high levels of cell wall bound EGII activity towards barley β-glucan. In 2004, the same research group co-displayed *T. reesei* endoglucanase II, cellobiohydrolase II and *A. aculeatus* β-glucosidase 1 simultaneously on the cell surface of *S. cerevisiae*. The yeast strain co-displaying EGII and CBHII showed significantly higher activity on amorphous cellulose (phosphoric acid-swollen cellulose) than the one displaying only EGII. The main end product formed was cellobiose. Simultaneous expression of all three cellulolytic genes (*EGII*, *CBHII* and *BGL1*) enabled the *S. cerevisiae* strain to directly produce ethanol from the amorphous cellulose (Fujita *et al.* 2004). However, anaerobic growth on cellulose as a result of synergistic expression has only been demonstrated in 2007 (Den Haan *et al.* 2007). The expression of an endoglucanase (*egI* of *T. reesei*) together with β-glucosidase (*bgl1* of *S. fibuligera*) by *S. cerevisiae* enabled the recombinant yeast to grow on phosphoric acid swollen cellulose (PASC), an amorphous cellulose substrate. The recombinant yeast was able to produce ethanol at concentrations of up to 1.0 $g.l^{-1}$ from 10 $g.l^{-1}$ PASC.

4. ENZYME IMPROVEMENT

Proteins have an evolutionary potential to acquire new specificities or functions that differ from that of the original protein. This enables the evolution of a protein through amino acid changes under controlled laboratory conditions, the main goal being new or improved function (Yuan *et al.* 2005). These changes would include the adaptation of protein functions to extreme conditions, improved recombinant protein biosynthesis as well as altered specificities and activities of enzymes. Mutational studies can also be used as a tool for unraveling protein structure-function relationships.

4.1 Rational design

Rational design can be applied when the structure of a protein, as well as the mechanism with which it operates, is known. Therefore the consequences of a change at a particular site can be predicted (Yuan *et al.* 2005, Bornscheuer and Pohl 2001). Since the protein structure is known, mutants can be planned, prepared and propagated in a host organism. Site-directed mutagenesis is becoming a powerful method for engineering enzymes, since information regarding protein structure and function is constantly increasing (Rubin-Pitel *et al.* 2006). During site-directed mutagenesis an oligonucleotide is constructed which is, apart from a desired mismatch, complementary to a piece of template DNA on a single stranded plasmid (Figure 11) (Carter 1986). After annealing, this mutagenic primer DNA is extended with the Klenow fragment of *E. coli* DNA polymerase I, to form a double stranded plasmid. The transformation of this plasmid to *E. coli* results in both wild-type and mutant progeny, which is identified either by hybridization with a probe or by sequencing. Multiple mutations, insertions and deletions can also be introduced using this approach.

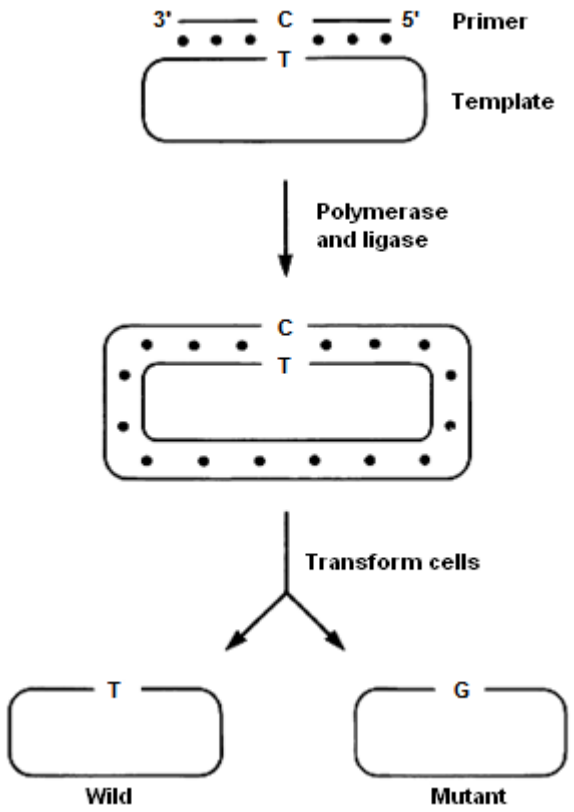

Figure 11: Site-directed mutagenesis by the incorporation of a mismatched nucleotide, resulting in both wild-type and mutant progeny (Carter 1986).

Site-directed mutagenesis is not only time-consuming and expensive, but also requires information regarding the protein before being applied (Sylvestre *et al*. 2005, Bornscheuer 1998). Protein structure and function is complex and unfortunately not available for the majority of proteins. Directed evolution presents a more practical and simple approach to generating mutants with improved properties (Tao and Cornish 2002).

4.2 Directed evolution

Natural evolution of proteins is induced by external factors such as irradiation and oxidation, as well as by genetic errors, including failures of DNA replication or repair (Rubin-Pitel *et al.* 2006). This leads to major diversity, due to the introduction of transitions, transversions, deletions, insertions and inversions. A transition occurs when a purine nucleotide is substituted by another purine, or a pyrimidine by another pyrimidine. When a purine nucleotide is substituted by a pyrimidine, or vice versa, it is referred to as a transversion. Deletions and insertions involve the elimination or addition, respectively, of nucleotides to a gene. An inversion occurs when a double-stranded DNA fragment undergoes a 180° rotation.

Directed evolution mimics evolution in nature, generating enzymes with altered activity, specificity and stability; on a laboratory timescale (Hibbert and Dalby 2005). It involves the generation of large mutant protein libraries and selection of desirable functions, without prior knowledge regarding the protein sequence, structure, or mechanism (Yuan *et al.* 2005, Brakmann 2001). A gene can either be randomly mutated or the gene fragments recombined (Bornscheuer 1998). Directed evolution has led to improved characteristics such as higher reaction rates and increased performance of enzymes at high temperatures, as well as improved expression, folding and secretion of recombinant enzymes produced by a heterologous host (Roodveldt *et al.* 2005).

The conventional methods used to generate mutations in genes include the use of chemical mutagens such as sodium bisulfite, methoxylamine, nitrous acid, hydroxylamine or treatment with ultraviolet light (Rubin-Pitel *et al.* 2006). DNA is directly damaged and incorrectly replicated to progeny (Neylon 2004). These methods however, are discontinuous and can lead to great cell damage (Rubin-Pitel *et al.* 2006). Applied methods of directed mutagenesis are based mainly on error-prone DNA replication and DNA shuffling (Roodveldt *et al.* 2005).

4.2.1 Error-prone polymerase chain reaction (PCR)

Error-prone PCR is one of the most popular methods currently applied in directed evolution. This method relies on the inaccuracy of DNA polymerases to incorporate the correct nucleotides during PCR (Rubin-Pitel *et al.* 2006). After a number of cycles, the reaction mixture contains a collection of DNA, differing slightly from the parental DNA. The standard steps of denaturation, annealing and primer extension are followed to amplify a small amount of DNA (Moore and Maranas 2000). The rate of mutations made spontaneously by DNA polymerases during incorporation of nucleotides is too low to be useful in the construction of a library of mutants. Therefore, error rates can be increased by generating mutagenic reaction conditions, such as using Mn^{2+} instead of Mg^{2+} as a co-factor, as well as including biased concentrations of dNTPs in the reaction mixture (Neylon 2004). During the extension step, where complementary nucleotides are added to the original template, unavoidable mismatches are made and incorporated into the new double stranded DNA (Moore and Maranas 2000). A common problem with this technique is that the mutation frequency needs to be optimally regulated, since deleterious mutations frequently overshadow beneficial ones. Also, diversity may be limited by the mutational biases of DNA polymerases (Rubin-Pitel *et al.* 2006).

4.2.2 DNA Shuffling

The first *in vitro* homologous recombination method, called DNA shuffling, was introduced in 1994 (Rubin-Pitel *et al.* 2006). In DNA shuffling random point mutations are introduced and recombined in a gene (Parikh and Matsumura 2005). The gene is randomly fragmented and reassembled into chimeric sequences by PCR, during which mutations are introduced due to recombination (Yaun *et al.* 2005). DNA shuffling occurs in three steps, starting off with parental sequences sharing desired traits (Figure 12). These sequences are randomly fragmented with DNase I where after fragments within a certain size range is reassembled again by thermocycling, a process similar to PCR. The double stranded fragments are denatured into single stranded chains, which in the next step anneal to other sufficiently complementary strands. Double stranded DNA is formed again, however, containing 3' or 5' overhangs. During polymerase extension, the 5' overhangs are filled in by the DNA polymerase in a 5'→3' manner, whereas the 3' overhangs stay unchanged. Denaturation, annealing and extension are repeated a number of times until DNA chains of the original length are obtained. These chimeras of double stranded DNA are then amplified by a normal PCR reaction (Moore and Maranas 2000).

For this method to be successful, the presence of regions with high sequence homology surrounding regions of diversity is needed. Another draw-back of homologous recombination methods such as DNA shuffling is the necessity of high sequence identity among parent genes (Rubin-Pitel *et al.* 2006).

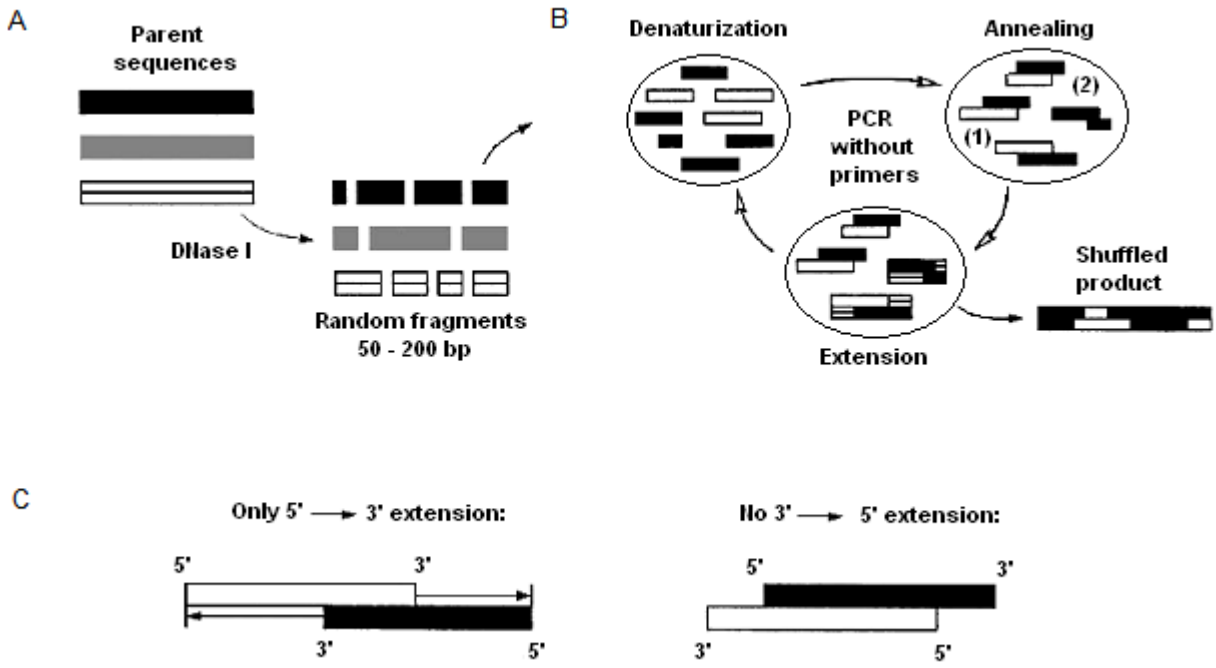

Figure 12: Illustration of the three steps of DNA shuffling, resulting in double stranded DNA chimeras (Moore and Maranas 2000). The parent sequences are fragmented by DNase I (A) and reassembled again by thermocycling (B). DNA polymerase fills in the 5' overhangs (C), while the 3' overhangs stay unchanged.

4.2.3 Mutagenic strains

Passing cloned genes through a mutator strain is a simple method to introduce random mutations in a gene (Bornscheuer 1998). Apart from being generated in the laboratory, mutators are commonly found in natural populations, where they can acquire beneficial mutations such as antibiotic resistance (Džidić *et al.* 2003). Mutator strains have higher spontaneous mutation rates than the wild type strains, due to defects in their DNA repair mechanisms (Džidić *et al.* 2003, Brakmann 2001). The majority of mutator abilities are found to be linked to defects in the *mutS, mutH, mutL or uvrD* genes, all of which play a role in the methyl-directed mismatch repair pathway of *E. coli* (Li *et al.* 2003). This system

corrects base mismatches in newly replicated DNA and also inhibits recombinatory events between species.

In this method of mutagenesis, the parental gene is cloned into a vector plasmid and transformed to the mutator strain cells. After a number of replications, the plasmid is retrieved from the mutagenic cells and further propagated in a repair competent strain for selection of improved properties. The mutation rate however is low, with approximately one mutation per 1000 base pairs per mutation cycle (Rubin-Pitel *et al.* 2006). Since the whole plasmid is mutated, other regions can also obtain defects or improved properties, which make screening for improvement of the gene itself, more complicated (Bornscheuer 1998).

During the process of replication, errors such as mismatch bases, deletions or insertions can occur (Yang 2000). Some of these errors escape proofreading by the polymerase protein and consequently the DNA mismatch repair system is responsible for maintaining genomic integrity by repairing the errors. It is also known as the mutHLS pathway, since the repair mechanism has been shown to depend mainly on three proteins: MutS, MutL and MutH (Džidić *et al.* 2003). To distinguish between the parental DNA and the new daughter strand, *E. coli* utilizes methyltransferases to add a methyl group to the d(GATC) sequence of the template DNA, generating a hemimethylated duplex with the newly synthesized strand (Yang 2000). The temporarily unmethylated strand serves as substrate for repair. A single hemimethylated d(GATC) sequence is sufficient to direct mismatch repair (Lyer *et al.* 2006).

The MutS protein (Figure 13) recognizes one of 8 different kinds of mismatches or a loop structure resulting from either a deletion or an insertion of 1-4 nucleotides in the new daughter strand (Kunkel and Erie 2005, Yang 2000). It has a high affinity for the most frequent polymerization errors such as G-T mismatches and single insertion–deletion errors, and lower affinity for other mismatches. Crystal structures of MutS bound to DNA reveal a direct interaction of the Phe and Glu residues of the protein with the mismatch (Junop *et al.* 2003).

The binding of the MutL protein to the DNA activates its ATPase activity (Kunkel and Erie 2005, Yang 2000). Although the interactions between MutL and MutS are not yet fully understood, MutL is suggested to be an ATP-operated signaling molecule, mediating interactions between MutS and MutH. Interaction of MutS with the MutL dimer activates MutH, a restriction endonuclease, which cleaves the

new, temporarily unmethylated strand at the GATC site, either on the 3' or the 5' side of the error, preferentially along the shortest path to the mismatch (Kunkel and Erie 2005, Modrich 1991). MutL also loads a helicase, UvrD, onto the DNA at the nick (Jun *et al.* 2006).

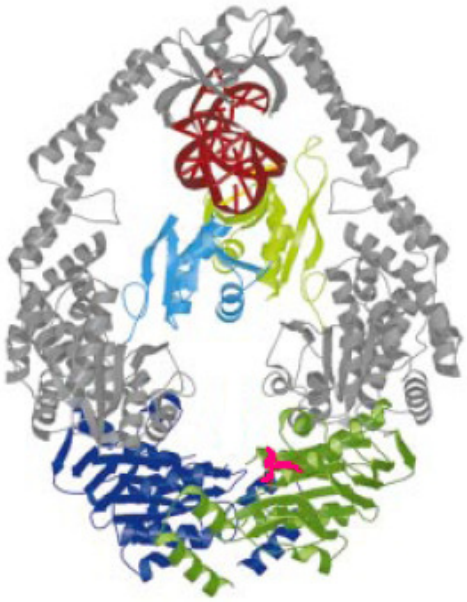

Figure 13: The dimer structure of the MutS protein. The protein (grey) is bound to a mismatch in DNA (red). The ATPase domains are coloured dark blue and green with the mismatch binding domains, light blue and light green. The ADP molecule is in pink (Adapted from Lamers *et al.* 2004).

DNA helicase II (UvrD) and single-strand DNA-binding protein (SSB) is also activated by the MutS-MutL complex and work together to generate single-stranded DNA between the nick and the mismatch, which is digested by either 3' or 5' exonucleases, depending on the location of the nick relative to the mismatch (Lyer *et al.* 2006, Kunkel and Erie 2005). The helicase unwinds the DNA as it moves along in a bidirectional manner through the hydrolysis of ATP. If the nick is made 3' to the mismatch, exonucleases ExoI, ExoVII or ExoX hydrolyzes the strand in a 3'- 5' direction, while exonuclease ExoVII or RecJ is responsible for 5'-3' hydrolytic activity (Jun *et al.* 2006, Lyer *et al.* 2006, Modrich 1991). The gap generated through the action of the exonucleases is stabilized by SSB, while DNA polymerase III correctly resynthesizes a new strand on the template DNA (Lyer *et al.* 2006, Kunkel and Erie 2005). DNA ligase reseals the nick and thereby restores the DNA helix structure (Figure 14).

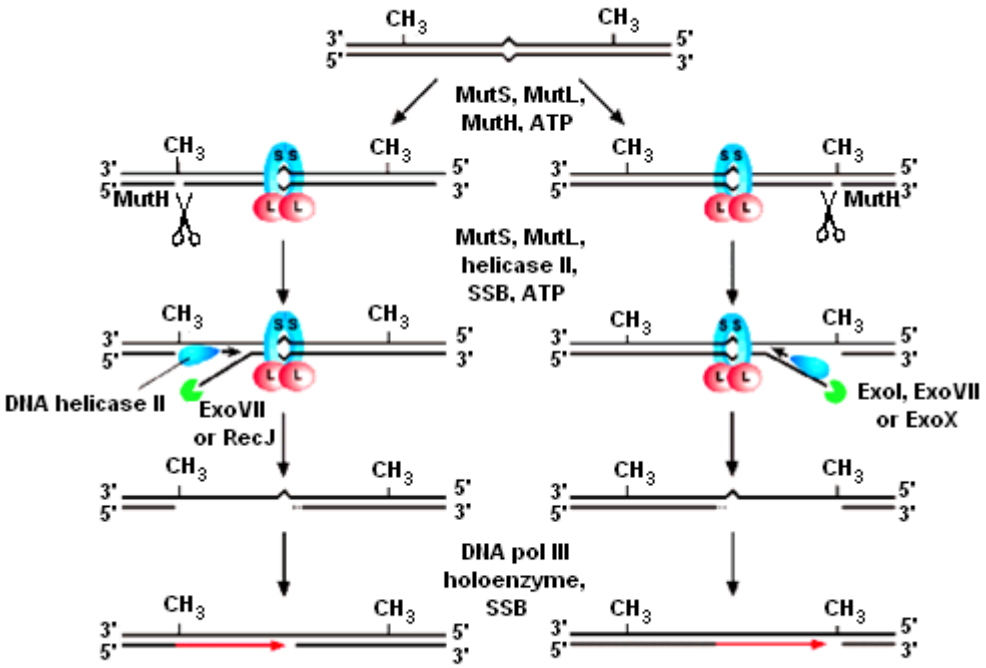

Figure 14: Schematic representation of *E. coli* DNA mismatch repair, indicating the three main repair proteins, MutS (S), MutL (L) and MutH, as well as all the other participating proteins (Lyer *et al.* 2006).

The popularity of using directed evolution in the laboratory is increasing due to the ease with which it can be applied to engineer proteins with improved characteristics. Directed evolution experiments have been successfully applied to optimize protein activities, such as binding, stability and enzyme selectivity, for use in industrial biocatalysts (Neylon 2004, Zhao *et al.* 2002). A wide range of techniques are currently available for creating protein libraries with altered activities and the choice of technique depends on accessibility, applicability and cost-effectiveness.

5. THIS STUDY

S. cerevisiae has a long association with the food and beverage industry (Ostergaard *et al.* 2000). This yeast is an attractive tool for expression of recombinant proteins due to the variety of vector systems and promoters available and the ease of product purification. Although science has made significant progress in the past two decades, some limitations associated with low product yields could not be overcome.

This study was undertaken in light of the increasing pressure on countries to implement green technology for waste disposal and the escalating energy demand. Both these problems could be addressed if waste (cellulose in particular) could be converted to bioethanol. The aim would be to provide a cost effective means of decreasing the quantity of waste produced annually, while producing an economically important commodity.

The yeast, *S. cerevisiae*, is currently under investigation for the possible use as host for bioethanol production through genetic engineering. In general, the levels of foreign gene expression of cellulases are low due to the limited secretion capacity. In this study the *E. coli* ES1301 *mutS* mutator strain was used to create mutations on the DNA sequence of the shuttle vector (containing the *egI of T. reesei*) in order to improve the levels of endoglucanase activity produced by the host *S. cerevisiae.* A transformant was obtained that produced higher levels of endoglucanase activity. Part of the mutated shuttle vector was sequenced and the possible relevant mutations had been identified. The altered *egI* gene, as well as the *egII* gene of *T. reesei*, was co-expressed with other cellulase genes in an attempt to find an optimum cocktail of cellulases for CBP.

6. REFERENCES

Antoni D, Zverlov VV, Schwarz WH (2007) Biofuels from microbes. Applied Microbiology and Biotechnology 77: 23-35

Aristidou A, Penttilä M (2000) Metabolic engineering applications to renewable resource utilization. Current Opinion in Biotechnology 11: 187-198

Bailey MJ, Siika-aho M, Valkeajärvi A, Penttilä ME (1993) Hydrolytic properties of two cellulases of *Trichoderma reesei* expressed in yeast. Biotechnology and Applied Biochemistry 17: 65-76

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

Bornscheuer UT (1998) Directed evolution of enzymes. Angewandte Chemie International Edition 37: 3105-3108

Bornscheuer UT, Pohl M (2001) Improved biocatalysts by directed evolution and rational protein design. Current Opinion in Chemical Biology 5: 137-143

Brakmann S (2001) Discovery of superior enzymes by directed molecular evolution. Chembiochem 2: 865-871

Carter P (1986) Site-directed mutagenesis. Biochemical Journal 237: 1-7

Coughlan MP, Hazlewood GP (1993) Hemicellulose and hemicellulases, Portland Press, London

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

Den Haan R, Rose SH, Lynd LR, Van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae.* Metabolic Engineering 9: 87-94

Domínguez Á, Fermiñán E, Sánchez M, González FJ, Pérez-Campo FM, García S, Herrero AB, San Vicente A, Cabello J, Prado M, Iglesias FJ, Choupina A, Burguillo FJ, Fernández-Lago L, Carmen López M (1998) Nonconventional yeasts as hosts for heterologous protein production. International Microbiology 1: 131-142

Džidić S, Bačun-Družina V, Petranović M (2003) The role of mismatch repair in bacterial evolution. Food Technology and Biotechnology 41: 177-182

Eriksson K-EL, Blanchette RH, Ander P (1990) Biodegradation of lignin. In: Eriksson K-EL, Blanchette RH, Ander P (eds) Microbial and enzymatic degradation of wood and wood components. Springer-Verlag, Berlin Heidelberg New York

Fujita Y, Ito J, Ueda M, Fukuda H, 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. Applied and Environmental Microbiology 70: 1207-1212

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

Galbe M, Zacchi G (2002) A review of the production of ethanol from softwood. Applied Microbiology and Biotechnology 59: 618-628

Gray KA, Zhao L, Emptage M (2006) Bioethanol. Current Opinion in Chemical Biology 10: 1-6

Gruno M, Väljamäe P, Pettersson G, Johansson G (2003) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. Biotechnology and Bioengineering 86: 503-511

Henrissat B, Driguez H, Viet C, Schülein M (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. Biotechnology 3: 722-726

Hibbert EG, Dalby PA (2005) Directed evolution strategies for improved enzymatic performance. Microbial Cell Factories 4, DOI 10.1186/1475-2859-4-29

Hill J, Nelson E, Tilman D, Polasky S, Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy of Sciences 103: 11206-11210

Hon DN-S (1995) Chemical modification of lignocellulosic materials. CRC Press, United States

Howard RL, Abotsi E, Jansen van Rensburg EL, Howard S (2003) Lignocellulose biotechnology: issues of bioconversion and enzyme production. African Journal of Biotechnology 2: 602-619

Ito J, Fujita Y, Ueda M, Fukuda H, Kondo, A (2004) Improvement of cellulose-degrading ability of a yeast strain displaying *Trichoderma reesei* endoglucanase II by recombination of cellulose-binding domains. Biotechnology Progress 20: 688-691

Junop MS, Yang W, Funchain P, Clendenin W, Miller JH (2003) *In vitro* and *in vivo* studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. DNA Repair 2: 387-405

Jun S-H, Gyun Kim T, Ban C (2006) DNA mismatch repair system, classical and fresh roles. FEBS Journal 273: 1609-1619

Karlsson J, Saloheimo M, Siika-aho M, Tenkanen M, Penttilä M, Tjerneld F (2001) Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei.* European Journal of Biochemistry 268: 6498-6507

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

Kirk TK, Cullen D (1998) Enzymology and molecular genetics of wood degradation by white-rot fungi. In: Young RA, Akhtar M (eds) Environmentally friendly technologies for the pulp and paper industry, John Wiley and Sons, New York, pp 273-308

Kirk TK, Higuchi T, Chang HM (1980) Lignin biodegradation: microbiology, chemistry, and applications. CRC Press, United States

Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. Angewandte Chemie International Edition 44: 3358-3393

Klemm D, Schmauder H-P, Heinze T (2002) Cellulose. In: Vandamme E, De Baets S, Steinbüchel A (eds) Biopolymers: biology, chemistry, biotechnology, application, Polysaccharide II, Wiley-VCH, Weinheim, Germany, pp 275-319

Kunkel TA, Erie DA (2005) DNA mismatch repair. Annual Review of Biochemistry 74: 681-710

Lamers MH, Georgijevic D, Lebbink JH, Winterwerp HHK, Agianian B, De Wind N, Sixma TK (2004) ATP increases the affinity between MutS ATPase domains, Implications for ATP hydrolysis and conformational changes. Journal of Biological Chemistry 279: 43879-43885

Li B, Tsui H-CT, LeClerc JE, Dey M, Winkler ME, Cebula TA (2003) Molecular analysis of *mutS* expression and mutation in natural isolates of pathogenic *Escherichia coli.* Microbiology 149: 1323-1331

Lin Y, Tanaka S (2006) Ethanol fermentation from biomass resources: current state and prospects. Applied Microbiology and Biotechnology 69: 627-642

Lyer RR, Pluciennik A, Burdett V, Modrich PL (2006) DNA mismatch repair: functions and mechanisms. Chemical Reviews 106: 302-323

Lynd LR (1996) Overview and evaluation of fuel ethanol from cellulosic biomass: technology, economics, the environment, and policy. Annual Review of Energy and the Environment 21: 403-465

Lynd LR, Van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology 16: 577-583

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002*)* Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews 66: 506-577

Mach RL, Zeilinger S (2003) Regulation of gene expression in industrial fungi: *Trichoderma.* Applied Microbiology and Biotechnology 60: 515-522

Medve J, Karlsson J, Lee D, Tjerneld F (1998) Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: adsorption, sugar production pattern, and synergism of the enzymes. Biotechnology and Bioengineering 59: 621-634

Modrich P (1991) Mechanisms and biological effects of mismatch repair. Annual Review of Genetics 25: 229-253

Moore GL, Maranas CD (2000) Modeling DNA mutation and recombination for directed evolution experiments. Journal of Theoretical Biology 205: 483-503

Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Research 32:1448-1459

Okada H, Tada K, Sekiya T, Yokoyama K, Takahashi A, Tohda H, Kumagai H, Morikawa Y (1998) Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. Applied and Environmental Microbiology 64: 555-563

Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, Ketudat Cairns JR (2006) Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. BMC Plant biology 6, DOI: 10.1186/1471-2229-6-33

Ostergaard S, Olsson L, Nielsen J (2000) Metabolic Engineering of *Saccharomyces cerevisiae.* Microbiology and Molecular Biology Reviews 64: 34-50

O'Sullivan A (1997) Cellulose: the structure slowly unravels. Cellulose 4: 173-207

Parikh MR, Matsumura I (2005) Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of β-fucosidase from β-galactosidase. Journal of Molecular Biology 352: 621-628

Pérez J, Muñoz-Dorado J, De la Rubia T, Martínez J (2002) Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. International Microbiology 5: 53-63

Pu Y, Zhang D, Singh PM, Ragauskas AJ (2007) The new forestry biofuels sector. Biofuels, Bioproducts and Biorefining 2: 58-73

Rabinovich ML, Melnik MS, Bolobova AV (2002) Microbial Cellulases (Review). Applied Biochemistry and Microbiology 38: 305-321

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick Jr. WJ, Hallet JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311: 484-489

Romanos MA, Scorer CA, Clare JJ (1992) Foreign gene expression in yeast: a review. Yeast 8: 423-488

Roodveldt C, Aharoni A, Tawfik DS (2005) Directed evolution of proteins for heterologous expression and stability. Current Opinion in Structural Biology 15: 50-56

Rubin-Pitel SB, Cho CM-H, Chen W, Zhao H (2006) Directed evolution tools in bioproduct and bioprocess development. In: Yang S-T (ed) Bioprocessing for value-added products from renewable resources: new technologies and applications, Elsevier Science, New York, pp 49-72

Saloheimo A, Henrissat B, Hoffrén A-M, Teleman O, Penttilä M (1994) A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. Molecular Microbiology 13: 219-228

Saloheimo M, Kuja-Panula J, Ylösmäki E, Ward M, Penttilä M (2002) Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A). Applied and Environmental Microbiology 68: 4546-4553

Saloheimo M, Nakari-Setälä T, Tenkanen M, Penttilä M (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. European Journal of Biochemistry 249: 584-591

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

Sandgren M, Ståhlberg J, Mitchinson C (2005) Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. Progress in Biophysics and Molecular Biology 89: 246-291

Schülein M (2000) Protein engineering of cellulases. Biochimica et Biophysica Acta 1543: 239-252

Sjöström E (1993) Wood Chemistry: Fundamentals and Applications, 2nd edn. Academic Press New York, New York

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

Sylvestre J, Blesa S, Delcourt M (2005) Massive mutagenesis: high-throughput, high precision library creation. Drug Plus International

Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T (1999) Molecular cloning and expression of the novel fungal β-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. The Journal of Biochemistry 125: 728-736

Tao H, Cornish VW (2002) Milestones in directed enzyme evolution. Current Opinion in Chemical Biology 6: 858-864

Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Tibtech 15: 160-166

Van Maris AJA, Abbott DA, Bellissimi E, Van den Brink J, Kuyper M, Luttik MAH, Wisselink HW, Scheffers WA, Van Dijken JP, Pronk JT (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 WH, Pretorius IS (1998) Engineering yeast for efficient cellulose degradation. Yeast 14: 67-76

Van Rooyen R, Hahn-Hägerdal B, La Grange DC, Van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. Journal of Biotechnology 120: 284-295

Van Wyk JPH (2001) Biotechnology and the utilization of biowaste as a resource for bioproduct development. TRENDS in Biotechnology 19: 172-177

Van Zyl WH, Lynd LR, Den Haan R, McBride JE (2007) Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae.* Advances in Biochemical Engineering/Biotechnology 108: 205-35.

Warren RAJ (1996) Microbial hydrolysis of polysaccharides**.** Annual Review of Microbiology 50: 183-212

Yang W (2000) Structure and function of mismatch repair proteins. Mutation research 460: 245-256

Yin J, Guangxing L, Xiaofeng R (2007) Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. Journal of Biotechnology 127: 335-347

Yuan L, Kurek I, English J, Keenan R (2005) Laboratory-directed protein evolution. Microbiology and Molecular Biology Reviews 69: 373-392

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

Zhang Y-HP, Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnology and Bioengineering 88: 797-824

Zhao H, Chockalingam K, Chen Z (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. Current Opinion in Biotechnology 13: 104-110

Websites cited:

Carbohydrate-Active enZYmes. Last updated: 1 August 2008. AFMB – CNRS –Universités Aix-Marseille I & II. Accessed: 19 August 2008 <www.cazy.org>

Nutrition Resources. Chemistry review: Carbohydrates. Copyright date: 2006. Jones and Bartlett Publishers. Accessed: 19 August 2008 <www.nutrition.jbpub.com/resources/chemistryreview9.cfm>

Online dictionary. die.net. Last updated 5 April 2008. Accessed: 25 August 2008 <www.dictionary.die.net>

Plant Cell Walls. Complex Carbohydrate Research Centre. Last updated: 20 August 2007. The University of Georgia. USA. Accessed: 25 August 2008 <www.ccrc.uga.edu/~mao/intro/ouline.htm>

Pipeline consideration for ethanol. Reviewed: 2002. Whims, J. Sparks Company Inc. Accessed: 18 November 2008 <www.agmrc.org>/NR/rdonlyres/4EE0E81C-C607-4C3F-BBCF-B75B7395C881/0/ksupipelineethl.pdf>

CHAPTER 3:

Construction of *Saccharomyces cerevisiae* strains co-expressing cellulase genes for efficient hydrolysis of amorphous cellulose

(Article to be submitted for publication in
Applied Microbiology and Biotechnology)

Construction of *Saccharomyces cerevisiae* strains co-expressing cellulase genes for efficient hydrolysis of amorphous cellulose

L. du Plessis • S.H. Rose • W.H. van Zyl*

*Department of Microbiology, University of Stellenbosch, Stellenbosch 7600, South Africa

ABSTRACT

The endoglucanase I and II genes (*egI* or *Cel5A* and *egII* or *Cel7B*) of *Trichoderma reesei* QM6a were successfully cloned and expressed in *Saccharomyces cerevisiae* under transcriptional control of the yeast *ENO1*-promotor and terminator sequences. Activity expressed by the *egI* gene was improved two-fold by random mutagenesis. Both endoglucanase genes were co-expressed with the synthetic, codon optimized cellobiohydrolase gene (*sCBHI*) from *T. reesei*, as well as a β-glucosidase (*bgl1*) from *Saccharomycopsis fibuligera* in *S. cerevisiae*. Levels of endoglucanase activity were lower when co-expressed with *sCBHI* or *bgl1*, suggesting a metabolic burden on the cells. Recombinant strains were able to hydrolyze phosphoric acid swollen cellulose (PASC), generating mainly cellotriose, cellobiose and glucose. A bottleneck was observed at the conversion from cellobiose to glucose, suggesting the β-glucosidase activity to be the rate-limiting factor. As a consequence, the recombinant strains were unable to produce enough glucose for growth on cellulose. The results of this study provide insight into further optimization of recombinantly expressed cellulase combinations for saccharification and fermentation of cellulose to ethanol.

1. INTRODUCTION

The depletion of the world's supply of non-renewable fuel feedstocks, rising oil prices and the negative impact fossil fuel emissions has on the environment, has led to an increasing need for the development

of an alternative renewable transportation fuel (Agbogbo and Coward-Kelly 2008, Kristensen *et al.* 2008). Bioethanol presents an attractive alternative, since the carbon dioxide released during combustion is recycled and converted into biomass through photosynthesis (Fujita *et al.* 2004). Lignocellulosic biomass has received considerable interest as a resource for the production of bioethanol due to its abundance and low cost (Lynd *et al*. 2005).

Plant biomass consists of 40-55% cellulose, 25-50% hemicellulose and 10-40% lignin, depending on the source (Sun and Cheng 2002). Cellulose presents the major source of fermentable sugars and is degraded through the concerted action of three types of enzymes: endoglucanases, exoglucanases (mainly cellobiohydrolases) and β-glucosidases (Den Haan *et al.* 2007b). Endoglucanases hydrolyze glycosidic bonds within internal amorphous sites in the cellulose polysaccharide chain, generating oligosaccharides of various lengths (Fujita *et al.* 2004). Cellobiohydrolases degrade the oligosaccharides from the reducing- or non-reducing ends leaving cellobiose as main end product. β-glucosidases hydrolyze soluble cellodextrins and cellobiose, generating glucose as the main end product.

The filamentous ascomycete, *Trichoderma reesei* (*Hypocrea jecorina*), is known to secrete a range of cellulases and hemicellulases required for degradation of lignocellulosic substrates (Sandgren *et al.* 2005). This fungus, however, does not have the ability to ferment the resulting sugars to ethanol. The yeast *Saccharomyces cerevisiae*, on the other hand, is known to produce ethanol and has a high ethanol tolerance. Other advantages to this host include its GRAS (Generally Regarded As Safe) status, the ease with which it can be genetically manipulated, the vast amount of information available regarding its molecular biology, its robustness in industrial processes, etc. (Den Haan *et al.* 2007b). Although *S. cerevisiae* cannot utilize cellulosic material, it has natural fermentation capability, making it an ideal host for expression of recombinant cellulase enzymes. The feasibility of enzymatic conversion of cellulose to ethanol depends on the cost-effectiveness of the process (Lynd *et al.* 2005). Therefore, a need arose to develop an organism with the ability to enzymatically degrade lignocellulose and ferment the resulting sugars to ethanol in a single step, called consolidated bioprocessing (CBP).

Five endo-β-1,4-glucanases (encoded by *egI, egII, egIII, egIV* and *egV*), two cellobiohydrolases (encoded by *cbhI* and *cbhII*) and a β-glucosidase (*bgl1*) from *T. reesei* have been efficiently secreted

into the culture medium by *S. cerevisiae* transformants (Van Rooyen *et al.* 2005, Lynd *et al.* 2002). Synergistic expression of cellulase enzymes for degradation of cellulosic substrates has been demonstrated by several studies (Fujita *et al.* 2004, 2002, Medve *et al.* 1998, Okada *et al.* 1998, Kleman-Leyer *et al.* 1996, Bailey *et al.* 1993, Woodward *et al.* 1988), but growth on cellulose has only recently been demonstrated (Den Haan *et al.* 2007b). The multicopy expression of the *egI* from *T. reesei* and the *bgl1* of *Saccharomycopsis fibuligera* by a recombinant strain of *S. cerevisiae,* enabled growth on an amorphous cellulosic substrate (phosphoric acid swollen cellulose, PASC) and the subsequent production of ethanol.

In this study we expressed the *T. reesei* endoglucanase genes, *egI* (*Cel7B*) and *egII* (*Cel5A*), in the yeast strain *S. cerevisiae* Y294 to determine efficiency of heterologous expression and levels of extracellular activity displayed by this host. We also aimed to improve the heterologous EGI activity produced by *S. cerevisiae* Y294 through random mutagenesis, using an *E. coli* mutator strain. Finally, we attempted simultaneous and synergistic saccharification of amorphous cellulose to enable growth on this substrate for eventual production of ethanol as end product, using recombinant yeast strains producing the *T. reesei* endoglucanases EGI and EGII, together with the synthetic codon optimized cellobiohydrolase I (sCBHI) and the β-glucosidase (BGL1) from *S. fibuligera*.

2. MATERIALS AND METHODS

2.1 Media and cultivation

All chemicals used were of analytical grade. *Escherichia coli* DH 5α (Takara Bio Inc.) was used as the host strain for the recombinant DNA manipulations and plasmid propagation. Bacterial cultivation (37°C) took place in Terrific Broth (12 $g.l^{-1}$ tryptone, 24 $g.l^{-1}$ yeast extract, 4 $ml.l^{-1}$ glycerol, 100 $ml.l^{-1}$ of phosphate buffer) containing 100 $\mu g.ml^{-1}$ ampicillin (Sambrook *et al.* 1989). The yeast strains *S. cerevisiae* Y294 and *S. cerevisiae* Y294[BGL1] were used as host strains for expression of the *T. reesei* endoglucanases (*egI* and *egII*) and synthetic cellobiohydrolase (*sCBHI*) genes. *S. cerevisiae* strains were aerobically cultivated on a rotary shaker (100 rpm) at 30°C in 125 ml Erlenmeyer flasks containing 30 ml synthetic complete (SC) medium (1.7 $g.l^{-1}$ yeast nitrogen base without amino acids and ammonium sulphate (Difco laboratories), 5 $g.l^{-1}$ $(NH_4)_2SO_4$, 20 $g.l^{-1}$ glucose and supplemented

with appropriate amino acids). Yeast strains were maintained on agar (20 $g.l^{-1}$) plates of the same composition. Autoselective *S. cerevisiae* strains were cultured under the same conditions in YPD medium (10 $g.l^{-1}$ of yeast extract, 20 $g.l^{-1}$ peptone and 20 $g.l^{-1}$ glucose). *S. cerevisiae* Y294[BGL1] were selected and maintained on YPC plates (10 $g.l^{-1}$ yeast extract, 20 $g.l^{-1}$ peptone, 20 $g.l^{-1}$ cellobiose (Sigma) and 20 $g.l^{-1}$ agar). All strains were tested for growth on YP-PASC agar plates (10 $g.l^{-1}$ yeast extract, 20 $g.l^{-1}$ peptone, 10 $g.l^{-1}$ PASC and 20 $g.l^{-1}$ agar) and SC-PASC agar plates. Phosphoric acid swollen cellulose (PASC) was prepared from Avicel PH-101 (Fluka) as described by Den Haan *et al.* (2007b). For enzymatic assays, stationary phase pre-cultures were used to inoculate the medium to approximately 1 x 10^6 $cells.ml^{-1}$. Samples were taken periodically to determine the optical density and dry cell weight (Den Haan *et al.* 2007b).

2.2 Strains and plasmids

The relevant genotypes of fungal, bacterial and yeast strains, as well as plasmids used and constructed in this study, are listed in Table 1.

Table 1: Microbial strains and plasmids used in this study.

Strains or plasmids	**Genotype**	**Source/Reference**
Strains:		
Trichoderma reesei QM6a	Wild-type	ATCC 13631
Escherichia coli DH5α	*supE44 ΔlacU169 (ϕ80lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Sambrook *et al.* (1989)
Escherichia coli ES1301 *mutS*	*lacZ*53 *mutS*201::Tn*5 thyA*36 *rha*-5 *metB*1 *deoC* IN(*rrnD-rrnE*)	Promega Corporation
Saccharomyces cerevisiae Y294	*α leu2*-3,112 *ura3*-52 *his3 trp1-289*	ATCC 201160
Saccharomyces cerevisiae Y294[BGL1]	*α leu2*-3,112 *ura3*-52 *his3 trp1-289,* multicopy genomic integration of *bgl1* of *Saccharomycopsis fibuligera*	This laboratory
***S. cerevisiae* Y294:**		
[*fur1::LEU2* yENO1]	*bla ura3/URA3 $ENO1_P$-$ENO1_T$*	Den Haan *et al.* (2007b)
[*fur1::LEU2* pLEGII]	*bla ura3/URA3 $ENO1_P$-egII-$ENO1_T$*	This work
[*fur1::LEU2* pLEGIIC]	*bla ura3/URA3 $ENO1_P$-egII-$ENO1_T$; $ENO1_P$ -sCBHI-$ENO1_T$**	This work
[*fur1::LEU2* pAZ40]	*bla ura3/URA3 $ENO1_P$-egI-$ENO1_T$*	Den Haan *et al.* (2007b)
[*fur1::LEU2* pLEM1]	*bla ura3/ URA3 $ENO1_P$-egI-$ENO1_T$*	This work

Table 1: Microbial strains and plasmids used in this study (continue).

***S. cerevisiae* Y294:**		
[*fur1::LEU2* pLEM1C	*bla ura3/URA3 ENO1*$_P$*-egI-ENO1*$_T$*;*	This work
	ENO1$_P$ *-sCBHI-ENO1*$_T$*	
pMEG1	*bla ura3/URA3 ENO1*$_P$*-egI-ENO1*$_T$)	This work
pAZ1	*bla ura3/URA3 ENO1*$_P$*-xyn2-ENO1*$_T$	This laboratory
pMPxyn	*bla ura3/URA3 ENO1*$_P$*-xyn2-ENO1*$_T$	This work
***S. cerevisiae* Y294[BGL1]:**		
[*fur1*::*LEU2* yENO1]	*bla ura3/URA3 ENO1*$_P$*-ENO1*$_T$	Den Haan *et al.* (2007b)
[*fur1*::*LEU2* pLEGII]	*bla ura3/URA3 ENO1*$_P$*-egII-ENO1*$_T$	This work
[*fur1*::*LEU2* pLEGIIC]	*bla ura3/URA3 ENO1*$_P$*-egII-ENO1*$_T$ *;*	This work
	ENO1$_P$ *-sCBHI-ENO1*$_T$*	
[*fur1*::*LEU2* pAZ40]	*bla ura3/URA3 ENO1*$_P$*-egI-ENO1*$_T$	Den Haan *et al.* (2007b)
[*fur1*::*LEU2* pLEM1]	*bla ura3/ URA3 ENO1*$_P$*-egI-ENO1*$_T$	This work
[*fur1*::*LEU2* pLEM1C]	*bla ura3/ URA3 ENO1*$_P$*-egI-ENO1*$_T$;	This work
	ENO1$_P$ *-sCBHI-ENO1*$_T$*	
Plasmids:		
pDrive	*bla*	Qiagen
pDrive-egII	*bla, egII*	This work
yENO1	*bla URA3 ENO1*$_P$*-ENO1*$_T$	Den Haan *et al.* (2007b)
YEpENOB-sCBHI	*bla URA3 ENO1*$_P$ *-sCBHI-ENO1*$_T$	This laboratory
pLEGII	*bla URA3 ENO1*$_P$*-egII-ENO1*$_T$	This work
pLEGIIC	*bla URA3 ENO1*$_P$*-egII-ENO1*$_T$ *;*	This work
	ENO1$_P$ *-sCBHI-ENO1*$_T$*	
pAZ40	*bla URA3 ENO1*$_P$*-egI-ENO1*$_T$	Den Haan *et al.* (2007b)
pLEM1	*bla URA3 ENO1*$_P$*-egI-ENO1*$_T$	This work
pLEM1C	*bla URA3 ENO1*$_P$*-egI-ENO1*$_T$ *;*	This work
	ENO1$_P$ *-sCBHI-ENO1*$_T$*	
pMEG1	*bla URA3 ENO1*$_P$*-egI-ENO1*$_T$)	This work
pAZ1	*bla URA3 ENO1*$_P$*-xyn2-ENO1*$_T$	This laboratory
pMPxyn	*bla URA3 ENO1*$_P$*-xyn2-ENO1*$_T$	This work
pDF1	*bla fur1*::*LEU2*	La Grange *et al.* (1996)

**ENO1*$_P$*-sCBHI-ENO1*$_T$ has been designated synCBHI
The mutated DNA fragments are underlined.

2.3 Construction of plasmids

Standard protocols were followed for DNA manipulations (Sambrook *et al.* 1989). The enzymes used for restriction digests and ligations were purchased from Inqaba Biotec and used as recommended by

the supplier. Digested DNA was eluted from agarose gels using the Zymoclean TM Gel Recovery Kit (Zymo Research). PCR reactions were carried out with a Perkin Elmer GeneAmp® PCR System 2400 using *Taq* DNA polymerase (Inqaba Biotec) according to the suppliers specifications. Table 2 contains a list of all the primers (Whitehead Scientific) used in the study.

Table 2: PCR primers used in this study for the cloning, sequencing and confirmation of positive clones. Restriction sites are underlined.

Primer name	**Sequence (5'-3')**	**Source accession nr.**
ENO1$_{PT}$		X99228
ENO1-L	GGATCCACTAGTCTTCTAGGCGGGTTATC	
ENO1-R	GGATCCAAGCTTGCGGCCGCAAAGAGGTTTAGACATTGG	
egI		AB003694
TREGI-L	GATCGAATTCAATGGCGCCCTCAGTTACAC	
TREG1-R	GTACAGATCTAGTCAACGCTCTAAAGGCATTG	
egII		M19373
TREGII-L	ATAGATCTGCGGCCGCTGCTCGATAATATCTCCC	
TREGII-R	GATCCTCGAGGCTCAGAGTGCTACTTTC	
TREGII-Overlap	AGGTAACGCAAGTGCCATCTGTGGTACAGCCAAA	
sCBHI		
sCBHI-L	GACTGAATTCATAATGGTCTCCTTCACCTCC	AAB50278
sCBHI-R	GACTCTCGAGTTACAAACATTGAGAGTAGTATGG	CAH10320
FUR1 disruption		M36485
FUR1-L	TCCGTCTGGCATATCCTA	
FUR1-R	TTGGCTAGAGGACATGTA	
Real-time PCR		
TRP1-L	CAATGATTACGGCATTGATATCG	YDR007W
TRP1-R	CTGAATCAAACAAGGGAATAAACG	
URA3-L	CAGGCGGCAGAAGAAGT	YEL021W
URA3-R	CGTAACCTTCATCTCTTCCAC	

The genomic copy of *egII* was isolated from *T. reesei* and the intron removed by means of overlap PCR. The *egII* cDNA was amplified and ligated into the pDrive-plasmid (Inqaba Biotec). Plasmid DNA was isolated, purified and DNA sequences of *egII* confirmed by means of automated sequencing. The plasmid pDrive-egII was enzymatically digested with *Eco*RI (Roche). The *egII* fragment was isolated and ligated into the corresponding *Eco*RI site in the *S. cerevisiae–E. coli* shuttle vector yENO1 (Den Haan *et al.* 2007b), generating pLEGII. The synthetic cellobiohydrolase I gene cassette (synCBHI) was retrieved from YEpENOB-sCBHI and cloned into the *Bam*HI site of pLEGII, now

designated pLEGIIC. The *egI* gene was previously cloned into the *Eco*RI and *Bgl*II sites of yENO1 under the control of the *ENO1* promoter and terminator to generate pAZ40 (Den Haan *et al.* 2007b). Plasmid pAZ40 was subjected to mutagenesis and designated pLEM1. Following partial digestion of pLEM1 with *Bam*HI, synCBHI was cloned into this site generating pLEM1C. After digestion of pLEM1 with *Eco*RI and *Xho*I the mutated *egI* fragment was ligated into the yENO1 plasmid, digested with the same enzymes. The resulting plasmid is referred to as pMEG1. pLEM1 was also digested with *Bam*HI and *Eco*RI, retrieving the mutated *ENO1* promoter from the plasmid. This fragment was cloned into pAZ1, to place the expression of the endo-xylanase gene (*xyn2* of *T. reesei*) under the control of the mutated *ENO1* promoter. The plasmid was designated pMPxyn. The relevant plasmid maps are shown in Figure 1.

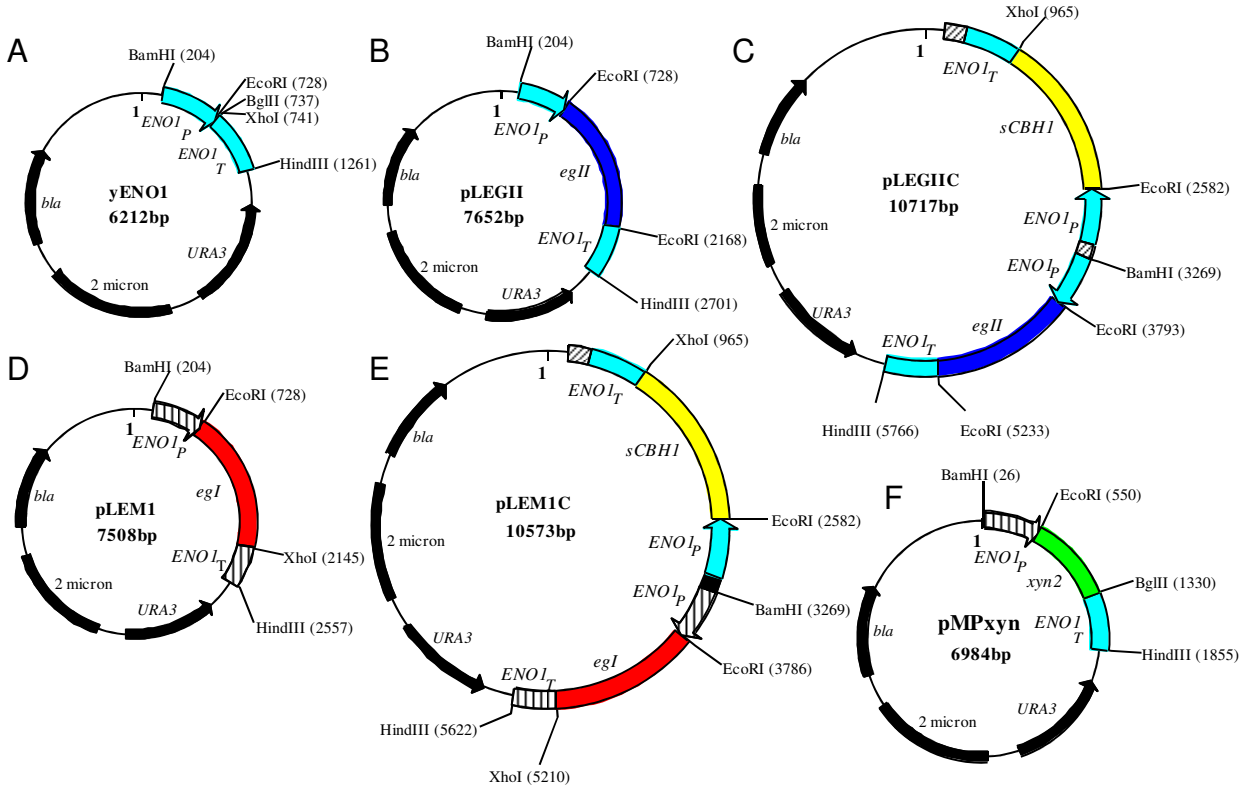

Figure 1: Plasmid maps of vectors used in this study including the orientation of cloned genes and relevant restriction enzyme positions. pLEGII (B) was constructed from yENO1 (A) by cloning the *egII* gene into the *ENO1*-promotor-terminator cassette. Insertion of synCBH yielded pLEGIIC (C). Insertion of the same synCBH fragment into the mutated *egI*-containing plasmid pLEM1 (D), generated pLEM1C (E). Cloning of the mutated *ENO1* promoter fragment into pAZ1 resulted in pMPxyn (F).

2.4 Yeast transformation

Yeast strains *S. cerevisiae* Y294 and Y294[BGL1] were transformed with the recombinant plasmids using the lithium acetate/DMSO method described by Hill *et al.* (1991). Transformants were selected for growth on SC^{-URA} plates. Autoselective strains were constructed by subsequent transformation with pDF1 (La Grange *et al.* 1996). Disruption of the *FUR1* gene results in maintenance of the *URA3*-bearing expression vector under non-selective conditions (Kern *et al.* 1990).

2.5 Mutagenesis

The plasmid pAZ40 was transformed to the DNA mismatch repair deficient strain, *E. coli* ES1301 *mut*S (Promega). Colonies were pooled and propagated overnight in liquid medium. The plasmid DNA was isolated and transformed to *S. cerevisiae* Y294. Transformants were screened for higher levels of enzyme activity using the reducing sugar assay (Miller 1959).

2.6 Plate assays

Recombinant yeast strains were screened for endoglucanase activity by transferring stationary phase cells onto selective agar plates containing 0.1 % (w/v) Ostazin Brilliant Red-hydroxyethyl cellulose (OBR-HEC) (Sigma). Cells were incubated at 30ºC for 48 hours. Extracellular endoglucanase activity was detected as a clearing zone surrounding the colonies.

2.7 Liquid enzyme activity assays

Enzyme activity was assessed colorimetrically by measuring the amount of reducing sugars released using dinitrosalicylic acid (DNS) (Miller 1959). One unit is determined as 1 µmole reducing sugar liberated per minute using glucose or xylose as standard. The substrate used for the liquid endoglucanase assays was 1% (w/v) CMC (Sigma) or 0.1% (w/v) barley β-glucan (Sigma) resuspended in 0.05 M sodium citrate buffer (pH 5). One percent (w/v) birchwood xylan (Roth) in 0.05 M citrate buffer (pH 5) was used to measure the endo-xylanase activity. All assays were performed in triplicate over a time period of 6 days.

2.8 Characterization of mutated EGI enzyme

The pH optimum was determined according to La Grange *et al.* (1996) using 1% CMC at 60°C. The substrate was buffered at pH 4, pH 4.5, pH 5, pH 5.5, pH 6 and pH 6.5 using 0.05 M sodium citrate buffer. The temperature optimum was determined using 1% CMC in 0.05 M citrate buffer (pH 5) at 50°C, 55°C, 60°C, 65°C and 70°C.

2.9 Isolation of plasmid DNA from *S. cerevisiae*

Plasmid DNA was extracted from the yeast, according to Hoffman and Winston (1987). The plasmid DNA was transformed to *E. coli* DH 5α and propagated for further analysis.

2.10 Relative copy number determination

Total DNA was extracted from *S. cerevisiae* Y294[pAZ40] and *S. cerevisiae* Y294[pLEM1] and used to determine and compare copy number by real-time polymerase chain reaction (RT-PCR). Two primer sets specific to the uracil encoding gene (*URA3*) and the tryptophan encoding gene (*TRP1*) were designed. The *URA3* is a single-copy gene in both plasmids pAZ40 and pLEM1, as well as in the *S. cerevisiae* chromosomal DNA. The *TRP1* is a single-copy gene on the *S. cerevisiae* genome. The primer sets were designed using Primer3 software [Rosen and Skaletsky 2000] based on the *S. cerevisiae URA3* and *TRP1* sequences. The sequences of the primers are shown in Table 2. The concentration of the total DNA extracted was measured with the NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, USA) and normalized to $2ng.\mu l^{-1}$ with deionised and distilled water (ddH_2O). Real-time PCR amplification and analysis was performed using a LightCycler instrument, with software version 4.05 (Roche Diagnostics), according to the manufacturers' specifications. Relative quantification presents the amount of target gene in a sample relative to another sample (i.e., a calibrator), which contains both target and reference genes with a constant ratio. In this experiment, *URA3* and *TRP1* acted as the target and reference genes, respectively.

2.11 Automated sequencing

Pure plasmid DNA was isolated from *E. coli* with the Zyppy Plasmid Miniprep Kit (Zymo Research). Sequence verification was done after sequence determination by the dideoxy chain termination method, with an ABI PRISM^{TM} 3100 Genetic Analyzer, using primers indicated in Table 2. The sequences were analyzed with the PC based DNAMAN (version 4.1, from Lynnon BioSoft) and Chromas (version 1.45).

2.12 Product formation

After 5 days of growth in YPD, cultures were filtered through Whatman glass microfibre filterpaper. PASC and sodium azide was added to 130 ml yeast filtrates to yield a final concentration of 1% and 0.05% respectively. Cellulase adsorption on PASC occurred during overnight incubation with stirring at 4°C. Cellulose containing the adsorbed enzyme was captured onto a Whatman filter paper. The filter paper was transferred to a capped tube containing 6 ml of citrate buffer (pH 5) and 0.05% sodium azide. A sample of one ml was taken at time zero. Degradation of PASC was allowed to continue for four hours on a roller at 37°C after which another sample was taken. Unhydrolyzed substrate was filtered through a 0.22 µm filter. The soluble sugars (glucose, cellobiose and cellotriose) were analyzed on a Dionex HPLC system with an anion exchange Carbopac PA100 column and pulsed amperometric detection. Glucose, cellobiose (Sigma) and cellotriose (Sigma) were used as standards.

3. RESULTS

3.1 Cloning and expression of cellulases in *S. cerevisiae*

The *T. reesei egII* and mutated *egI* genes were cloned onto multi-copy, episomal vectors under transcriptional control of the constitutive *S. cerevisiae* enolase I gene (*ENO1*) promoter (Table 1). These plasmids were transformed to both *S. cerevisiae* Y294 and *S. cerevisiae* Y294[BGL1] (which contained the *S. fibuligera bgl1* gene integrated into the genome). Presence of the endoglucanases was confirmed by means of plate assays. The synthetic *sCBHI* gene (codon optimized for expression in *S. cerevisiae*) expressed under the control of the *ENO1* promoter, was cloned into the two plasmids

mentioned above for concurrent expression in both *S. cerevisiae* strains. PCR analysis was performed to confirm the presence of the genes using total DNA as template. Autoselective strains were constructed by disrupting the *FUR1* gene of all the *S. cerevisiae* strains using plasmid pDF1 (La Grange *et al.* 1996, Kern *et al.* 1990). The disruption was confirmed by PCR amplification of the *fur1::LEU2* DNA fragment using extracted genomic DNA (data no shown). The yeast strain containing the yENO1 plasmid acted as a negative control strain.

3.2 Enzyme activity measurements

3.2.1. Endoglucanase II

Recombinant yeast strains were screened on OBR-HEC plates for secreted endoglucanase II activity. OBR-HEC hydrolysis was detected by the appearance of clear zones in the red agar (Figure 2).

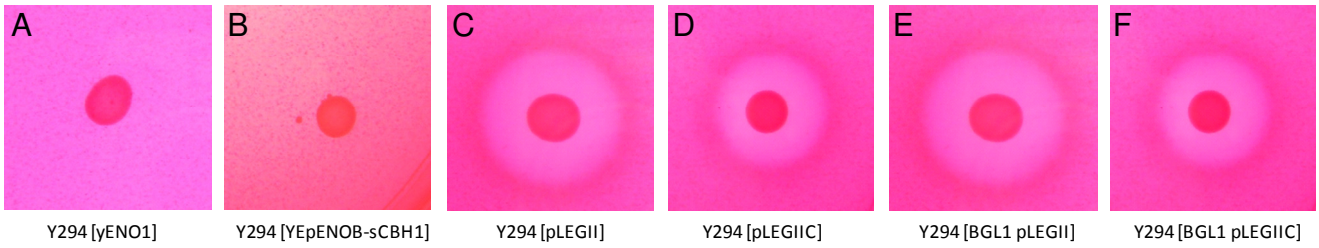

Figure 2: Recombinant *S. cerevisiae* Y294 and Y294[BGL1] strains displaying extracellular EGII activity (C-F) on SC^{-URA}-agar plates containing OBR-HEC after 48 hours of incubation at 30°C. Strains Y294[yENO1] (A) and Y294[YEpENOB-sCBH1] (B) did not display hydrolysis, thus lacking EGII activity.

The activity levels of the secreted recombinant endoglucanase enzymes were initially quantified in SC^{-URA} medium. Activity is expressed in katals per ml, with 1 katal being the amount of enzyme needed to produce one mole of reducing sugar from the substrate per second under the assay conditions (Bailey *et al.* 1992). EGII displayed no detectable activity (when the strains were cultivated in SC^{-URA}) towards the synthetic substrate Carboxymethyl Cellulose (CMC) (data not shown), yet some activity was detected towards barley β-glucan (Figure 3). However, cultivation of the *S. cerevisiae* strains in YPD medium resulted in similar levels of activity for both CMC and barley β-glucan, the highest being 19.6 $nkat.ml^{-1}$ and 22.3 $nkat.ml^{-1}$ respectively.

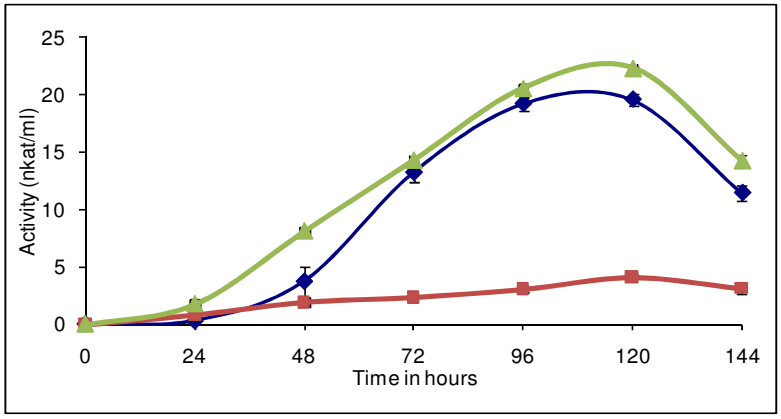

Figure 3: Time course of EGII activity towards barley-β-glucan produced by ■ *S. cerevisiae* Y294[pLEGII] cultivated in SC^{-URA} and ▲ *S. cerevisiae* Y294[*fur1::LEU2* pLEGII] cultivated in YPD, as well as EGII activity towards CMC produced by ♦ *S. cerevisiae* Y294[*fur1::LEU2* pLEGII) cultivated in YPD. Assays were performed in triplicate and error bars represent the standard deviation. Values have been normalized by deducting negative cotrol values.

4.2.2 Endoglucanase I

The secreted endoglucanase I activity produced by the different *S. cerevisiae* strains used in this study was observed as halos on OBR-HEC plates (Figure 4).

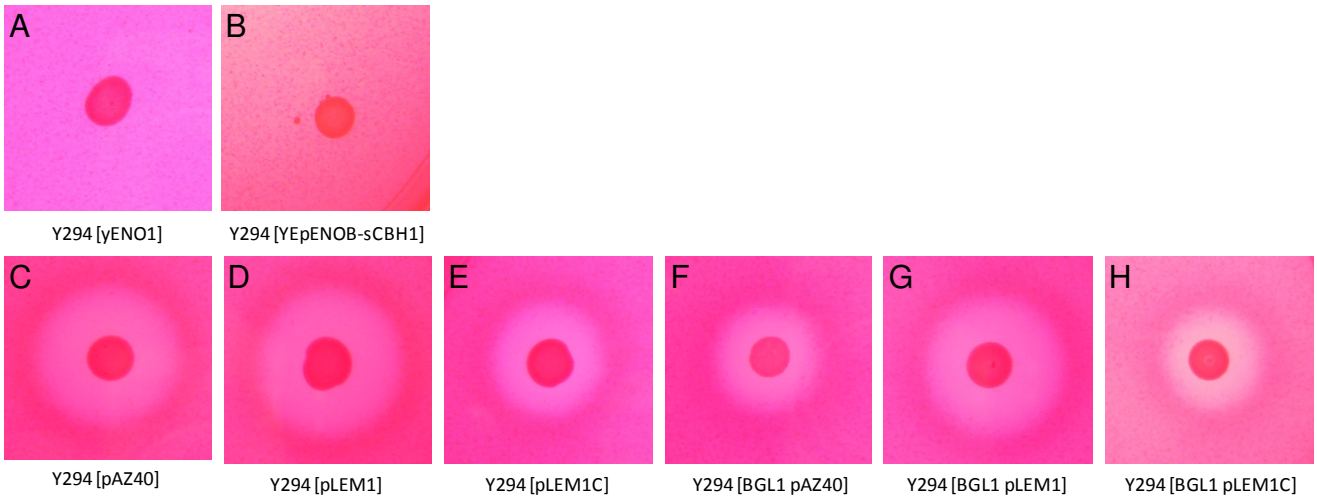

Figure 4: EGI activity (C-H) of recombinant *S. cerevisiae* Y294 and Y294[BGL1] strains on OBR-HEC-containing SC^{-URA}-agar plates after 48 hours of incubation at 30°C. Y294[yENO1] (A) and Y294[YEpENOB-sCBH1] (B) did not show any extracellular EGI activity.

After subjecting the pAZ40 plasmid to mutagenesis, it was transformed to *S. cerevisiae* Y294. Liquid assays were performed in order to screen for a transformant displaying improved levels of EGI activity. After screening approximately 200 yeast colonies, one such strain was identified. This strain was consequently designated *S. cerevisiae* Y294[pLEM1]. Maximum activity of *S. cerevisiae* Y294[pLEM1] was reached after 120 hours at 21.8 nkat.ml^{-1}, which is 2.1 times more than the highest level obtained with *S. cerevisiae* Y294[pAZ40] (10.3 nkat.ml^{-1}) (Figure 5). The mutated *egI* gene expressed under the control of the original *ENO1* promoter and terminator cassette, *S. cerevisiae* Y294[pMEG1], reached 16.9 nkat.ml^{-1}, which is still 1.6 times more activity than that obtained by *S. cerevisiae* Y294[pAZ40], suggesting an altered gene sequence. Activity levels per dry cell weight confirmed improved levels (15.4 nkat.g^{-1} DCW on day four) of EGI activity for strain *S. cerevisiae* Y294[pLEM1], compared to *S. cerevisiae*[pAZ40] (7.6 nkat.g^{-1} DCW).

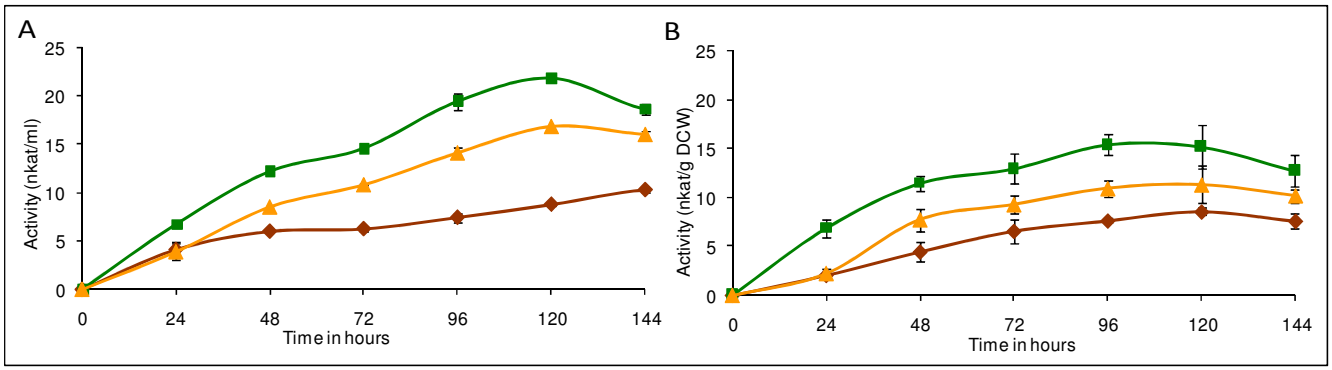

Figure 5: Comparison of extracellular EGI activity towards CMC produced over a period of 144 hours by ♦ *S. cerevisiae* Y294[pAZ40], ■ *S. cerevisiae* Y294[pLEM1], ▲ *S. cerevisiae* Y294[pMEG1] cultivated in SC^{-URA}, expressed as (A) total secreted endoglucanase activity and (B) endoglucanase activity per dry cell weight (DCW). Assays were performed in triplicate and error bars represent the standard deviation. Values have been normalized by deducting the negative control values.

Higher levels of activity were obtained by strains cultivated in YPD (Figure 6). *S. cerevisiae* Y294[pLEM1] peaked at 25.7 nkat.ml^{-1} after 120 hours of growth. This is a 1.7 times improvement on the endoglucanase activity. However, determined per dry cell weight, activity levels were lower than that obtained in synthetic media, the highest level reached by *S. cerevisiae* Y294[pLEM] (7.8 nkat.g^{-1} DCW), a 1.5 times improvement.

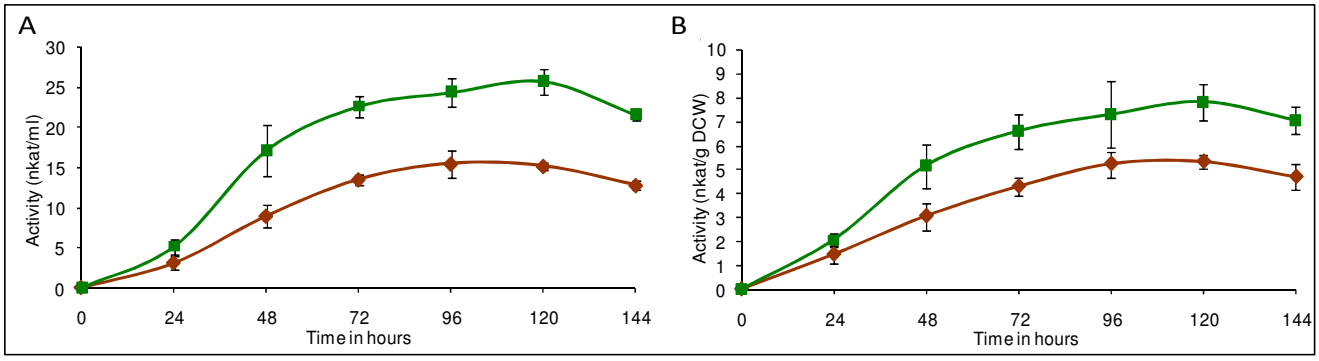

Figure 6: Extracellular EGI activity towards CMC produced over a period of 144 hours by ♦ *S. cerevisiae* Y294[*fur1::LEU2* pAZ40] and ■ *S. cerevisiae* Y294[*fur1::LEU2* pLEM1] cultivated in YPD, expressed as (A) total endoglucanase activity and (B) endoglucanase activity per dry cell weight. Assays were performed in triplicate and values have been normalized by deducting the negative contol values. The error bars represent the standard deviation.

Reducing sugar assays were performed to determine the effect of the mutated *ENO1* promoter on endo-xylanase activity, comparing levels produced by *S. cerevisiae* Y294[pAZ1] to that of *S. cerevisiae* Y294[pMPxyn] (Figure 7). The latter proved to display higher levels, suggesting an altered *ENO1* promoter sequence. These results imply that mutations in the promoter region are contributing to the overall higher levels of endoglucanase activity produced by *S. cerevisiae* Y294[pLEM1].

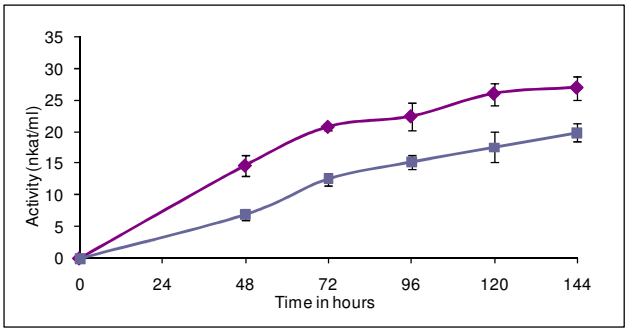

Figure 7: Time course of endo-xylanase activity towards birchwood xylan over a period of 144 hours by ■ *S. cerevisiae* Y294[pAZ1] and ♦ *S. cerevisiae* Y294[pMPxyn]. Assays were performed in triplicate and values have been normalized by deducting the negative control values. The error bars represent the standard deviation.

3.3 Characterization of mutated EGI

The recombinant EGI activity, produced by all strains, peaked at pH 4.5 to 5 (Figure 8). The optimum temperature of the mutated EGI remained unchanged, with the highest level of activity obtained at

60°C. Thus, although the mutations in the endoglucanase gene caused higher levels of extracellular activity, the pH and temperature optima were unaffected.

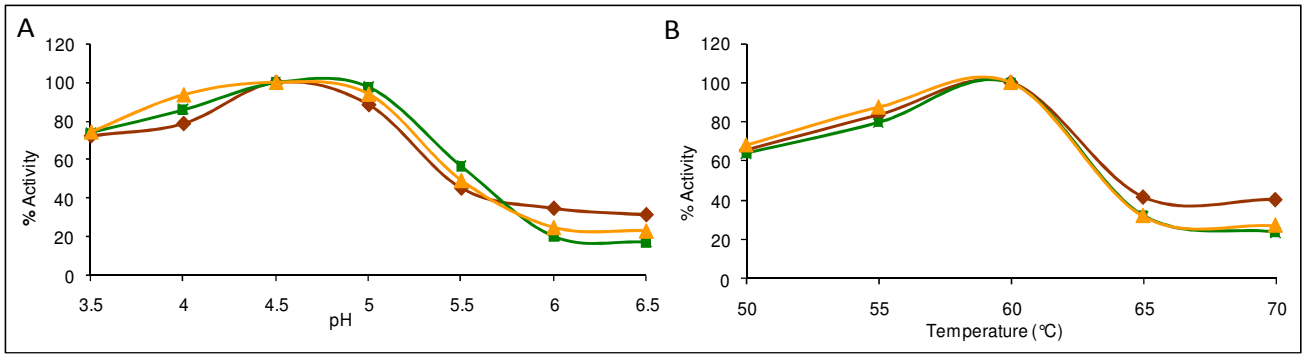

Figure 8: Effect of (A) pH and (B) temperature on the activity of the EGI produced by ■ *S. cerevisiae* Y294[pLEM1], ▲ *S. cerevisiae* Y294[pMEG1] and ♦*S. cerevisiae* Y294[pAZ40]. Activity is expressed as a percentage of the highest value.

3.4 Relative copy number determination

The entire pAZ40 plasmid was amplified and mutated by the mutator strain. Therefore, it is possible that mutations were induced in other regions on the plasmid (apart from the *egI* gene and promoter), such as the 2µ fragment. As the 2µ is the major determinant of the copy number of the plasmid in the yeast, real-time PCR was used to determine the relative copy number of pAZ40 and pLEM1 in *S. cerevisiae* transformants. The *URA3* and *TRP1* genes were used as the target and reference genes, respectively. The results are shown in Table 3. The ratio of the Ct value of *URA3* and *TRP1* is similar in the two *S. cerevisiae* strains, *S. cerevisiae* Y294[pAZ40] and *S. cerevisiae* Y294[pLEM1].

Table 3: Ratio of threshold values (Ct) of the *URA3* target gene to the *TRP1* reference gene.

Strain	***URA3*^{Ct}:*TRP1*^{Ct}**
S. cerevisiae Y294[pAZ40]	1:1.38 ± 0.03
S. cerevisiae Y294[pLEM1]	1:1.37 ± 0.05

3.5 Automated sequencing

Sequencing of the mutated promoter on plasmid pLEM1 revealed three point mutations: one A to G transition and two T to C transitions (Figure 9). The binding site for the RNA polymerase II is

unaffected by the changes in the promoter sequence. In the *egI* gene only one A to G transition mutation was detected, which was located in the carbohydrate binding module (Figure 10). The deduced amino acid sequence revealed an amino acid change from a tyrosine to a cysteine.

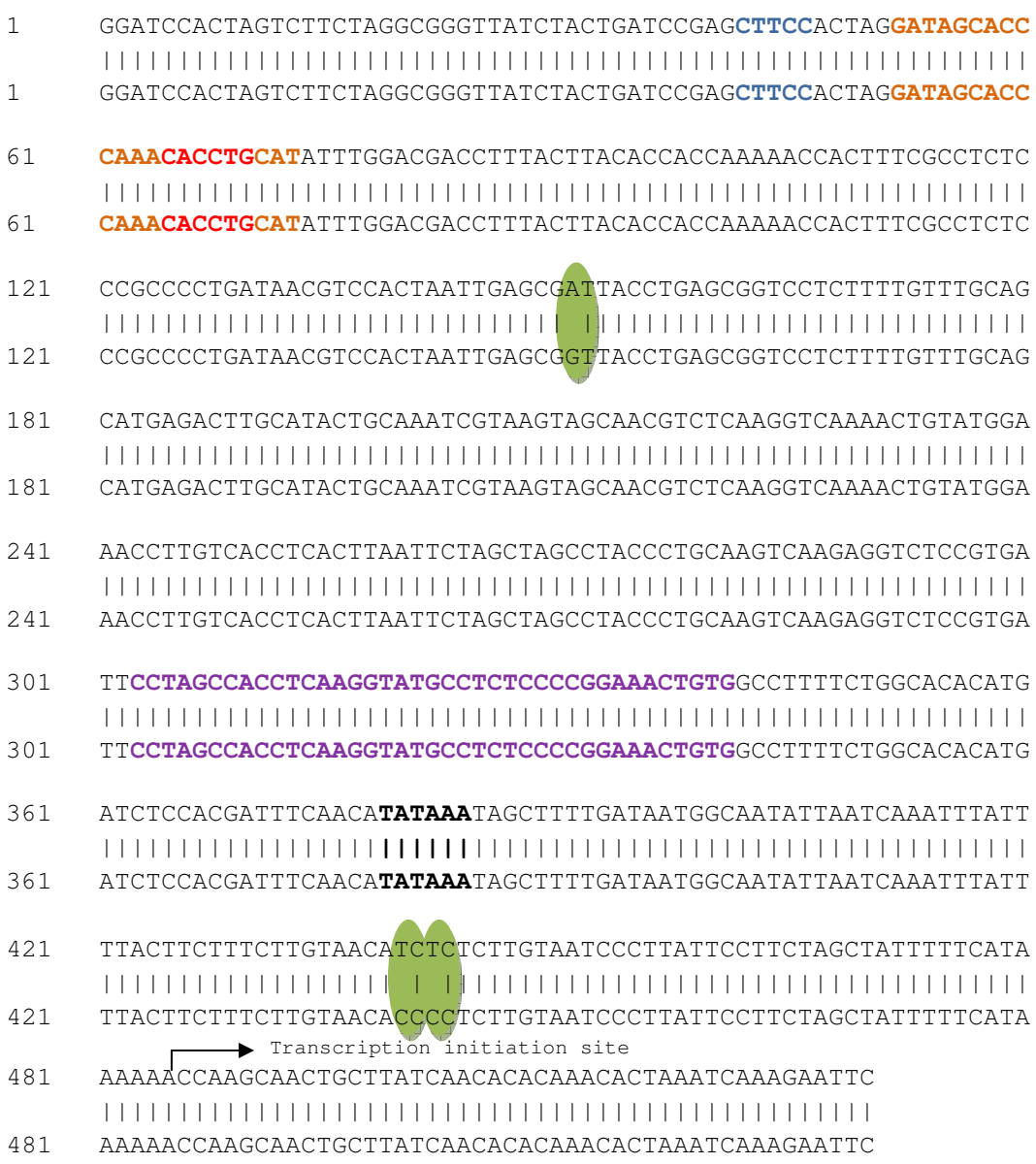

Figure 9: The sequence alignment of the *ENO1* promoter regions of pAZ40 (upper line) and pLEM1 (lower line) revealed a 99.4% identity. Three single base mutations were identified (indicated by green circles). The TATA box where RNA polymerase II binds during the assembly of the transcription initiation complex is indicated in bold type. The E box (5'-CANNTG-3') where the transcriptional regulator Sgc1p binds is indicated in red, and is located in the upstream activation sequence (UAS) (orange) (Chen and Lopes 2007, Machida *et al.* 1989). The Gcr1p protein, required for high-level glycolytic gene expression in *S. cerevisiae*, interacts with the CTTCC sequence motif (blue) (Huie *et al.* 1992, Cohen *et al.* 1987). The upstream repressor sequence is indicated in purple.

```
1      ATGGCGCCCTCAGTTACACTGCCGTTGACCACGGCCATCCTGGCCATTGCCCGGCTCGTC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1      ATGGCGCCCTCAGTTACACTGCCGTTGACCACGGCCATCCTGGCCATTGCCCGGCTCGTC

61     GCCGCCCAGCAACCGGGTACCAGCACCCCCGAGGTCCATCCCAAGTTGACAACCTACAAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61     GCCGCCCAGCAACCGGGTACCAGCACCCCCGAGGTCCATCCCAAGTTGACAACCTACAAG

121    TGTACAAAGTCCGGGGGGGTGCGTGGCCCAGGACACCTCGGTGGTCCTTGACTGGAACTAC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121    TGTACAAAGTCCGGGGGGGTGCGTGGCCCAGGACACCTCGGTGGTCCTTGACTGGAACTAC

181    CGCTGGATGCACGACGCAAACTACAACTCGTGCACCGTCAACGGCGGCGTCAACACCACG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181    CGCTGGATGCACGACGCAAACTACAACTCGTGCACCGTCAACGGCGGCGTCAACACCACG

241    CTCTGCCCTGACGAGGCGACCTGTGGCAAGAACTGCTTCATCGAGGGCGTCGACTACGCC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241    CTCTGCCCTGACGAGGCGACCTGTGGCAAGAACTGCTTCATCGAGGGCGTCGACTACGCC

301    GCCTCGGGCGTCACGACCTCGGGCAGCAGCCTCACCATGAACCAGTACATGCCCAGCAGC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301    GCCTCGGGCGTCACGACCTCGGGCAGCAGCCTCACCATGAACCAGTACATGCCCAGCAGC

361    TCTGGCGGCTACAGCAGCGTCTCTCCTCGGCTGTATCTCCTGGACTCTGACGGTGAGTAC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361    TCTGGCGGCTACAGCAGCGTCTCTCCTCGGCTGTATCTCCTGGACTCTGACGGTGAGTAC

421    GTGATGCTGAAGCTCAACGGCCAGGAGCTGAGCTTCGACGTCGACCTCTCTGCTCTGCCG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421    GTGATGCTGAAGCTCAACGGCCAGGAGCTGAGCTTCGACGTCGACCTCTCTGCTCTGCCG

481    TGTGGAGAGAACGGCTCGCTCTACCTGTCTCAGATGGACGAGAACGGGGGCGCCAACCAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481    TGTGGAGAGAACGGCTCGCTCTACCTGTCTCAGATGGACGAGAACGGGGGCGCCAACCAG

541    TATAACACGGCCGGTGCCAACTACGGGAGCGGCTACTGCGATGCTCAGTGCCCCGTCCAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541    TATAACACGGCCGGTGCCAACTACGGGAGCGGCTACTGCGATGCTCAGTGCCCCGTCCAG

601    ACATGGAGGAACGGCACCCTCAACACTAGCCACCAGGGCTTCTGCTGCAACGAGATGGAT
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601    ACATGGAGGAACGGCACCCTCAACACTAGCCACCAGGGCTTCTGCTGCAACGAGATGGAT

661    ATCCTGGAGGGCAACTCGAGGGCGAATGCCTTGACCCCTCACTCTTGCACGGCCACGGCC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
661    ATCCTGGAGGGCAACTCGAGGGCGAATGCCTTGACCCCTCACTCTTGCACGGCCACGGCC
```

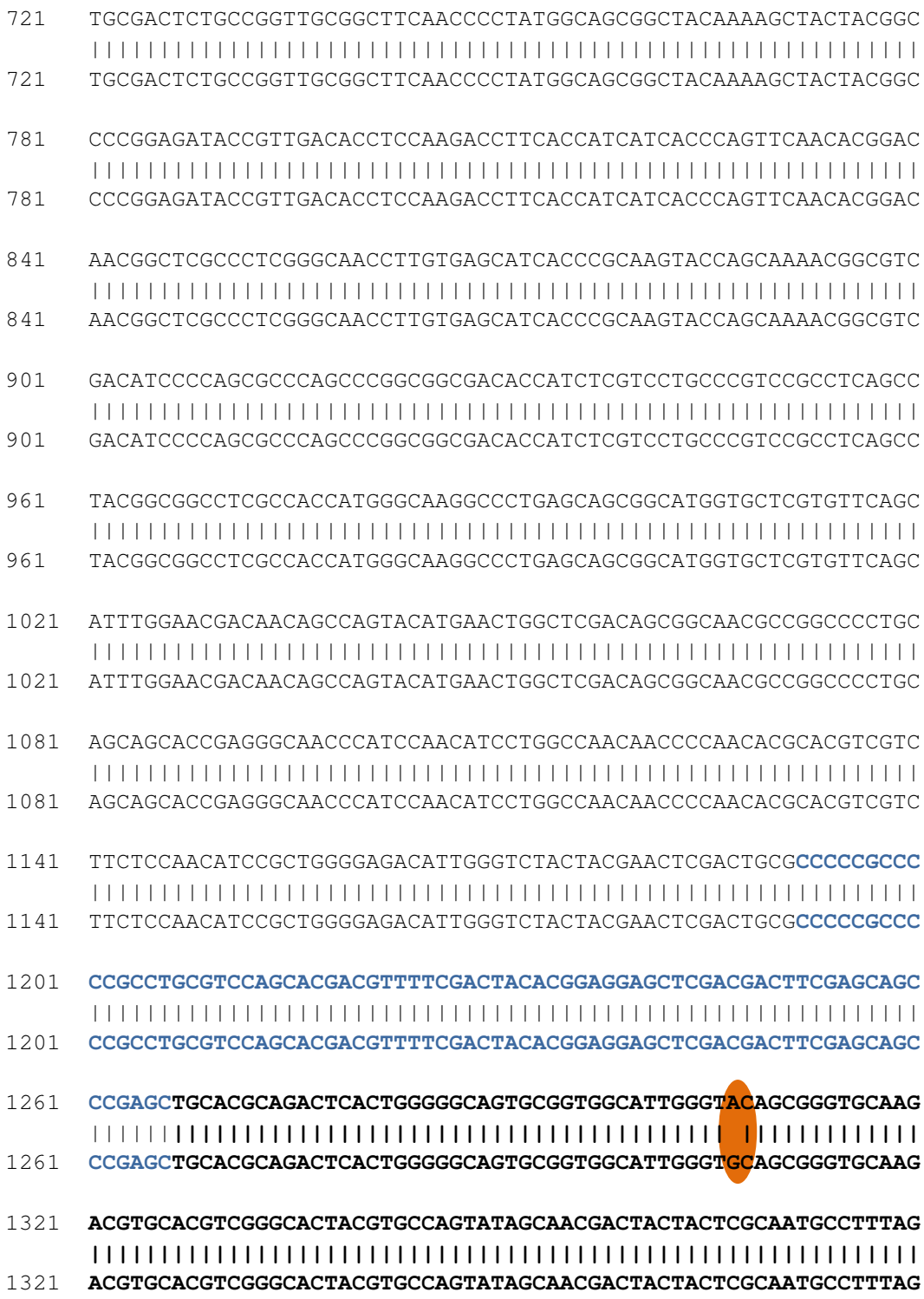

Figure 10: The sequence alignment of the *egI* genes of pAZ40 (upper line) and pLEM1 (lower line) revealed a 99.93% homology. One single base mutation has been detected in the carbohydrate binding module (bold type) and is indicated with a brown circle. The linker region (blue) separates the CBM from the catalytic domain. The catalytic domain is located between the signal peptide (green) and the linker region (www.cazy.org).

3.6 Co-expression of cellulases and growth on PASC

Growth on amorphous cellulose as sole carbon source by a *S. cerevisiae* strain expressing both an *egI* and *bgl1* episomally has been demonstrated recently (Den Haan *et al.* 2007b). The *S. cerevisiae* recombinant strains constructed in this study was tested for growth on both SC and YP medium, supplemented with 1% PASC as sole carbon source. However, no growth was detected on the amorphous cellulose substrate. To determine whether this could be due to metabolic load caused by gene co-expression, endoglucanase assays were performed with the cellulase co-expressing strains cultivated in YPD. Endoglucanase (EGI and EGII) activities were generally lower in the strains expressing additional cellulase genes (Figure 11). Co-expression with the *sCBHI*, had a greater influence on endoglucanase activity levels than co-expression with the multi-copy integrated *BGL1*.

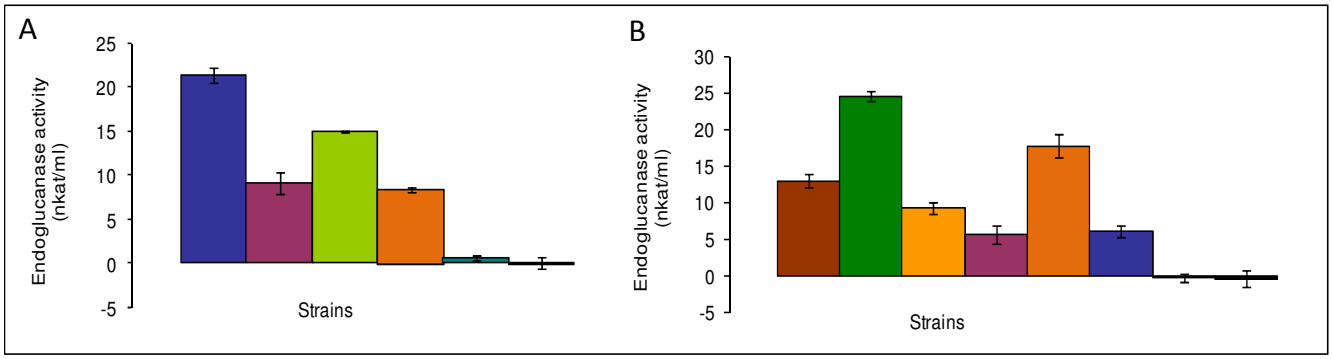

Figure 11: (A) EGII activity produced by recombinant *S. cerevisiae* strains: ■ Y294[*fur1::LEU2* pLEGII], ■ Y294[*fur1::LEU2* pLEGIIC], ■ Y294[BGL1 *fur1::LEU2* pLEGII], ■ Y294[BGL1 *fur1::LEU2* pLEGIIC], ■ Y294[*fur1::LEU2* YEpENOB-sCBHI], ■ Y294[BGL1]. (B) EGI activity by recombinant *S. cerevisiae* strains: ■ Y294[*fur1::LEU2* pAZ40], ■ Y294[*fur1::LEU2* pLEM1], ■ Y294[*fur1::LEU2* pLEM1C], ■ Y294[BGL1 *fur1::LEU2* pAZ40], ■ Y294[BGL1 *fur1::LEU2* pLEM1], ■ Y294[BGL1 *fur1::LEU2* pLEM1C], ■ Y294[*fur1::LEU2* YEpENOB-sCBHI], ■ Y294[BGL1].

3.7 Product formation

The products formed after hydrolysis of PASC were analyzed with HPLC. For all strains tested the main end-products formed on PASC were glucose, cellobiose and cellotriose (Table 4). The main product for all strains, with the exception of *S. cerevisiae* Y294[*fur1::LEU2* pAZ40] and *S. cerevisiae* Y294[BGL1 *fur1::LEU2* pLEGII], was cellobiose.

Although significant amounts of cellotriose were detected for strains producing EGII, a different pattern was observed for those strains producing EGI. For these strains little, if any, cellotriose was detected, which is similar to the results obtained in a previous study (Karlsson *et al.* 2002).

Table 4: The ratio of products formed relative to glucose in strains co-expressing *sCBHI* and *BGL1* with *EGI/EGII*.

Yeast Strains	**Glucose : Cellobiose : Cellotriose**
S. cerevisiae Y294[*fur1::LEU2* pLEGII]	1.0 : 7.4 : 2.9
S. cerevisiae Y294[*fur1::LEU2* pLEGIIC]	1.0 : 11.4 : 0.2
S. cerevisiae Y294[BGL1 *fur1*::*LEU2* pLEGII]	1.0 : 1.5 : 1.9
S. cerevisiae Y294[BGL1 *fur1*::*LEU2* pLEGIIC]	1.0 : 5.3 : 1.0
S. cerevisiae Y294[*fur1::LEU2* pAZ40]	1.0 : 0.1 : 0.0
S. cerevisiae Y294[*fur1::LEU2* pLEM1]	1.0 : 4.0 : 0.0
S. cerevisiae Y294[*fur1::LEU2* pLEM1C]	1.0 : 10.9 : 0.1
S. cerevisiae Y294[BGL1 *fur1*::*LEU2* pLEM1]	1.0 : 3.5 : 0.0
S. cerevisiae Y294[BGL1 *fur1*::*LEU2* pLEM1C]	1.0 : 6.3 : 0.2

4. DISCUSSION

The yeast, *S. cerevisiae*, has a long association with the wine and brewing industries for its ability to produce ethanol. It therefore comes as no surprise that it is currently still the organism of choice for the production of bioethanol through CBP. The feasibility of CBP would ultimately depend on the successful conversion of inexpensive lignocelluloses to glucose by means of a consortium of enzymes expressed at the optimal ratio. In this study we expressed the *T. reesei* endoglucanase genes, *egII* (*Cel5A*) and *egI* (*Cel7B*), in the yeast strain *S. cerevisiae* Y294. The EGI activity was improved through random mutagenesis using the *E. coli* ES1301 *mutS* mutator strain. We also constructed recombinant yeast strains co-displaying these *T. reesei* endoglucanases with a synthetic cellobiohydrolase (sCBHI), as well as a β-glucosidase (BGL1) from *S. fibuligera*. Growth of *S. cerevisiae* strains due to this synergistic expression was tested on an amorphous cellulose substrate, PASC. Hydrolysis products produced by the recombinant strains were also determined.

The results of this study indicate that *S. cerevisiae* Y294 was able to produce functional *T. reesei* EGII. Although no activity was detected towards CMC when cells were cultured in selective synthetic medium, activity was detected towards barley β-glucan (Figure 3). This substrate preference could be attributed to the difference in structure of the two substrates. The cellulose derivative, carboxymethylcellulose, has a number of carboxymethyl groups (-CH_2COONa) attached to the cellulose molecule, formed by treatment with sodium hydroxide and chloroacetic acid (Biswal and Singh 2004). CMC is thus a synthetic cellulose derivative, whereas barley β-glucan is a natural cellulose-like polysaccharide. Its structure differs from CMC in that it has, in addition to β-1,4-linkages (70%), also β-1,3-linkages (30%) between glucose residues (Christensen *et. al.* 2001).

Even though no activity could be detected towards CMC in minimal medium, the activity levels were similar to that obtained towards barley β-glucan when cells were cultivated in rich medium. This observation could be compared to, and partly explained by results obtained in a study by Stals *et al.* (2004). This group found that the glycosylation of the *T. reesei* cellobiohydrolase I (Cel7A) varies when the fungus is cultivated under different conditions. The protein was fully N-glycosylated (at all three glycosylation sites of the catalytic domain) when isolated from minimal medium, but when the fungus was cultivated in rich medium, only one or two sites contained high-mannose chains. A difference in the *O*-glycosylation pattern in the linker region was also observed. Grown in minimal medium, there were 19-29 mannose substitutions detected on the linker peptide, whereas it contained only about 16-23 mannose residues in rich medium. *S. cerevisiae* elongates the mannose chain in heterologous proteins by adding even more mannose residues (Romanos *et al.* 1992). Based on this, we speculate that, when *S. cerevisiae* Y294 [pLEGII] was grown in minimal medium, the secreted protein was hyperglycosylated on all of its glycosylation sites, rendering it unable to access the more bulky structure of CMC for efficient hydrolysis. It is possible that hyperglycosylation takes place to a lesser extend when cultivation takes place in rich medium, enabling more efficient degradation of CMC.

To obtain higher cell biomass and consequently higher levels of produced heterologous enzyme, we constructed *S. cerevisiae* autoselective strains by disrupting the *FUR1* gene of *S. cerevisiae* (La Grange *et al.* 1996). Instead of following the pyrimidine salvage pathway, the yeast is forced to utilize the episomally produced *URA3* gene product to synthesize uridine 5'-phosphate *de novo*, even

when cultivated in complex medium such as YPD. Disruption of the *FUR1* gene has an additional advantage; the transformants can be cultivated in less expensive media because they do not require selective pressure to maintain the plasmids. The highest endoglucanase activity produced by *S. cerevisiae* Y294[pLEGII] was obtained in YPD medium, resulting in 19.6 nkat.ml^{-1} and 22.3 nkat.ml^{-1} on CMC and barley β-glucan, respectively (Figure 3).

The activity of the *T. reesei* EGI was improved by using random mutagenesis utilizing the *E. coli* mutator strain. After transformation of the mutated *egI* containing plasmid (pLEM1) to *S. cerevisiae* Y294, transformants were screened for improved levels of endoglucanase activity. One transformant (*S. cerevisiae* Y294[pLEM1]) was identified which produced endoglucanase levels of 21.8 nkat.ml^{-1}, double that of the strain producing the wild-type gene (*S. cerevisiae* Y294[pAZ40]) (Figure 5). Autoselective strains cultured in rich media produced a similar activity pattern (Figure 6). However, the specific activities were low, suggesting that the higher levels of activity obtained in YPD was due to the higher biomass production.

Expression of the mutated *egI* gene under the control of an unaltered promoter (*S. cerevisiae* Y294[pMEG1]) also resulted in higher levels of endoglucanase activity, although it was lower than that produced by *S. cerevisiae* Y294[pLEM1]. Sequencing of the mutated endoglucanase gene revealed a single A to G transition located in the carbohydrate binding module (CBM) (Figure 10). The deduced amino acid sequence shows a change from a tyrosine to a cysteine due to this point mutation. The function of the CBM is to bring the cellulase catalytic domain in close proximity of the cellulose surface and maintaining it there for the course of hydrolysis (Linder *et al.* 1999).

The CBM in fungal enzymes are generally classified as family 1 and usually consist of less than 40 amino acids. Binding to cellulose involves the hydrophobic interaction of the sugar rings of cellulose and aromatic side chains of tyrosines, phenylalanines or tryptophans on a planar strip of the CBM. The CBM of *T. reesei* CBHI has been the focus of most studies, and the three dimensional structure has been solved (Linder *et al.* 1995). Since the amino acid sequences of the EGI and CBHI CBMs are similar, the latter could provide a model for studying the structure and function of the EGI CBM (Figure 12). Mattinen *et al.* 1997 have shown that replacing a tyrosine at position 5 with an alanine resulted in a distorted CBM structure, although a mutation at position 31 barely affected the structure.

Thus it is proposed that the substitution of a tyrosine at position 13 in this study caused a change in structure of the EGI CBM, resulting in an altered binding affinity of the enzyme. Three disulphide bridges in the EGI CBM stabilize the structure (Linder *et al.* 1999). The cysteine in the EGI CBM at position 16 forms a disulfide bridge with another cysteine near the N-terminus (Linder *et al.* 1995). Therefore, it is possible that the newly introduced cysteine at position 13 in the mutated CBM is rather involved in this interaction, again potentially altering the conformation, the stability or binding affinity of the CBM.

The EGI produced by *S. cerevisiae* Y294[pAZ40], *S. cerevisiae* Y294[pLEM1] and *S. cerevisiae* Y294[pMEG1] all have a pH optimum between 4.5 and 5 (Figure 8). The optimum temperature for EGI was at 60°C for all strains tested. Therefore it can be concluded that the mutations had no effect on these characteristics of the protein. This could be expected, since an amino acid substitution (tyrosine→cysteine) was identified in the CBM, and not in the catalytic domain.

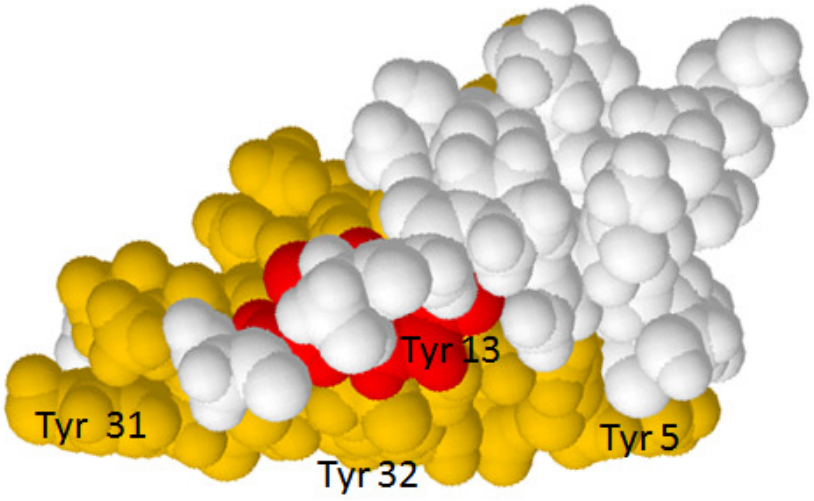

Figure 12: The structure of CBHI CBM. The flat cellulose binding surface is formed by three antiparallel β-sheets (yellow). The position of the tyrosine which was substituted with a cysteine in EGI is indicated in red. Note that the EGI CBM does not contain the tyrosine at position 5, but a tryptophan instead (Linder *et al.* 1995). The structure was drawn using the program Rasmol (version 2.6).

The mutation in the *egI* was responsible for improved activity, but this was not the only contributing factor for overall higher levels of endoglucanase activity produced by *S. cerevisiae* Y294[pLEM1]. It has previously been shown that the constitutive (glycolytic) *ENO1* promoter, successfully promotes heterologous protein expression (Van Rooyen *et al.* 2005, Shiba *et al.* 1998). To determine whether the mutations in the promoter region (Figure 9) had an effect on the increased levels of the EGI, the *T. reesei* xylanase gene was used as reporter gene. The strain *S. cerevisiae* Y294[pMPxyn] produced higher levels of endo-xylanase activity than *S. cerevisiae* Y294[pAZ1], suggesting improved promoter function due to the mutation (Figure 7). Three point mutations had been identified by automated sequencing: one A to G transition and two T to C transitions. Although the polymerase II binding consensus (TATA element) is unaltered, these mutations could have an effect on the binding and function of assisting transcription factors.

S. cerevisiae has nine basic helix-loop-helix (bHLH) proteins, which are classified as transcriptional regulators (Chen and Lopes 2007). The bHLH proteins interact with upstream activation sequence (UAS) elements including the E box motif. One E box is present in the *ENO1*-promoter used in this study. It has been shown that the bHLH protein Sgc1p functions exclusively through this specific E box element. Sgc1p activates the expression of glycolytic genes such as enolase 1 (*ENO1*). The Gcr1p transcription factor ensures high-level expression of these genes and interacts with the CTTCC sequence motif, located in the UAS region (Huie *et al.* 1992). Results from mutational studies performed by Chen and Lopes (2007) demonstrated that all nine *S. cerevisiae* bHLH proteins were required for maximal expression of the *ENO1* gene. Further studies regarding their binding sites and specific roles in transcriptional regulation could possibly resolve the potential effects of the mutations induced in the promoter.

The copy number of an episomal plasmid is determined by the 2µ region on the plasmid, the size of the plasmid and the stringency of the selection used. Since an alteration in plasmid copy number could be a contributing factor to higher levels of expressed enzyme, real-time PCR was performed to compare the relative copy number of *S. cerevisiae* Y294[pLEM1] to *S. cerevisiae* Y294[pAZ40] (Table 3). Results showed similar plasmid copy numbers for both strains, estimated at about 20 copies per cell. The 2µ-based plasmids (YEp) generally have about 25 copies per cell (Romanos *et al.* 1992). The lower

copy number of strains used in this study is attributed to the larger size of the plasmids, which is a result of the presence of the gene cassettes.

Co-expression of endoglucanases (*egI* and *egII)* with the synthetic *T. reesei* cellobiohydrolase, the *S. fibuligera* β-glucosidase, or both, led to lower levels of endoglucanase activity compared to separate expression (Figure 11). This 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). This was due to a reduction in cell yields, suggesting an increased metabolic burden enforced on the cells.

Analysis of the hydrolysis products derived from an amorphous cellulose substrate, PASC, was found to be similar to that observed in a previous study (Karlsson *et al.* 2002). EGII produced glucose, cellobiose and cellotriose as end products, while mainly glucose and cellobiose were detected for strains expressing EGI. An overall high level of cellobiose was observed for all strains. This bottleneck was even larger when co-expressed with *sCBHI*. Den Haan *et al.* (2007a) observed that the heterologous production of the *T. reesei* CBHI by *S. cerevisiae* resulted in low titers of functionally secreted enzyme, even though the specific activity was found not to differ significantly to that of the native enzyme. Therefore the *sCBHI* used in this study has been synthetically designed to be codon optimized (CAI = 0.666) for efficient translation in *S. cerevisiae*. However, it has been reported that cellobiose has a strong inhibitory affect on *T. reesei* cellulases (Gruno *et al.* 2003), which makes the effective conversion of this product to glucose essential. Co-expression with *BGL1* slightly alleviated the metabolic load (Table 4), although it was not effective enough to produce sufficient glucose for growth on amorphous cellulose.

It has recently been shown that the episomal multicopy expression of the *egI* (of *T. reesei*) together with *bgl1* (of *S. fibuligera*) by *S. cerevisiae* enabled the recombinant yeast to grow on phosphoric acid swollen cellulose, PASC (Den Haan *et al.* 2007b). The strains constructed in this study were unable to grow on PASC. This could be explained by the *bgl1* not being expressed at such high levels when integrated into the chromosome as was the case with strain *Saccharomyces cerevisiae* Y294[BGL1]. In a study by McBride *et al.* (2005) yeast strains *S. cerevisiae* N96[ySFI] and *S. cerevisiae* Y294[ySFI], both expressing the *bgl1* gene of *S. fibuligera*, were tested for growth on cellobiose. The strain

containing the multicopy episomal *bgl1* exhibited a faster growth rate than the strain containing the integrated *bgl1*. The β-glucosidase activity was about five-fold higher with multicopy expression as compared to integrative expression. It was experimentally confirmed that the amount of β-glucosidase activity was the main limitation affecting the growth of the strain containing the integrated *bgl1*. In the study by Den Haan *et al.* (2007b) it was found that most of the β-glucosidase activity was cell associated, even though a secretion signal (*T. reesei xyn2* secretion signal) preceded the *bgl1* gene. Therefore, it is also possible that not all the BGL1 enzymes took part in the hydrolysis; some enzymes could have remained in the cell filtrate since only the supernatant was used to obtain the information for Table 4.

The use of recombinant *S. cerevisiae* strains co-expressing cellulases was assessed for their ability to degrade amorphous cellulose. Cellulose degradation is the first milestone to achieve towards conversion of cellulose to ethanol through simultaneous saccharification and fermentation. Although no growth was observed on amorphous cellulose, the study has led to a greater understanding of co-expression of enzymes by the yeast *S. cerevisiae*. This knowledge can be applied in future studies for optimization of the ratio at which the individual cellulases (in cellulase cocktails) should be recombinantly expressed by *S. cerevisiae*.

5. ACKNOWLEDGEMENTS

The authors would like to thank the National Research Foundation (NRF) for financial support. We would also like to express gratitude to Nico de Villiers from *Omic Solutions* for assistance in the real-time PCR experiments and Anscha Zietsman and Dr Riaan den Haan for supplying plasmids pAZ1 and YEpENOB-sCBHI, as well as strain *Saccharomyces cerevisiae* Y294[BGL1].

6. REFERENCES

Agbogbo F, Coward-Kelly G (2008) Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast, *Pichia stipitis*. Biotechnology Letters DOI: 10.1007/s10529-008-9728-z

Bailey MJ, Biely P, Poutanen K (1992) Interlaboratory testing of methods for assay of xylanase activity. Journal of Biotechnology 23: 257-270

Bailey MJ, Siika-aho M, Valkeajärvi A, Penttilä ME (1993) Hydrolytic properties of two cellulases of *Trichoderma reesei* expressed in yeast. Biotechnology and Applied Biochemistry 17: 65-76

Biswal DR, Singh RP (2004) Characterisation of carboxymethyl cellulose and polyacrylamide graft copolymer. Carbohydrate Polymers 57: 379-387

Chen M, Lopes JM (2007) Multiple basic helix-loop-helix proteins regulate expression of the *ENO1* gene of *Saccharomyces cerevisiae*. Eukaryotic Cell 6:786-796

Christensen BE, Ulset A-S, Beer MU, Knuckles BE, Williams DL, Fishman ML, Chau HK, Wood PJ (2001) Macromolecular characterisation of three barley β-glucan standards by size-exclusion chromatography combined with light scattering and viscometry: an inter-laboratory study. Carbohydrate Polymers 45: 11-22

Cohen R, Yokoi T, Holland JP, Pepper AE, Holland MJ (1987) Transcription of the constitutively expressed yeast enolase gene *ENO1* is mediated by positive and negative *cis*-acting regulatory sequences. Molecular and Cellular Biology 7: 2753-2761

Den Haan R, McBride J, La Grange DC, Lynd LR, Van Zyl WH (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 SH, Lynd LR, Van Zyl WH (2007b) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metabolic Engineering 9: 87-94

Fujita Y, Ito J, Ueda M, Fukuda H, 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. Applied and Environmental Microbiology 70: 1207-1212

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

Gruno M, Väljamäe P, Pettersson G, Johansson G (2003) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. Biotechnology and Bioengineering 86: 503-511

Hill J, Ian KA, Donald G, Griffiths DE (1991) DMSO-enhanced whole cell yeast transformation. Nucleic Acids Research 19: 5791

Hoffman CS, Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57: 167-272

Huie MA, Scott EW, Drazinic CM, Lopez MC, Hornstra IK, Yang TP, Baker HV (1992) Characterization of the DNA-binding activity of GCR1: In Vivo evidence for two GCR1-binding sites in the upstream activating sequence of *TPI* of *Saccharomyces cerevisiae*. Molecular and Cellular Biology 12: 2690-2700

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

Kern L, de Montigny J, Jund R, Lacroute F (1990) The *FUR1* gene of *Saccharomyces cerevisiae*: cloning, structure and expression of wild-type and mutant alleles. Gene 88: 149-157

Kleman-Leyer KM, Siika-Aho M, Teeri TT, Kirk TK (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. Applied and Environmental Microbiology 62: 2883-2887

Kristensen JB, Thygesen LG, Felby C, Jørgensen H, Elder T (2008) Cell-wall structural changes in wheat straw pretreated for bioethanol production. Biotechnology for Biofuels DOI: 10.1186/1754-6834-1-5

La Grange DC ,Claeyssens IM, Pretorius IS, Van Zyl WH (2001) Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* xylosidase (*xlnD*) and the *Trichoderma reesei* xylanase II (*xyn2*) genes. Applied and Environmental Microbiology 67: 5512-5519

La Grange DC, Pretorius IS, Van Zyl WH (1996) Expression of a *Trichoderma reesei* beta-xylanase gene (*XYN2*) in *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 62: 1036-1044

Linder M, Lindeberg G, Reinikainen T, Teeri TT, Pettersson G (1995) The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. FEBS Letters 372: 96-98

Linder M, Nevanen T, Teeri TT (1999) Design of a pH-dependent cellulose-binding domain. FEBS Letters 447: 13-16

Lynd LR, Van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology 16: 577-583

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews 66: 506-577

Machida M, Jigami Y, Tanaka H (1989) Purification and characterization of a nuclear factor which binds specifically to the upstream activation sequence of *Saccharomyces cerevisiae* enolase 1 gene. European Journal of Biochemistry 184: 305-311

Mattinen M-L, Kontteli M, Kerovuo J, Linder M, Annila A, Lindeberg G, Reinikainen T, Drakenberg T (1997) Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*. Protein Science 6: 294-303

McBride JE, Zietsman JJ, Van Zyl WH, Lynd LR (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: 93-101

Miller GL (1959) Use of dinitrosalicyclic acid reagent for determination of reducing sugar. Analytical chemistry 31: 426-428

Medve J, Karlsson J, Lee D, Tjerneld F (1998) Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: adsorption, sugar production pattern, and synergism of the enzymes. Biotechnology and Bioengineering 59: 621-634

Okada H, Tada K, Sekiya T, Yokoyama K, Takahashi A, Tohda H, Kumagai H, Morikawa Y (1998) Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. Applied and Environmental Microbiology 64: 555-563

Romanos MA, Scorer CA, Clare JJ (1992) Foreign gene expression in yeast: a review. Yeast 8: 423-488

Sambrook J, Fritsch E F, Maniatis T. 1989. Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, NY

Sandgren M, Ståhlberg J, Mitchinson C (2005) Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. Progress in Biophysics and Molecular Biology 89: 246-291

Shiba Y, Fukui F, Ichikawa K, Serizwa N, Yoshikawa H (1998) Process development for high-level secretory production of carboxypeptidase Y by *Saccharomyces cerevisiae*. Applied Microbiology and Biotechnology 50: 34-41

Stals I, Sandra K, Geysens S, Contreras R, Van Beeumen J, Claeyssens M (2004) Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. Glycobiology 14: 713-724

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

Van Rooyen R, Hahn-Hägerdal B, La Grange DC, Van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. Journal of Biotechnology 120: 284-295

Woodward J, Lima M, Lee NE (1988) The role of cellulase concentration in determining the degree of synergism in the hydrolysis of microcrystalline cellulose. Biochemical Journal 255: 895-899

Websites cited:

Carbohydrate-Active enZYmes. Last updated: 1 August 2008. AFMB – CNRS –Universités Aix-Marseille I & II. Accessed: 19 August 2008 <www.cazy.org>

Home - of the Evolving Code Wiki. Evolvingcode. Last updated: 29 May 2007. Freeland lab, Biological Sciences Department at UMBC, University of Maryland. Accessed: 26 August 2008. <www.evolvingcode.net>

CHAPTER 4
GENERAL DISCUSSION AND CONCLUSIONS

1. DISCUSSION

The depleting fossil fuel reserves and the adverse effects of greenhouse gas emissions on the environment, has created the need for the development of alternative energy sources. Liquid biofuels produced from biomass could provide a more sustainable source of energy for transportation. Lignocellulosic biomass is an attractive source of fermentable sugars for conversion to bioethanol, since it is renewable, inexpensive, geographically distributed and its combustion results in low greenhouse gas emissions. Currently, research is focusing on the development of a recombinant microorganism with the ability to degrade lignocellulose and ferment the resulting sugars in one step, referred to as CBP (Consolidated Bioprocessing). CBP is proposed to be the most cost-effective means of bioethanol production from cellulosic biomass.

S. cerevisiae has for many years been associated with the beer, wine and baking industries, and has established itself as a popular host for recombinant expression studies (Ostergaard *et al.* 2000). Apart from its recognized safety (GRAS status), other advantages include the ease of transformation with foreign DNA and its ability of perform posttranslational modifications of heterologous proteins. Many cellulase genes have successfully been expressed, and in recent years co-expressed, in *S. cerevisiae*, for the production of ethanol from lignocellulosic biomass. The most commonly used cellulases are sourced from the filamentous fungus, *T. reesei*, which possesses one of the best studied cellulase enzyme systems, producing all the enzymes needed for efficient breakdown of crystalline cellulose to glucose. All five endoglucanases and two cellobiohydrolases have been cloned and successfully expressed in *S. cerevisiae* (Lynd *et al.* 2002).

To increase the efficiency of cellulose hydrolysis, cellulase performance (production or activity levels) can be improved by directed evolution applied to each cellulase enzyme. The main advantage of directed evolution is that it does not require any information on enzyme structure or its interaction with a substrate. Table 1 contains a list of cellulases with altered properties obtained using directed evolution. Directed evolution was applied in this study by using an *E. coli* strain with a defective DNA mismatch repair system. The activity of the major endoglucanase of *T. reesei* (EGI) was improved two-fold, with the point mutation identified in the carbohydrate binding module (CBM). The functioning of the constitutive yeast enolase 1 promoter, *ENO1*, was also altered.

The cellulase enzymes need to work synergistically to obtain effective cellulose hydrolysis. Synergistic cellulosic saccharification and fermentation to ethanol has been demonstrated by yeast strains co-expressing different cellulase components (Den Haan *et al.* 2007, Fujita *et al*. 2004, 2002). We therefore decided to investigate and evaluate the co-expression of the three main components of the cellulase system, that could assist in the future construction of a yeast strain producing cellulases (for growth on cellulose and consequent fermentation to ethanol) at the required ratio. The *T. reesei* endoglucanases (*egI* and *egII*) were co-expressed with a synthetic cellobiohydrolase (*cbhI* from *T. reesei*) and a β-glucosidase (*bgl1* from *S. fibuligera*). The recombinant *S. cerevisiae* strains were able to degrade PASC to glucose, cellobiose and cellotriose, but the glucose yield was insufficient to enable growth on cellulose as sole carbon source.

Table 1: Cellulases with altered properties, obtained through various techniques (Zhang *et al.* 2006).

Enzyme	**Altered property**	**Technique**	**Reference**
Endoglucanase	Thermal stability	Family shuffling	Murashima *et al.* 2002
Endoglucanase	Activity	DNA shuffling	Kim *et al.* 2000
Endoglucanase	Alkali pH	Error-prone-PCR	Wang *et al.* 2005
Endoglucanase	-	Family shuffling	Catcheside *et al.* 2003
β-glucosidase	Cold adoption	DNA shuffling	Lebbink *et al.* 2000
β-glucosidase	Thermal stability	Error-prone-PCR	Gonzalez-Blasco *et al.* 2000
β-glucosidase	Thermal stability	Error-prone-PCR and Family shuffling	Arrizubieta and Polaina 2000
β-glucosidase	Activity	Error-prone-PCR	McCarthy *et al.* 2004
β-glycosidase	Activity	Family shuffling	Kaper *et al.* 2002
β-glucosidase (glycosynthase)	Activity	Error-prone-PCR	Kim *et al.* 2004
Endoglucanase (glycosynthase)	Activity	Cassette mutagenesis	Lin *et al.* 2004

2. CONCLUSIONS

To summarize, the following can be concluded from this study:

- The *T. reesei* endoglucanases (*egI* and *egII*) were functionally expressed under transcriptional control of the *ENO1*-promotor-terminator cassette by *S. cerevisiae* Y294.
- Secreted EGI activity was improved approximately two-fold by random mutagenesis, using the *Escherichia coli* ES1301 *mutS* mutator strain.
- The optimum pH and temperature profiles of the mutated EGI were shown to be unaltered.
- The mutation contributing to improved levels of EGI activity was located in the carbohydrate binding module of the enzyme and was identified as an amino acid change from a tyrosine to a cysteine.
- Data suggest improved function of the constitutive *ENO1* promoter due to three point mutations identified in the sequence of the promoter.
- Co-expression of *egI* and *egII* with the synthetic *sCBHI* and the *bgl1* led to lower levels of endoglucanase activity, indicating an additional metabolic load or even saturation of the secretion capacity of the cell.
- The strains were used to hydrolyze PASC and the end products were identified with HPLC. Strains expressing *egII* mainly produced glucose, cellobiose and cellotriose, while only glucose and cellobiose were detected for strains expressing *egI*. Overall high levels of cellobiose were detected, suggesting ineffective BGL1 activity.
- Recombinant *S. cerevisiae strains* co-expressing different combinations of cellulases were unable to grow on amorphous cellulose as sole carbon source, probably due to the limited BGL1 activity.

3. FUTURE RESEARCH

Cellulases most often used in recombinant expression for degradation of cellulosic substrates are obtained from *T. reesei*. Although this organism contains the complete repertoire of cellulases for complete cellulose hydrolysis, not all of them are heterologously produced at high levels. Therefore, research is not restricted to the use of only *T. reesei* as source of cellulases. It has been shown, for

example, that the β-glucosidase from *T. reesei* is expressed at inefficient levels by *S. cerevisiae* as host, and therefore studies have been conducted to evaluate other sources of this component of the cellulase system. Comparison of four β-glucosidases expressed by recombinant yeasts suggested the β-glucosidase from *Saccharomycopsis fibuligera* to be the best candidate for heterologous expression directed at cellulose degradation (Van Rooyen *et al.* 2005). Other genera that have received attention with regard to their cellulolytic capabilities, include *Aspergillus, Fusarium, Geotricum, Penicillium, Clostridium, Cellulomonas, Thermobifida*, etc. (Lynd *et al.* 2002).

A disadvantage of *S. cerevisiae* as recombinant host is its tendency to hyperglycosylate secreted heterologous proteins, resulting in low levels of active extracellular protein. Therefore, alternative yeasts are also being investigated for use as host organisms for expression of cellulase genes (Ostergaard *et al.* 2000). Other non-*Saccharomyces* yeasts under investigation for the production of heterologous proteins include *Kluveromyces lactis*, *Yarrowia lipolytica*, *Hansenula polymorpha* and *Pichia pastoris*, the latter being the most studied alternative yeast system (Domínguez *et al.* 1998).

Improvement in cellulase performance by directed evolution is a powerful strategy for more efficient degradation of cellulose. A relatively simple approach to improvement is the random mutagenesis of an individual cellulase, before selecting or screening for new or improved properties. The success of directed evolution depends on the property sought, or simply put: "You get what you screen for". The largest limitation to directed evolution (as in this study), is that the screening of mutants is based on their ability to degrade soluble substrates, whereas cellulose is a solid, insoluble composite of lignocelluloses biomass. However, it is a starting point, from which a greater understanding of substrate interaction and hydrolysis could lead to better design of strategies for crystalline cellulose degradation. Future studies can focus on developing experimentally simpler, faster and more effective techniques for the screening of larger libraries in the search for improved properties (Hibbert *et al.* 2005). A step in the right direction is the rapid improvement in computational approaches. Thereby, libraries can be created that contain fewer redundant enzyme variants, by elimination of sequences that are incompatible with the particular protein fold. Also, computer modeling can be used for creating novel or altered enzymes, an example being the redesign of an enzyme active site for improved catalytic activity (Hibbert *et al.* 2005).

With regards to this study in particular, several approaches need to be adjusted in the quest to developing an ideal CBP microorganism. Although the endoglucanases were successfully expressed by *S. cerevisiae* transformants, the co-expression thereof with other cellulases did not, as theoretically expected, enable the yeast to grow on cellulose as sole carbon source. Also, results suggested a metabolic burden imposed on the cells, due to expression of three heterologous proteins. Therefore, future studies could focus on optimizing cellulase cocktail mixtures produced by the host organisms, to achieve the most effective synergistic activity. In terms of cellulase improvement, this study has demonstrated the benefit of random mutagenesis for cellulase improvement. The limited activity of the β-glucosidase resulted in a bottleneck in the cellulose degradation, indicating that there is room for improvement for the BGL1, which might also be obtained by means of random mutagenesis. Since screening was based only on higher levels of activity, future studies on mutated enzymes can focus on altering other enzyme properties, such as pH and temperature optima, stability, substrate specificity and affinity, etc.

Methods currently used for saccharification and fermentation of lignocellulosic biomass are still inefficient and expensive (Rubin 2008). Although major milestones have been achieved in the development of a microorganism with the ability to utilize lignocellulosic substrates and produce ethanol, a great deal of research still needs to be conducted in an attempt to construct the ideal recombinant organism for CBP. Issues to be addressed include the extension of substrate and product range, improvement of process performance, improvements in product yield and elimination of toxic by-products (Ostergaard *et al.* 2000).

4. REFERENCES

Arrizubieta MJ, Polaina J (2000) Increased thermal resistance and modification of the catalytic properties of a beta-glucosidase by random mutagenesis and in vitro recombination. Journal of Biological Chemistry 275: 28843-28848

Catcheside DE, Rasmussen JP, Yeadon PJ, Bowring FJ, Cambareri EB (2003) Diversification of exogenous genes in vivo in Neurospora. Applied Microbiology and Biotechnology 62: 544-549

Den Haan R, Rose SH, Lynd LR, Van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae.* Metabolic Engineering 9: 87-94

Domínguez Á, Fermiñán E, Sánchez M, González FJ, Pérez-Campo FM, García S, Herrero AB, San Vicente A, Cabello J, Prado M, Iglesias FJ, Choupina A, Burguillo FJ, Fernández-Lago L, Carmen López M (1998) Nonconventional yeasts as hosts for heterologous protein production. International Microbiology 1: 131-142

Fujita Y, Ito J, Ueda M, Fukuda H, 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. Applied and Environmental Microbiology 70: 1207-1212

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

Gonzalez-Blasco G, Sanz-Aparicio J, Gonzalez B, Hermoso JA, Polaina J (2000) Directed evolution of beta-glucosidase A from *Paenibacillus polymyxa* to thermal resistance. Journal of Biological Chemistry 275: 13708-13712

Hibbert EG, Dalby PA (2005) Directed evolution strategies for improved enzymatic performance. Microbial Cell Factories 4, DOI 10.1186/1475-2859-4-29

Kaper T, Brouns SJ, Geerling AC, De Vos WM, Van der Oost J (2002) DNA family shuffling of hyperthermostable beta-glycosidases. Biochemical Journal 368: 461-470

Kim YS, Jung HC, Pan JG (2000) Bacterial cell surface display of an enzyme library for selective screening of improved cellulose variants. Applied and Environmental Microbiology 66: 788-793

Kim YW, Lee SS, Warren RA, Withers SG (2004) Directed evolution of a glycosynthase from *Agrobacterium* sp. Increases its catalytic activity dramatically and expands its substrate repertoire. Journal of Biological Chemistry 279: 42787-42793

Lebbink JH, Kaper T, Bron P, Van der Oost J, De Vos WM (2000) Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* beta-glucosidase CelB by directed evolution. Biochemistry 39: 3656-3665

Lin H, Tao H, Cornish VW (2004) Directed evolution of a glycosynthase via chemical complementation. Journal of the American Chemical Society 126: 15051-15059

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS (2002*)* Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews 66: 506-577

McCarthy JK, Uzelac A, Davis DF, Eveleigh DE (2004) Improved catalytic efficiency and active site modification of 1,4-beta-D-glucan glucohydrolase A from *Thermotoga neapolitana* by directed evolution. Journal of Biological Chemistry 279: 11495-11502

Murashima K, Kosugi A, Doi RH (2002) Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD. Molecular Microbiology 45: 617-626

Ostergaard S, Olsson L, Nielsen J (2000) Metabolic Engineering of *Saccharomyces cerevisiae.* Microbiology and Molecular Biology Reviews 64: 34-50

Rubin EM (2008) Genomics of cellulosic biofuels. Nature 454: 841-845

Van Rooyen R, Hahn-Hägerdal B, La Grange DC, Van Zyl WH (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. Journal of Biotechnology 120: 284-295

Wang T, Liu X, Yu Q, Zhang X, Qu Y, Gao P (2005) Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. Biomolecular Engineering 22: 89-94

Zhang Y-H, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. Biotechnology Advances 24: 452-481