Regulation of genes encoding enzymes involved in plant cell wall deconstruction in *Trichoderma reesei*

Laure Nicolas Annick Ries, BSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2013

Abstract

This study describes the regulation of genes encoding plant cell wall-degrading enzymes in the presence of different carbon sources from the biotechnologically important fungus *Trichoderma reesei*. It was shown that different carbon sources influence fungal growth rate, biomass production and subsequent enzyme secretion. Several genes were identified and suggested to play a role in the development of conidia and in maintaining polarised growth.

RNA-sequencing studies showed an increase in transcript levels of genes encoding enzymes involved in plant cell wall degradation (CAZy) as well as of genes encoding lipases, expansins, hydrophobins, G-protein coupled receptors and transporters when mycelia were cultivated in the presence of a lignocellulosic substrate (wheat straw). The encoded non-CAZy proteins were proposed to have accessory roles in carbohydrate deconstruction. A model for solid substrate recognition in *T. reesei* was described, based on the comparison with the one proposed for *Aspergillus niger*. Post-transcriptional regulation mediated by regulatory RNAs was identified for nearly 2% of all *T. reesei* genes, including genes encoding cell wall-degrading enzymes.

Transcriptional regulation studies confirmed that transcription patterns of genes encoding enzymes involved in polysaccharide degradation differed between different carbon sources and that they are fine-tuned and dependent on factors such as culture conditions, consumption rate, assimilation of glucose and the presence of several transcription factors. The analysis of the structure of chromatin in the promoter and coding regions of one of these genes, *cbh1*, revealed different nucleosome positioning patterns under repressing (glucose) and inducing (sophorose, cellulose) conditions.

CRE1, the carbon catabolite repressor in *T. reesei* was shown to be involved in the repression of many CAZy and non-CAZy encoding genes. Furthermore, CRE1 was also shown to be important for nucleosome positioning within the *cbh1* coding region under repressing conditions and proposed to do so by interaction with (a) yet unidentified protein(s).

Publications

Ries, L., Belshaw, N. and Archer, D. (2010) Regulation of cellulase gene expression in *Trichoderma reesei*. Proceedings of IMC9, Edinburgh, P3.217.

Ries, L., Pullan, S., Delmas, S. and Archer, D. (2012) Comparison of the expression of genes required for carbohydrate deconstruction in *Trichoderma reesei* and *Aspergillus niger* when exposed to wheat straw. Proceedings of the BSBEC Grant holders Workshop, Crewe, p. 112.

Ries, L., Pullan, S. T., Delmas, S., Malla, S., Blythe, M. J., and Archer, D. B. (2013) Genome-wide transcriptional response of *T. reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. Manuscript in preparation.

Ries, L., Belshaw, N. J., Ilmén, M., Penttilä, M. E. and Archer, D. B. (2013) The role of CRE1 in nucleosome positioning within the *cbh1* promoter and coding regions of *Trichoderma reesei*. Manuscript in preparation.

Acknowledgements

Table of Contents

7

Abbreviations

1G/2G	First generation/second generation
ABF	Arabinofuranosidase
ACE	Activator of cellulases
ADP	Adenosine diphosphate
A. nidulans	*Aspergillus nidulans*
A. niger	*Aspergillus niger*
ARP2/3	Actin-related proteins 2/3
ATP	Adenosine triphosphate
AVC	Apical vesicle cluster
BGA	Beta-galactosidase
BGL	Beta-glycosidase
BiP	Binding protein
BLR	Blue light regulator
BSA	Bovine serum albumin
BXL	Beta-xylosidase
bZIP	Basic leucine zipper
CAE	*cbh2* activating element
CAF	Chromatin assembly factor
cAMP	Cyclic adenosine monophosphate
CAZy	Carbohydrate active enzyme
CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CCR	Carbon catabolite repression
CD	Catalytic domain

cDNA/gDNA	Complementary DNA/genomic DNA
CE	Carbohydrate esterase
CFEM	Conserved fungal specific extracellular membrane-spanning
CHD	Chromodomain helicase DNA-binding
CRE	Carbon catabolite repressor
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGL	Endoglucanase
ER	Endoplasmic reticulum
ERAD	ER-associated protein degradation
ERO1	ER-oxidoreductin 1
GH	Glycoside hydrolase
GNAT	GCN5-related N-acetyltransferase
GPCR	G-protein coupled receptor
GPRK	G-protein coupled receptor kinase
GTP	Guanosine triphosphate
GTPase	GTP hydrolysing enzyme
HAP	Heme activator protein
HAT	Histone acetyltransferase
HDAC	Histone deacetyltransferase
HFB	Hydrophobin
H. jecorina	*Hypocrea jecorina*
HKMT	Histone lysine methyltransferase

HMG	High mobility group
HsbA	Hydrophobic surface binding protein A
HSP70	Heat shock protein 70
IGV	Integrative Genomics Viewer
INO80	Inositol biosynthesis
IRE1	Inositol-requiring protein 1
KEGG	Kyoto encyclopaedia of genes and genomes
LSD	Lysine-specific demethylase
MAPK	Mitogen-activated protein kinase
MAT	Mating-type
MFS	Major Facilitator superfamily
M. grisea	*Magnaporthe grisea*
MH site	Micrococcal nuclease hypersensitive site
MNase	Micrococcal nuclease
MOD1	Morphology and development
mRNA	Messenger RNA
MS	Mass spectrometry
NAT	Natural antisense transcript
NC	No carbon source
NDR	Nucleosome-depleted region
NFR	Nucleosome-free region
N. crassa	*Neurospora crassa*
NTG	N-nitrosoguanidine
ORF	Open reading frame
PAK	P21-activated kinase
PCR	Polymerase chain reaction

PDI	Protein disulphide isomerase
P (D/X/C/G)A	Potato (dextrose/xylose/cellobiose/glycerol) agar
PEG	Polyethylene glycol
PL	Polysaccharide lyase
PRMT	Protein arginine methyltransferase
qRT-PCR	Real-time PCR
RAC1	Ras-related C3 botulinum toxin substrate 1
RCAF	Replication-coupled assembly factor
RGS	Regulator of G-protein signalling
RISC	RNA-induced silencing complex
RITS	RNA-induced initiation of transcriptional gene-silencing
RNA	Ribonucleic acid
rRNA/tRNA	ribosomal RNA/transfer RNA
RNAi	RNA interference
RNAP	RNA-polymerase
RPKM	Reads per kilobase of exon model per million mapped reads
RT-PCR	Reverse-transcription PCR
SAM	S-adenosyl-L-methionine
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
siRNA	Silencing RNA
SNF	Sucrose non-fermenting
S. pombe	*Schizosaccharomyces pombe*
ssPCR	Strand-specific PCR

SSPE	Saline sodium phosphate EDTA
SUMO	Small ubiquitin-like modifier
SWI/ISWI	Switch/Imitation switch
SWO	Swollenin
T. atroviride	*Trichoderma atroviride*
T. aurantiacus	*Thermoascus aurantiacus*
TCA cycle	Tricarboxylic acid cycle
TCM	*Trichoderma* complete medium
TEM	Transmission electron microscopy
TF	Transcription factor
T. harzianum	*Trichoderma harzianum*
T. koningii	*Trichoderma koningii*
TMM	*Trichoderma* minimal medium
T. reesei	*Trichoderma reesei*
TSS	Transcription start site
T. terrestris	*Thielavia terrestris*
T. virens	*Trichoderma virens*
T. viride	*Trichoderma viride*
UDP	Uridine diphosphate
UPR	Unfolded protein response
UV	Ultraviolet
XlnR	Xylanolytic regulator
XYN	Xylanase
XYR1	Xylanase regulator 1

Chapter 1: General introduction

1.1 Context of the research in this thesis

The need to replace fossil fuels with more environmentally friendly alternative fuels such as bioethanol has informed government policy and let to new research (Dashtban *et al*., 2009). The production of bioethanol from lignocellulosic materials encompasses several steps, one of which involves the degradation of pre-treated biomass by fungal enzymes. Most of the commercially-available enzymes for that purpose are secreted by the filamentous fungus *Trichoderma reesei*. Despite the commercial importance of those proteins, the regulation of the corresponding genes remains under-explored. Thus, *T. reesei* has been used in this project to study the regulation of genes important for carbohydrate degradation at the transcriptional and post-transcriptional level when grown on various different substrates, including wheat straw, a potential source of lignocellulosic material for bio-ethanol production.

1.2 First and second generation biofuels

Fossil fuels are the main energy source of the 21st century but burning of those materials is accompanied by many, recognised problems affecting the global economic and environmental health. Replacing fossil fuels with biofuels will help to reduce global CO_2 emissions, have a favourable greenhouse gas profile, decrease dependence on foreign oil and depleting oil resources, avoid security issues and promote local economies (Banerjee *et al.*, 2010; Fukuda *et al.*, 2009). Currently bioethanol and biodiesel are produced from sucrose in sugar cane and sugar beet or hydrolysed starch (first generation bioethanol, 1G) (Weber *et al.*, 2010). Worldwide 1G bio-ethanol production reached ~74 billion litres in 2009 and was expected to reach 85.9 billion litres in 2010 (Ayrinhac *et al.*, 2011).

The production of biofuels from lignocellulosic plant biomass, that does not compete for food-use, is called second generation (2G) and is an environmentally friendlier alternative to petroleum-based energy sources (Alvira *et al.*, 2010). In contrast to 1G bioethanol, lignocellulosic wastes do not compete with food crops or use additional arable land and they are less expensive than conventional agricultural feed stocks (Alvira *et al.*, 2010; Banerjee *et al.*, 2011; Margeot *et al.*, 2009). Lignocellulosic wastes are produced by the forestry, pulp and paper and agriculture industries in

addition to municipal and animal wastes and are energy sources for conversion to biofuels (Dashtban *et al.*, 2009). Other potential energy sources are grasses such as Switchgrass and Miscanthus (Dodd and Cann, 2009) as well as hardwood and softwood (Zhu *et al.*, 2010). The use of lignocellulose to generate 2G bioethanol is a more logical alternative to 1G biofuels, as it is the most abundant material on the planet and its production has been estimated at 1×10^{10} MT per year (Alvira *et al.*, 2010). Research on biofuel production is on-going but process optimisation is required as the costs of 2G biofuels cannot compete with oil yet.

1.3 Lignocellulose composition

Lignocellulose is the major structural component of plant cell walls (Figure 1.1) and is made of cellulose (40% – 60%), hemicelluloses (20% - 40%) and lignin (10% - 25%) and small quantities of other materials such as proteins and pectin (Dashtban *et al.*, 2009; Mathew *et al.*, 2008). Plant wall composition ratios depend on the plant type and species (Zhu *et al.*, 2010).

Cellulose is the most abundant biopolymer on earth and consists of long chains of glucose connected by β-1,4-glycosidic bonds. These chains can be coupled to each other by hydrogen bonds, hydrophobic interactions and van der Waal's forces to form crystalline structures known as microfibrils (Dashtban *et al.*, 2009). Microfibrils can then be stacked on top of each other to create a highly recalcitrant, crystalline cellulose structure (Liu *et al.*, 2010). Hemicelluloses and lignin act as a glue to bind the cellulose-based structure, and so building a robust matrix which makes up the plant cell wall (Zhu *et al.*, 2010).

Hemicelluloses are the second most common polymers in lignocelluloses and are found predominantly in hardwoods and cereals. In contrast to cellulose, hemicelluloses consist of a heterogeneous mix of pentose sugars (xylose, arabinose) and hexose sugars (mannose, glucose, galactose) and several acids (Dashtban *et al.*, 2009). Hemicellulose composition is very variable and depends on plant type and species (Dodd and Cann, 2009). Xylose is the most abundant sugar in hemicellulose followed by arabinose (Seiboth *et al.*, 2011). The backbone of hemicellulose is made of either xylan, consisting of β-1,4 linked xylose monomers, or mannan which consists of β-1,4-linked glucose and mannose units (Juhasz *et al.*, 2005). Depending on the source of the hemicellulose, the backbone chain can be decorated with sugar residues such as arabinose and galactose residues; and/or acids such as glucuronic,

coumaric and ferulic acids as well as acetyl groups (Dodd and Cann, 2009; Knob *et al*, 2010).

Arabinoxylans which include cop plants such as wheat, have a xylose backbone with arabinose residues linked to the *O*-2 and *O*-3 of xylose (Dodd and Cann, 2009). Xylose residues can be singly or doubly substituted with arabinose residues and the 5'-OH of arabinofuranosyl groups can be esterified to ferulic or coumaric acid groups. Ferulic and coumaric acids can form covalent links to lignin or other ferulic and coumaric acid groups of arabinose residues, thus re-inforcing the plant cell wall structure. Glucuronoxylans which include hardwoods, herbs and woody plants contain a xylose backbone with glucuronic acid linked to the *O*-2 of xylose (Dodd and Cann, 2009). Arabinoglucuronoxylans are mainly found in grasses such as *Miscanthus giganteus* and contain arabinose, glucuronic acids and acetyl side chains linked to the xylose backbone.

Lignin is the third major component of lignocellulosic biomass is comprised of three aromatic alcohols: mainly coniferyl, sinapyl and p-coumaryl alcohols (Dashtban *et al.*, 2009). Lignin is the material which confers rigidity to plant cell walls by cross-linking through hydroxycinnamic acids, such as ρ-coumaric acid and ferulic acid, to cell wall polysaccharides and also to xylan, thus acting as protection against cell wall degrading enzymes (Dodd and Cann, 2009). It is the most recalcitrant material found in lignocellulosic biomass.

Plant-cell walls consists of a heterogenous mix of polysaccharides which possess variations in the type and linkage of the backbone sugars as well as the presence of other sugars and/or acetate decorations; this is all dependent on the plant species, cell type and differentiation state (Boraston *et al.*, 2004).
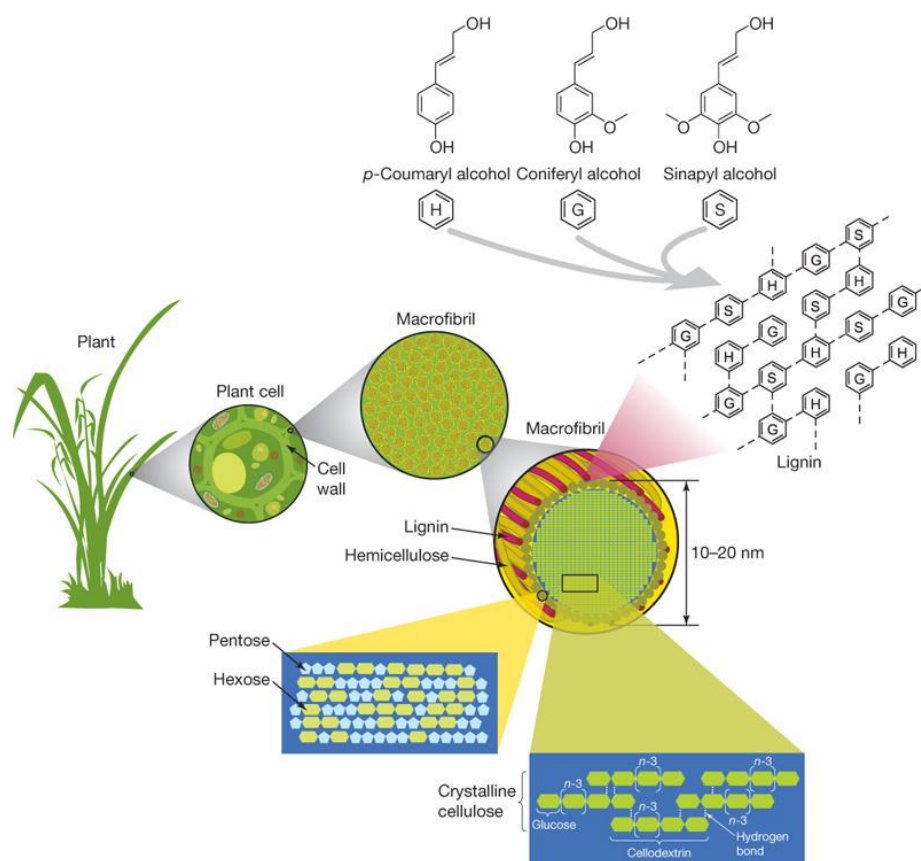
Figure 1.1: Diagrammatic representation of the structure of plant cell wall and associated cellulose, hemicellulose and lignin components. Taken from Rubin (2008).

1.4 Plant biomass degradation by fungal enzymes

The glycosidic linkages of plant cell wall sugars exhibit half-lives at room temperature on the order of 20 million years as was shown by assessing spontaneous hydrolysis of β-O-glucofuranoside bonds found in cellulose chains at different temperatures and between pH 7 and 14 (Dodd and Cann, 2009; Wolfenden *et al.*, 1998). In nature, the degradation of plant biomass is catalysed by enzymes from various different microorganisms including many species of bacteria and fungi (Dodd and Cann, 2009). Saprobic fungi, including some species of *Trichoderma* and *Aspergillus,* play important roles in the degradation of dead biomass by secreting cell-free, non-complexed enzymes (Elkins *et al.*, 2010).

Carbohydrates can exist in the form of mono-, di-, oligo- and polysaccharides, allowing a great number of possible combinations resulting in many different structures. Plant cell wall carbohydrates contribute to overall structure and rigidity, protection against viral infections and lytic enzymes and also act as

food storage compounds (Henrissat and Davies, 1997). In order to be able to degrade the diverse array of carbohydrate bonds present in plants, filamentous fungi such as *T. reesei* have evolved a wide range of enzymes able to cleave these bonds (Bourne and Henrissat, 2001). These enzymes are known as glycoside hydrolases (GHs), carbohydrate esterases (CEs) and polysaccharide lyases (PLs) and are classed and described in the carbohydrate active enzyme database (CAZy - www.cazy.org).

The majority of genes encoding plant cell wall-degrading enzymes in the genomes of filamentous fungi such as *T. reesei* and *A. niger* are glycoside hydrolases. Currently 130 families and 14 clans of GHs have been described (www.cazy.org; see below for details). GHs can act on bonds between two carbohydrates or on bonds between a carbohydrate and a non-carbohydrate moiety and many of these enzymes are poly-specific. One of the first classifications proposed for GHs was based on their substrate specificity as defined by the Union of Biochemistry and Molecular Biology; this is expressed by an EC number for each enzyme. This classification does not reflect the capability of some enzymes to act on several different substrates and does not give any information on their structure. In the CAZy system, glycoside hydrolases are classed into families based on their primary amino acid sequence which allows proteins with conserved folds and catalytic mechanisms to be grouped together in the same family (Henrissat and Davies, 1997). Furthermore, 3D structures of proteins are often more conserved than their amino acid sequence and therefore, glycoside hydrolases from different sequence-based families can be grouped into the same clan. The CAZy database gives detailed information about the characteristic folds found in each clan. Glycoside hydrolases can also be divided into three different types when considering their topological arrangement of the catalytic site: a tunnel suited for processive exo-attack, a cleft allowing endo-attack and a crater/pocket required for end-on-attack (Hilden and Johansson, 2004).

Glycoside hydrolases are globular proteins consisting of a catalytic domain (CD) and some enzymes are attached to a carbohydrate binding module (CBM) via a highly glycosylated linker region. CBMs can further be classed into 64 different families based on their primary amino acid sequence and vary in size between 4 and 20 kDa. CBMs generally have three roles which are important for enhanced catalysis: the proximity effect (bringing the CD domain in close contact with the substrate), a targeting function and a disruptive function (Boraston *et al.*, 2004; Hervé *et al.*, 2010). For GHs which have a CBM, these modules are essential for a fully functional enzyme as the loss of the Cel7a CBM (*T. reesei*) has been shown to severely diminish enzyme activity on crystalline cellulose (Hildén and Johansson, 2004). All CBMs use

aromatic or polar residues to bind to the hydrophobic surface of the substrate. Anchoring of the enzyme to the substrate allows either 1D or 2D (1 or 2-dimensional) movement across the carbohydrate surface allowing the enzyme to slide around to find new areas suitable for degradation. There are also differences in the binding affinity: Cel7a and Cel6a from *T. reesei* (encoded by *cbh1* and *cbh2*) both have family 1 CBMs but the rate of desorption is very different (Hildén and Johansson, 2004).

Glycoside hydrolases operate via two different mechanisms (inverting or retaining) using water for acid/base assisted catalysis and nucleophilic attack (Figure 1.2). Both mechanisms use two carboxylic acids, a glutamate and an aspartate residue, and procede through oxocarbonium ion-like transition states (Knob *et al.*, 2010). The inverting reaction employs a one-step displacement mechanism through acid/base catalysis and the anomeric configuration of the substrate is inverted. The retaining mechanism on the other hand, is a two-step displacement reaction where a covalent glucosyl-enzyme intermediate is formed. During the first step of the mechanism (glycosylation), the acid-base catalyst protonates the glycosidic oxygen and the nucleophile then attacks the substrate to form the covalent enzyme-glycosyl intermediate (Knob *et al.*, 2010). In the second step (deglycosylation), the acid/base catalyst deprotonates a water molecule which then attacks the substrate and releases a sugar molecule. In this mechanism, the substrate retains its original conformation.

Apart from the GHs, polysaccharide lyases (PLs) and carbohydrate esterases (CEs) also cleave bonds found between plant cell wall components but they use a different mechanism than GHs. PLs cleave uronic acid-containing polysaccharide chains via a β-elimination mechanism to generate an unsaturated hexenuronic acid and a new reducing end (www.cazy.org). CEs catalyse the de-O or de-N-acetylation of substituted saccharides releasing sugar moieties. PLs and CEs can also be classed into different families based on their amino acid sequence (www.cazy.org).

In nature, filamentous fungi are presented with an array of polysaccharides and they therefore secrete many cell wall-degrading enzymes which are able to deconstruct the respective carbon source. As a result, hetero-synergy occurs between the enzymes, meaning that different hydrolases, esterases and lyases act synergistically on the main chain and the side chains of carbohydrates in order to degrade the substrate as much as possible. Fungi like *T. reesei* produce a multiplicity of enzymes with very similar functionalities (e.g. β-glucosidases, β-xylosidases) which are encoded by one gene each. These enzymes exhibit a diversity of physicochemical properties, structures,

specific activities and yields (Knob *et al.*, 2010). These so-called iso-enzymes may present different effectiveness in hydrolysing hemicelluloses and cellulose. Furthermore, broad substrate specificities are common amongst these enzymes, especially concerning hemicellulases. For example, the *T. reesei* BXL1 not only possesses xylosidase activity but also arabinofuranosidase and glucosidase catalytic activity.

Inverting mechanism for a β-glycosidase:
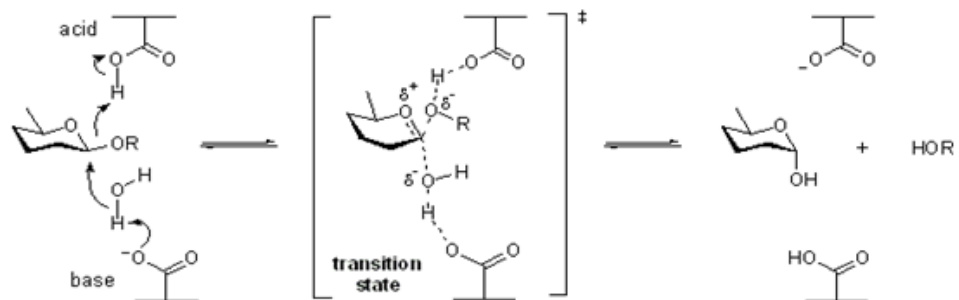
Retaining mechanism for a β-glycosidase:
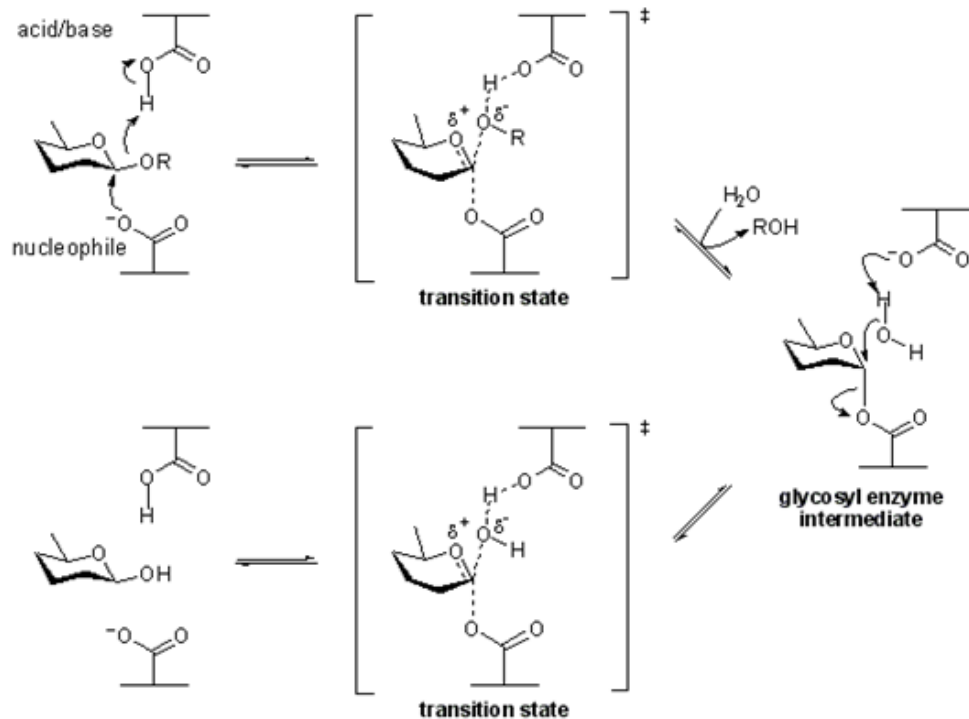
Figure 1.2: Diagram of the inverting and retaining mechnisms used in acid/base catalysis by glycoside hydrolases. Taken from www.cazypedia.org.

The three major types of GHs required for lignocellulosic biomass degradation are described below. Accessory enzymes such as swollenins, hydrophobins (discussed in Chapter 4) and copper-dependent monooxygenases (known as GH61 family enzymes, although not true GHs, see below) as well as GHs which cleave less frequent chemical bonds (e.g. arabinofuranosidases) also exist which help the breakdown of lignocellulosic biomass (LB) and are potential targets for lowering the overall enzyme production cost (Banerjee *et al*, 2010).

1.4.1 Cellulases

Cellulases are enzymes which degrade the cellulose fraction of plant cell walls. Three different types act in synergy to degrade cellulose into its individual glucose monomers (Figure 1.3). Endoglucanases (EGLs) start with cleaving internal glycosidic bonds within the cellulose chain, hence initiating cellulose breakdown and providing more free chain ends. Cellobiohydrolases (CBHs), then act on the reducing (CBH1) and non-reducing (CBH2) ends of the cellulose chains, releasing di-saccharides (cellobiose) or oligosaccharides which are further hydrolysed into glucose monomers by β-glucosidases (BGLs) (Dashtban *et al.*, 2009; Liu *et al.*, 2010; Mathew *et al.*, 2009).

1.4.2 Hemicellulases

Hemicellulases are enzymes which degrade the hemicellulose part of the lignocellulosic material. Due to the more variable nature of hemicellulose, more enzymes exist which act on it. Xylanases and mannanases cleave the β-1,4 linkages in xylan or mannan and release oligosaccharides which can further be hydrolysed into xylose and mannose monomers by β-xylosidases and β-mannosidases. B-xylosidases hydrolyse xylooligosaccharides and xylobiose from the non-reducing end liberating xylose, therefore preventing inhibition of xylanases by these substrates (Knob *et al.*, 2010). Furthermore, β-xylosidases also possess transglycosylation activity wherein monosaccharide units or alcohols (e.g. glycerol) are attached to or cleaved from xylose units. Endo-processive arabinanases cut randomly the backbone of arabinan, whereas exo-acting α-L-arabinofuranosidases release terminal, non-reducing α-L-arabinose residues (Seiboth *et al.*, 2011). Arabinofuranosidase as well as glucosidase activity can be found in some β-xylosidases (e.g. BXL1 of *T. reesei*) and some xylanases can also degrade arabinan (Seiboth *et al.*, 2011). Additional enzymes which act on hemicellulose residues include: α-galactosidase (releases galactose), α- and β-glucuronidases (remove glucuronic acid from xylose residues at the terminal, non-reducing end), acetyl xylan esterases and ferulic/coumaric esterases (Dashtban *et al*, 2009; Dodd and Cann, 2009). The α-glucuronidases are thought to work downstream of xylanases as they depend on them to generate their cognate substrates.

Ferulic/coumaric acid esterases, belonging to the carbohydrate esterase (CE) family 1, cleave ester bonds between ferulic and coumaric acid groups, thus breaking down some of the links between lignin and sugars and allowing access of other enzymes to the xylan backbone; acetyl xylan esterases release acetyl groups attached to the xylan backbone.

<u>1.4.3 Ligninases</u>

Ligninases are enzymes which degrade lignin. They are classified into two families: the phenol oxidases (laccases) and the peroxidases (lignin or manganese peroxidases) (Dashtban *et al.*, 2009). White-rot basidiomycetes are the most efficient producers of ligninases together with soft and brown-rot fungi. Lignin degradation is initiated through the production of free oxygen radicals by white, soft and brown-rot fungi. The oxygen radicals are produced through the Fenton reaction and then attack the lignin fraction of the plant cell wall as well as plant cell wall-associated saccharides and thus allow other proteins to access the lignocellulosic structure (Dashtban *et al.*, 2009). The genome of *T. reesei* encodes one secreted laccase (Transcript I.D. 54239) which has 43% identity at the protein level to the brown 2 laccase (Gene I.D. Afu1g15670) from *Aspergillus fumigatus* (*T. reesei* v.2.0 JGI database). This protein has not been characterised yet and no peroxidases are found in the genome of *T. reesei* and ligninases will not be discussed further in this project.
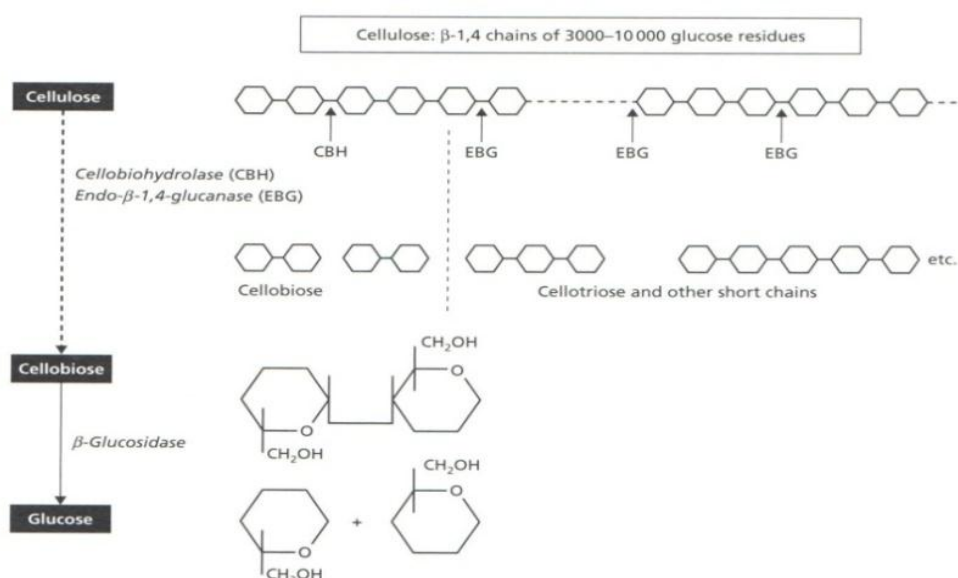
Figure 1.3: Diagram of the degradation of cellulose by the three different cellulases. Taken from Deacon (2006).

1.4.4 Copper-dependent monooxygenases (Family 61 of glycoside hydrolases)

The glycoside hydrolase 61s are a rather enigmatic type of enzyme family. Attention was first drawn to these proteins by their capacity to significantly enhance cellulase activity. Harris *et al.* (2010) showed that mixing the fermentation broth of *T. terrestris* with Celluclast (commercial cellulolytic cocktail from *T. reesei*: Novozymes) afforded a 2-fold reduction of total protein loading required for a 91% cellulose hydrolysis. Furthermore, a GH61 family member from *T. aurantiacus* also significantly increased CBH1, CBH2 and EGL1 activity. Despite their role in lignocellulose hydrolysis, GH61 enzymes lack the structure and residues typical for the catalytic site found in other glycoside hydrolases, and thus are probably not "true" glycoside hydrolases. There is no evidence of a cleft, tunnel, crevice or any catalytic acidic residues in these enzymes; they display instead a β-sheet fold with an extended planar face, the centre of which contains the N-terminal histidine with a metal ion binding site. GH61 enzymes are copper-dependent oxidoreductases which require the presence of small molecule redox-active cofactors such as gallate and ascorbate for enhanced catalytic activity and they also strongly bind a copper ion which is essential for enzymatic function (Quinlan *et al.*, 2011). A type II copper ion is bound at the N-terminal site, held in place by a PEG (polyethylene glycol) molecule, two histidines (one of which is methylated) and a tyrosine. GH61 family members initiate cellulose breakdown by generating oxidised and non-oxidised cellodextrin products which are easily acted upon by cellulases and thus differ significantly in their mode of action from conventional GHs which use acid/base catalysis to break glycosidic bonds.

The family of GH61s is unique to plant cell wall degrading filamentous fungi (the only exception is a bacterial GH61 enzyme found in *Crytpococcus neoformans*) and seems to be very ancient as they were found before the divergence of the two major fungal phyla, the Ascomycota and the Basidiomycota at around 600 million years ago (Harris *et al.*, 2010). It has been shown that most GH61s are unable to degrade lignocellulosic compounds without other enzymes present; the only exception to this is the *T. reesei* GH61A which was shown to have very limited endoglucanase activity (Karkehabadi *et al.*, 2008). The *T. reesei* genome encodes six GH61 proteins of which two have been characterised in more detail: GH61A possesses a CBM1 module whereas GH61B (EGL7) lacks a CBM module altogether and is secreted as a monomer.

As mentioned above, GH61 family members do not show much conservation with other glycoside hydrolases but instead were shown to be structurally

similar to the *Serratia marcescens* chitin binding protein 21 (CBP21) belonging to the CBM33 family; this protein stimulates chitinase activity but possesses no hydrolytic activity itself (Karkehabadi *et al*, 2008).

1.5 Biofuel production from plant biomass deconstruction

In terms of industrial applications, the released sugars from the degradation of cellulose and hemicellulose (glucose, mannose, galactose, xylose and arabinose) can all be fermented to biofuels. The degradation of lignin into aromatic compounds serves as the basis for the production of compounds such as adhesives and biopolymers (Zhu *et al.*, 2010). Furthermore lignocellulose degradation can result in the release of xylooligosaccharides which are used in food industries; and xylose can be converted to xylitol, a natural food sweetener (Knob *et al.*, 2010).

Ethanol production from lignocellulosic plant material comprises 4 main steps (Figure 1.4): 1) pre-treatment of the raw material, 2) enzymatic hydrolysis 3) fermentation and 4) purification and distillation of ethanol and other compounds (Margeot *et al*, 2009). Research in all four steps is done in concordance with scaling-up and integrating processes to minimise energy and water demand; as well as establishing reliable data for cost estimation and the determination of environmental and socio-economic impacts (Margeot *et al*, 2009).

1.5.1 Pre-treatments

The aim of pre-treatment is to solubilise the lignin and hemicellulose fractions and disrupt the crystalline structure of cellulose in order to render cellulose more accessible to enzymes. Various pre-treatment technologies have been developed and they can be classified into biological (lignin-degrading fungi), physical (mechanical to decrease particle size), chemical (acids, alkalis and solvents) and physico-chemical [steam or CO_2 explosion, ammonia fibre explosion (AFEX), wet oxidation, microwave or ultrasound pre-treatment, liquid hot water and sulphite pre-treatment to overcome recalcitrance of lignocellulose (SPORL)] pre-treatments (Alvira *et al.*, 2010; Zhu *et al.*, 2010). A combination of these methods is also possible (Zhu *et al.*, 2010).

1.5.2 Enzymatic hydrolysis

During the second step, the pre-treated raw material is subjected to enzymatic hydrolysis by cellulases, hemicellulases and even ligninases. They release hexose and pentose sugars from the pre-treated biomass and can also

act on any remaining solid fraction. Cellulases and hemicellulases (free, non-complexed enzymes) are mainly obtained from *Trichoderma* spp. and *Aspergillus* spp. but can also be purified from *Humicola* spp., *Fusarium* spp. or from bacteria such as *Bacillus subtilis* and *Clostridium thermocellum* in the form of the cellulosome (Banerjee *et al.,* 2010; Weber *et al.*, 2010). Depending on the pre-treatment method used and the resulting fraction of polysaccharides obtained from it (Zhu *et al.*, 2010); cellulases and/or hemicellulases are used to solubilise the released polysaccharides fraction (Banerjee *et al.*, 2009; Alvira *et al.*, 2010).

1.5.3 Fermentation and product purification

The monomeric sugars (arabinose, xylose, glucose, mannose, galactose) released during enzymatic hydrolysis can then be used to be fermented into valuable products such as ethanol (Dashtban *et al.*, 2009). The preferred organism to be applied during fermentation is *Saccharomyces cerevisiae* which is able to ferment hexose sugars (e.g. glucose) to ethanol. *S. cerevisiae* has been genetically engineered to contain genes from other xylose-fermenting yeasts and arabinose-metabolising microorganisms in order to ferment pentose sugars such as xylose and arabinose to ethanol (Dashtban *et al.*, 2009; Fukuda *et al.*, 2009). Separation of ethanol from the bioreactor is carried out by distillation which yields up to 93% pure ethanol.
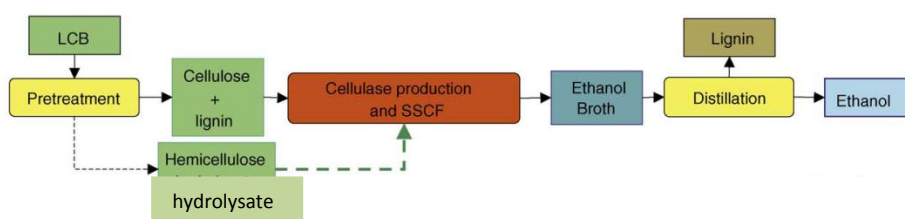
Figure 1.4: Flow diagram illustrating the concept of lignocellulosic biomass conversion to ethanol. Dotted lines represent optional process configurations (depends on pre-treatment used). LCB = lignocellulosic biomass, SSCF = simultaneous saccharification and co-fermentation process. Taken from Margeot *et al*. (2009).

1.6 *Trichoderma reesei* taxonomy, genome and industrial applications

T. reesei is a mesophilic, filamentous fungus and the anamorph of *Hypocrea jecorina* and grows aerobically in the soil on dead plant biomass (Figure 1.5). It is classified into the division of *Ascomycota*, the class of *Sordariomycetes*, the order of *Hypocreales* and the genus *Trichoderma*. At present, the International Sub-commission on *Trichoderma/Hypocrea* lists 104 species which have been described at the molecular level (Schuster *et al.*, 2010). Most species in the *Trichoderma* genus are saprobes; but plant and mushroom pathogens (*T. koningii*) and species which feed on other fungi (*T. viride*, *T. harzianium*, *T. atroviride*) have also been identified. The applications of *Trichoderma spp.* are numerous; some are important biological control agents in agriculture (*T. atroviride* and *T. virens*) and others have been used for enzyme production (*T. reesei*). Proteins such as β-glucanases are used to improve the brewing process, hemicellulases and pectinases are macerating enzymes in fruit juice production and cellulases are mainly applied in baking, malting and grain alcohol production (Schuster *et al.*, 2010).

Only recently *T. reesei* industrial strains (QM6a, RutC30, QM9414, and TU-6) were shown to be heterothallic (*MAT1-2* locus) and capable of mating with *H. jecorina* isolates with a *MAT1-1* mating locus in the presence of light (Seidl *et al.*, 2009).

The sequenced and annotated genome of *T. reesei* was published in 2008, and within it are encoded 201 glycoside hydrolases, 103 glycosyl transferases, 36 CBMs, 22 carbohydrate esterases and 5 polysaccharide lyases; this is less than the numbers found in some other Sordariomycetes and Eurotiomycetes such as *Aspergillus spp*. (Martinez *et al.*, 2008; Häkkinen *et al*., 2012). This trend is further confirmed when looking at the number of genes encoding cellulases (10) and hemicellulases (16). Most of the genes which encode plant cell wall-degrading enzymes are found in clusters within the genome. This facilitates the regulation of co-expression of these genes through chromatin-associated changes (see Chapter 5) and allows the secretion of enzymes which work together to efficiently degrade lignocellulosic biomass once they are secreted (Martinez *et al.*, 2008).

Although the genome of *T. reesei* encodes fewer enzymes involved in carbohydrate degradation, it is very successful in the niche it occupies. This is partially due to secondary metabolite secretion as well as lytic and proteolytic enzyme secretion. More than 100 metabolites with antibiotic activity such as polyketides, pyrones and terpenes have been found in *Trichoderma spp*

(Schuster *et al.*, 2010). *T. reesei* was first isolated from the Solomon islands in World War II because of its degradation of cotton canvas and garments of the US army (Seidl *et al.*, 2009). Nowadays, the enzyme mix secreted by *T. reesei* is employed in the pulp and paper (to reduce chlorine utilisation), textile and food industries (to improve digestibility and arabinoxylan degradation) (Knob *et al.*, 2010). Cellulases account for approximately 20% of the global enzyme market (Matthew *et al.*, 2009). Cellulases secreted by *T. reesei* include two CBHs (CBH1 and CBH2), five EGs (EGL1 – EGL5) and two BGLs (BGL1 and BGL2) (Liu *et al*, 2010). The CBHs and EGs consist of up to 90% of the total number of the secreted enzyme mix with CBH1 constituting up to 60% of the number of enzymes secreted by *T. reesei* strains (Rahman *et al.*, 2009). BGL typically makes up less than 1% of the total number of enzymes secreted (Margeot *et al.*, 2009). Secreted protein levels of up to 100 g/L have been reported for some *T. reesei* strains under production conditions (Banerjee *et al.*, 2009). Due to its industrial importance and the already widespread application of its enzymes, *T. reesei* is an important fungus for use of its enzymes in biofuel production.
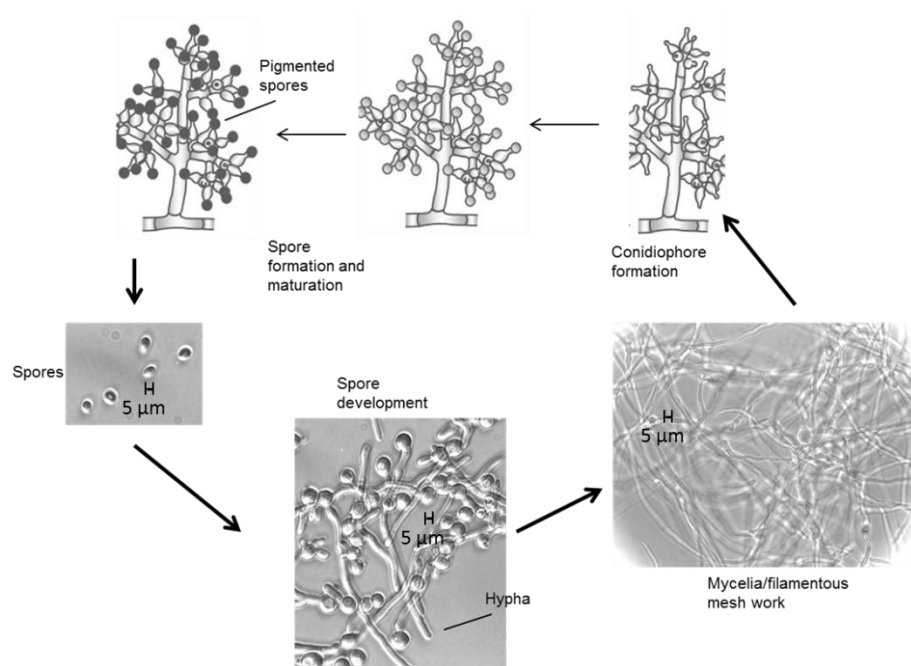
Figure 1.5: Asexual life cycle of *T. reesei*. The top 3 pictures are adapted from Steyaert *et al.*, (2010). The bottom 3 pictures were taken (Chapter 3, Materials and Methods) of QM6a spores (0 h) and mycelia (8 h and 22 h) when grown in TMM supplemented with 2% (w/v) glucose.

1.7 Induction of individual *T. reesei* cellulase and hemicellulase-encoding genes

There are several different carbon sources which have been shown to induce expression of genes encoding cellulases and hemicellulases. Glucose depletion has also been found to induce these genes anytime between 10 h and 30 h after it has all been used up (Margolles-Clark *et al.*, 1997; Mach *et al.*, 2003; Kubicek *et al.*, 2009).

Cellulose can serve as an inducer of cellulase-encoding genes. Cellulose is a polymer and cannot be taken up directly by the fungus but several different theories exist which might explain the secretion of cellulases in the presence of an insoluble substrate. One of these theories suggests there is a basal level of expression of cellulase-encoding genes, mainly *cbh1* and *cbh2* and other genes such as *cel5b*, which can start the degradation of cellulose in the absence of an inducer (Kubicek *et al.*, 2009). Further cleavage of the cellulose chains by these enzymes releases low molecular weight products such as soluble oligosaccharides or disaccharides (cellobiose, sophorose) which can be taken up by the cell and act as inducers. Conidia of *T. reesei* were proposed to contain CBH2 as well as CBH1, EGL2 and BGL1 on the cell surface, and those enzymes may cleave off soluble inducers from cellulose (Suto *et al.*, 2001). In this way, spores can induce the cellulase genes, germinate and break down crystalline cellulose (Suto *et al.*, 2001). Furthermore, it is thought that the release of an inducing sugar from the fungal cell wall under starvation conditions, a response which is unrelated to carbon catabolite repression, can also induce the cellulase genes. Glucose depletion in pre-grown cultures was shown to induce cellulases whereas growth in media with no carbon source did not induce the cellulolytic enzymes (Ilmén *et al.*, 1997). Never the less, the formation of a soluble inducer is required for full cellulase gene expression.

One of the strongest cellulase-inducing molecules in *T. reesei* is sophorose - a β-1,3 linked glucose disaccharide - a compound that is thought to be formed from the transglucosylation activity of BGL1 in the presence of cellobiose and cellulose (Mach *et al.*, 2003), although conflicting reports exist. The genome of *T. reesei* encodes 9 BGLs in total (Appendix A3 Table A.3.T.1). Sophorose induces an intracellular BGL, which was partially purified from the cell extract and showed sophorose hydrolysing activity at pH 6.5. This enzyme is likely to be involved in increasing glucose concentrations through hydrolysing sophorose (Suto *et al.*, 2001; Schuster *et al.*, 2011). Induction by sophorose is affected by various different parameters such as its concentration and rate of uptake and in the presence of glucose it is not taken up at all (Kubicek *et al.*, 2009). As sophorose is soluble, it induces cellulases much quicker (4 h to 6 h)

than in the presence of long chains of insoluble cellulose (20 h) polymers (El-Gogary *et al.*, 1989; Torigoi *et al.*, 1996). Cellobiose and sophorose can be taken up by a cellobiose permease which has low affinity and a low rate of uptake for sophorose. This suggests that sophorose is preferably taken up at very low concentrations and that the permease competes with β-glucosidases for this compound (Kubicek *et al.*, 2009). Recently, Zhou *et al*. (2012) showed that deletion of the intracellular β-glucosidases BGL2 and CEL1B (Appendix A3 Table A.3.T.1) as well as BGL1 led to a significant delay in *cbh1* induction. They further showed that addition of cellobiose to cultures of mutant strains growing on Avicel cellulose rescued this induction defect. Zhou *et al*. (2012) hypothesised that intracellular and extracellular β-glucosidases work together to ensure the correct balance of cellobiose production and metabolism which is required for cellulase and endoglucanase induction in the presence of cellulose. Further support for this theory comes from observations that transglycosylation activities of β-glucosidases were only shown *in vitro* and in the presence of very high concentrations of sugars (Zhou *et al*., 2012).

Another substrate which has been found to be capable of inducing the expression of genes that encode hydrolytic enzymes such as *cbh1*, and *bga1* (β-galactosidase) in *T. reesei* is lactose - although induction by this substrate is not as efficient as induction by sophorose (Ilmén *et al.*, 1997; Portnoy *et al.*, 2011). Lactose is not present in the natural environment of *T. reesei* but glycosydic bonds are found in plant cell walls which resemble the one in lactose, e.g. galactan side chains of pectins and arabinogalactans (Stricker *et al.*, 2007). Extracellular β-galactosidase (GH family 35), encoded by *bga1*, cleaves this substrate into D-galactose and D-glucose which can then both be metabolised. *T. reesei* does not possess a lactose permease, thus cleavage of lactose prior to uptake is essential. D-glucose can directly enter glycolysis whereas D-galactose can either enter the Leloir pathway or an alternative pathway where it has been shown to induce cellulase gene expression (Figure 1.6) (Seiboth *et al.*, 2007). The first step in the alternative pathway is the reduction of D-galactose to galactitol by XYL1 (xylose reductase 1), an enzyme which has also been shown to be important for pentose (e.g. D-xylose and L-arabinose) utilisation. Galactitol is the main inducer for extracellular *bga1* and the expression of this gene can also be up-regulated in the presence of L-arabinose, D-xylose, D-galactose and lactose (Fekete *et al.*, 2007). Furthermore, L-sorbose, an intermediate formed in the alternative pathway, has been shown to regulate and induce cellulase gene expression at the transcriptional level (Kubicek *et al.*, 2009; Furukawa *et al.*, 2009). The first step in the Leloir pathway converts β-D-glactose to α-D-galactose by mutarotase activity which is absent in *T. reesei* when grown in the presence of lactose (Fekete *et al.*, 2008). Instead, chemical mutarotation can convert β-D-glactose

to α-D-galactose but this depends on the surrounding water environment and presents a rate limiting step for lactose metabolism; thus the alternative pathway is preferred. Furthermore, increasing mutarotase activity in *T. reesei* transformants impaired cellulase gene expression in the presence of lactose (Fekete *et al.*, 2008). The second step of the Leloir pathway is catalysed by a galactokinase encoded by *gal1*; deletion of this gene abolished cellulase expression. Cellulase gene expression levels in lactose are lower than in the presence of cellulose and sophorose. It has also been shown that cellulase gene induction is highest during low growth rates in the presence of lactose due to extracellular BGA1 activity (Seiboth *et al.*, 2007).

Xylans as well as xylobiose, xylo-oligosaccharides and L-arabitol induce the expression of different hemicellulose-encoding genes such as *xyn1*, *xyn2*, *bxl1* and genes that encode other hemicellulose-degrading enzymes such as arabinofuranosidases (Aro *et al.*, 2003). The low constitutive basal expression of *xyn2* has led to sufficient enzyme capable of cleaving inducing oligosaccharides such as xylobiose or D-xylose from the xylan backbone, and which can then fully induce a sub-set of hemicellulase genes required for the degradation of hemicelluloses into individual sugars such as D-xylose, L-arabinose or D-galactose (Margolles-Clark *et al.*, 1997; Stricker *et al.*, 2008). The first step in the metabolism of these sugars is the same and involves the reduction to xylitol, L-arabitol and galactitol by XYL1. As mentioned before, galactitol serves as an inducer of *bga* (the product of which can also have BXL activity) and it is thought that L-arabitol is also one of the intracellular inducers of genes encoding xylanases and arabinolytic enzymes (Mach-Aigner *et al.*, 2011). D-xylose also acts intracellularly as an inducer of hemicellulose-encoding genes but this depends on its concentration (Mach-Aigner *et al.*, 2010). Higher levels of hemicellulolytic gene expression are achieved with lower concentrations of xylose, ranging between 0.5mM and 1mM. Furthermore, regulation of this pathway is controlled by CCR (carbon catabolite repression) which indirectly exerts control on *xyl1* and directly represses *xyn1*.
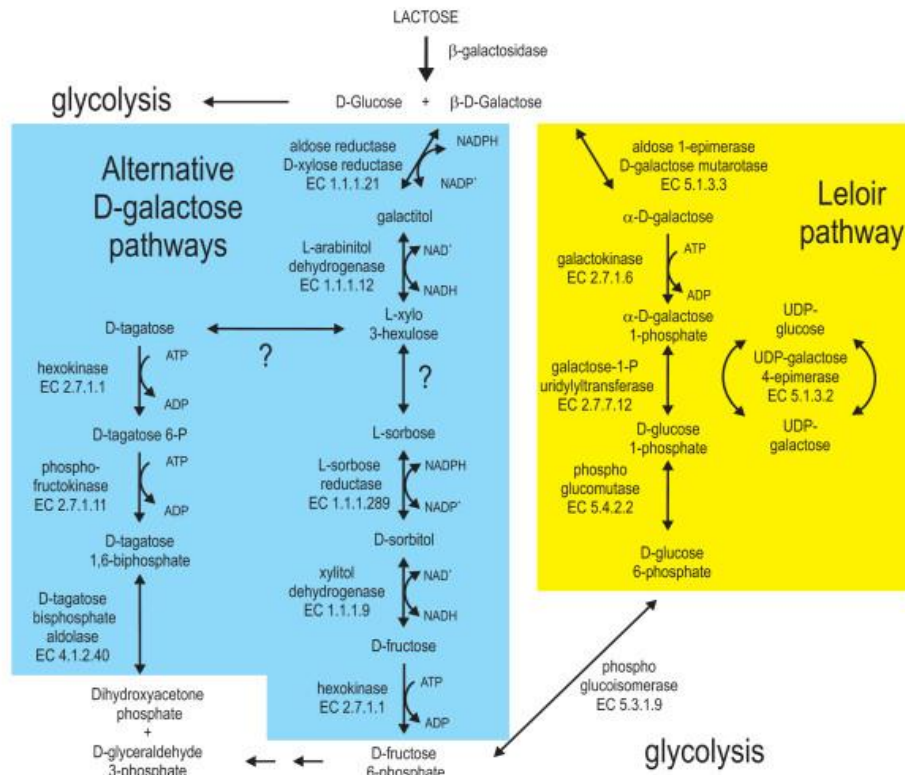
Figure 1.6: Flow diagram of the two pathways for lactose metabolism in *T. reesei*. All steps of the alternative pathway (blue box) which starts with the conversion of galactose to galactitol as well as the Leloir pathway (yellow box) are shown. Taken from Kubicek *et al*. (2009).

1.8 Challenges in enzymatic saccharification by *T. reesei* and other fungi

In each step of bio-ethanol production, improvements have been made but barriers remain for it to become a cost-competitive process which can be applied at an industrial scale. *T. reesei* is important for the second step of enzymatic saccharification and hence only this step will be discussed hereafter.

Enzymatic hydrolysis of lignocellulosic material should be cheap and should use an array of highly efficient and thermostable enzymes capable of degrading cellulose, hemicelluloses and lignin for the production of sugars and aromatic compounds. The cost of production of these enzymes is a major limiting factor for industrial bio-ethanol production (Jeoh *et al.*, 2008; Klein-

Marcuschamer *et al.*, 2012). Several strategies are employed to reduce initial enzyme load or to produce higher amounts of enzyme.

Fungi have been genetically mutagenised by UV (ultra-violet) light, NTG (N-nitrosoguanidine) or SDM (site directed mutagenesis) and screened for secretion of higher amounts of cellulases (Dashtban *et al.*, 2009; Ike *et al.*, 2010). For example, one of the limitations in the *T. reesei* cellulase system is carbon catabolite repression (CCR). In the presence of glucose most cellulase and hemicellulase encoding genes are repressed. A good example of this is the expression of *cbh1* from *T. reesei* which is inhibited by glucose and cellobiose. To overcome this, new *T. reesei* strains are screened for high cellulase secretion in glucose (and other substrates) after rounds of UV irradiation (Ike *et al.*, 2010). The most prominent example of this was the screening for the hypercellulolytic enzyme secretor strain RutC30 after several rounds of UV irradiation and NTG (Le Crom *et al.*, 2009) and in which carbon catabolite repression is abolished due to the truncated repressor CRE1. CRE1 binding sites can also be deleted within promoters, thus rendering cellulase gene promoters resistant to CCR (Matthew *et al*, 2008; Le Crom *et al.*, 2009).

Individual enzymes have also been made more efficient and can be employed in enzyme mixes (Ayrinhac *et al.*, 2011; Piscitelli *et al.*, 2011). For example, one of the drawbacks of the *T. reesei* cellulase secretion system is that it secretes very low amounts of BGL1; hence BGL1 activity is very low within the secreted enzyme mix. Although several genes within the genome encode β-glucosidases (Appendix A3 Table A.3.T.1), their expression and protein secretion is not sufficient for industrial application. To improve this, *bgl1*, the most prominent β-glucosidase, has been placed under the control of other *T. reesei* promoters (Rahman *et al.*, 2009) to increase the amounts of BGL1 secreted; or β-glucosidases have been engineered by L-shuffling to be more efficient (Ayrinhac *et al.*, 2011). In the latter technique, gene variants are recombined without the use of any polymerase, employing an appropriate ligase instead, in order to generate DNA fragments. Ligated and non-ligated fragments are denatured again and undergo cycles of hybridisation-denaturation-ligation until full-length genes are built. In this way, parental gene fragments are recombined randomly and Ayrinhac *et al.* (2011) generated a BGL, which had a 242-fold increased k_{cat} for ρNPGlc (ρ-Nitrophenyl β-D-glucopyranoside), from a *Chaetomium globosum* glucosidase putative gene, *T. reesei bgl1* gene and a *Neurospora crassa* putative glucosidase gene. The *cbh1* promoter is one of the strongest with a high rate of expression; it has been used to drive expression of BGL and EGs to improve total cellulase secretion (Matthew *et al.*, 2008; Liu *et al.*, 2008).

Research has also focused on protein secretion systems which can secrete enzymes efficiently and in improving the *T. reesei* secretion system (Banerjee *et al.*, 2010; Piscitelli *et al.*, 2011). The use of homologous recombination (Rahman *et al.*, 2009) or heterologous protein expression systems (e.g. yeast) provide high yields of more (thermostable) proteins (Dashtban *et al.*, 2009, Jeoh *et al.*, 2008; Fukuda *et al.*, 2009). Introducing genes which encode other GHs not found in *T. reesei* into its genome thus enhancing the versatility of the secreted enzyme cocktail would be one example (Martinez *et al.*, 2008). After the discovery of a sexual cycle in *T. reesei*, sexual recombination could be used as a genetic tool to improve industrial *T. reesei* strains (Seidl *et al.*, 2009). Research also focuses on finding new, efficient enzymes from other organisms (bioprospecting) by mining databases or random collection of microorganisms which grow on plant biomass in nature (Banerjee *et al.*, 2010).

Based on the strategies described above another major goal of research is to engineer or find enzyme cocktails by the co-culturing of microorganisms for CBP (consolidated bioprocessing) and SSF (simultaneous saccharification and fermentation) – this is a combination of various types of engineered or native enzymes secreted by fungi, bacteria or both which act synergistically to degrade cellulose, hemicelluloses and lignin efficiently (Banerjee *et al.*, 2010; Elkins *et al.*, 2010). Research on ligninases (Piscitelli *et al.*, 2011) but also on cellulases and hemicellulases has been conducted as different cellulases have different degrees of resistance to inhibition by lignin, whereas hemicellulases and glucosidases are less affected (Alvira *et al.*, 2010). Furthermore, some animals are also capable of producing these enzymes (without harbouring symbiotic bacteria and fungi in their guts) as was shown for the marine wood borer, *Limnoria quadripunctata* (isopod crustacean known as Gribble). The hepatopancreas (organ involved in the production of digestive enzymes, nutrient uptake, carbon-reserve storage, ion transport, osmoregulation, heavy metal sequestration and production of the oxygen binding pigment hemocyanin) transcriptome of this organism has been shown to contain GHs from 12 families as well as hemocyanins which may play a role in lignin degradation (King *et al.*, 2010). This model is promising as it allows screening for new enzymes from organisms across all three kingdoms of life and also mirrors to some degree the array of enzymes found in nature where many different bacteria and fungi are found side by side.

1.9 The main aims and objectives of this project

T. reesei is one of the main organisms used for industrial enzyme production and a promising candidate for future biomass saccharification. Despite the numerous applications of *T. reesei* enzymes, much of the mechanism of regulation of the genes encoding plant cell wall-degrading enzymes remains unknown. This study therefore set out to test the hypothesis that the type of carbon source detected by the fungus elicits different regulatory and transcriptional responses of genes encoding enzymes involved in plant biomass degradation. To gain a deeper understanding of the regulation of these genes, the main aims and objectives of this project were to:

- Investigate the influence of different sugars, which are found in plant cell walls, on spore development and mycelial growth in *T. reesei*
- Characterise the expression levels of genes which may be involved in spore development and establish whether their expression changes in the presence of different carbon sources
- Establish a possible link between the expression of genes encoding plant cell wall-degrading enzymes and fungal growth rates in the presence of different carbon sources
- Sequence and analyse the transcriptome of *T. reesei* when grown in a simple carbon source (glucose) and a complex lignocellulosic substrate (wheat straw) in order to: 1) find all GH, CE and PL-encoding genes significantly up-regulated in the presence of wheat straw; 2) discover genes encoding other, non-CAZy enzymes which may possibly be involved in detecting and signalling the presence of a lignocellulosic substrate as well as enzymes mediating the attachment of the fungus to the substrate
- Compare all findings of the previous point to the proposed model of carbohydrate degradation in *A. niger* in order to find similarities and differences in the mode of lignocellulose degradation employed by both industrially relevant fungi
- Investigate the regulation of different cellulase and hemicellulase-encoding genes in the presence of substrates which either induce or repress the *T. reesei* hydrolytic system
- Describe the chromatin structure in the *cbh1* promoter and coding regions in *T. reesei* and characterise changes in the structure which are related to gene inducing/repressing conditions.
- Investigate whether CRE1 is one of the factors involved in maintaining and/or contributing to observed changes in the chromatin structure in the *cbh1* promoter and/or coding regions during inducing and repressing conditions

Chapter 2: General Materials and Methods

2.1 Chemicals and reagents

The majority of chemicals used in this project were obtained from Sigma (Poole, Dorset, U.K.) and were of analytical grade or molecular biology grade. Other suppliers are referred to in the text. All solutions were made up in dd (double distilled) H_2O and autoclaved at 121°C for 15 min or 117°C (if carbon source is present) for 30 min. Millipore 2 μm filters were used to filter-sterilise certain solutions.

2.2 Strains

The wild type *Trichoderma reesei* QM6a (Druzhinina *et al.*, 2010), RUTC30 (see Figure 2.1), QM6a Δ*cre1* (wild-type with *cre1* knockout) and QM6a *cre1-1* (wild-type with the truncated *cre1*, from RUTC30) were used throughout this project. All strains were supplied by Roal Oy, Finland.
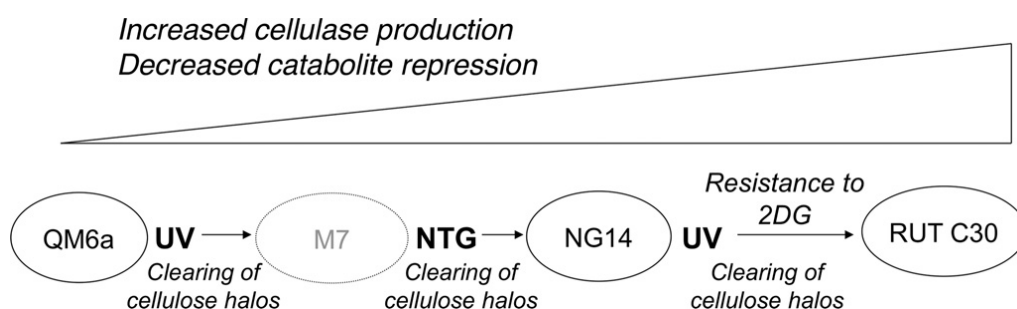
Figure 2.1: Mutational steps and intermediate strains leading to RUTC30. RUTC30 is a hypercellulosic mutant and carries many mutations (see Le Crom *et al.*, 2009) amongst which is a truncated *cre1* gene (NTG = N-nitrosoguanidine, UV = ultraviolet light). Taken from Le Crom *et al.* (2009).

2.3 Maintenance of *T. reesei* strains

All agar slopes were made by pouring 8 mL of molten agar into 20 mL Universal bottles (Sterilin) and leaving the agar to set at an angle. Conidia (glycerol stocks at -80°C) of *T. reesei* strains were resurrected and adapted on potato agar slopes (4.0 g/L potato extract, 15.0 g/L agar) complemented with 20.0 g/L of one of the following carbon sources: dextrose, cellobiose, xylose or glycerol (referred to as PDA, PCA, PXA or PGA). The spores were left to germinate for one week at 28°C. Once conidia were present they were harvested by washing the surface of the slopes with 2 mL 0.01% (w/v) Tween 80 and 100 µL of this spore suspension was used to inoculate new slopes. They were left to germinate at 28°C for another week and conidia from these slopes were used throughout experiments. Glycerol stocks were made by mixing 1 mL of original spore suspension with 500 µl of 60% (w/v) glycerol in a sterile cryovial, snap-freezing the mixture on dry ice and storing it at -80°C.

2.4 Cultivation of *T. reesei* strains

All strains were cultured in 250 mL Erlenmeyer flasks at 28°C, 150 rpm (time is indicated for each experiment) in *Trichoderma* complete medium (TCM): 10 g/L KH_2PO_4, 2 g/L $(NH_4)_2SO_4$, 2 g/L Bactopeptone, 2.5 g/L cellulose SigmaCell Type 20 (0.25% w/v) or 20 g/L (2% w/v) of the respective carbon source, 0.005 g/L $FeSO_4.7H_2O$, 0.0016 g/L $MnSO_4.H_2O$, 0.0014 g/L $Zn.SO_4.H_2O$, 0.0037 g/L $CoCl_{12}.6H_2O$, 0.6 g/L $MgSO_4$, 0.6 g $CaCl_2$; or in *Trichoderma* minimal medium (TMM) which is the same than TCM but without Bactopeptone and with 15 g/L KH_2PO_4 and 5 g/L $(NH_4)_2SO_4$. To obtain mycelia, TCM or TMM were inoculated with 10^5 spores/mL. Spore suspensions were obtained by washing the surface of the slopes as described in 2.3. Following vortexing of the spore suspension, a 10 µL sample was removed and an average count of the number of spores was determined using an "Improved Neubauer" haemocytometer (Marienfeld Laboratory Glassware, Lauda-Koenigshofen, Germany). The amount of spores present in 5 small squares was counted and the average taken. This was then divided by 16 and multiplied by 4 x 10^6. The resulting number was multiplied by the dilution factor which gives the concentration in spores/mL. The total number of spores wanted in a flask (e.g. 10^5 spores/mL in 100 mL TCM or TMM) was then divided by the spore concentration and the resulting amount of spore suspension added to each flask.

Wheat straw (Cordiale variety) was obtained from the Bioenergy and Brewing Science department (Sutton Bonington, Leicestershire, U.K.). The Cordiale wheat straw was milled using the Laboratory Mill 3600 (Perten, Sweden) and

passed through a sieve with a mesh size of 700 μm. Then, 5 g of the pre-milled straw was ball-milled in 80 mL stainless steel grinding bowls with 10-mm-diameter steel balls in a Planetary Mill (Pulverisette 5 classic line, Fritsch, Germany), at 400 rpm for 20 min. The resulting average particle size was ~75 μm. The Cordiale wheat straw is composed of 37% cellulose, 32% hemicelluloses, 31% lignin and retained a crystallinity of ~25% after ball milling.

Sophorose was obtained from AmsBio (Abingdon, Oxford, U.K.) and TMM or TCM was complemented with 250 mg/L (0.025% w/v) sophorose and 20 g/L (2% w/v) glycerol.

2.5 Harvest and storage of mycelia

Mycelia were separated from the culture liquid by pouring the content of each flask through Miracloth (Calbiochem® Merck, Darmstadt, Germany). They were rinsed twice with sterile ddH$_2$O before being blotted dry between paper towels and snap frozen in liquid nitrogen (University of Nottingham, Nottingham, U.K.).

Mycelia were then used in an assay, stored at -80°C or freeze-dried overnight.

2.6 Genomic DNA extraction from *T. reesei*

Mycelia were ground to a fine powder under liquid Nitrogen and samples of 100 mg or 200 mg of powder were weighed out. Each sample was re-suspended in 800 μL fungal DNA extraction buffer [50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 50 mM NaCl, 1% (w/v) SDS] and incubated at 65°C for 30 min. Samples were left to cool down for 5 min before being mixed with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) and shaken vigorously for 30 s at RT. Phases were separated by centrifugation at 15,000xg for 10 min. The aqueous (upper layer) phase was removed and the previous steps were repeated. The aqueous layer was removed again and the DNA precipitated with 0.7 volumes of isopropanol (Fisher Scientific, Loughborough, Leicestershire, U.K.) for 15 min and pelleted by centrifugation at 15,000xg for 20 min. Pellets were washed with 70% (v/v) ethanol, air dried and dissolved in 100 μL TE buffer [10 mM Tris (pH 8), 1 mM EDTA] on ice for one hour or overnight at 4°C. RNA was degraded with 50 mg/mL RNAse A (Qiagen, Crawley, Sussex, U.K.) at room temperature for 1 h. Purified DNA was

quantified on the Nanodrop ND-1000 spectrophotometer (Labtech, Ringmer, Sussex, U.K.) and stored at -20°C.

2.7 RNA extraction from *T. reesei*

Mycelia were ground to a fine powder under liquid N_2 and samples of 100 mg of powder were weighed out. Each 100 mg was re-suspended in 1 mL TriZol (Invitrogen, Carlsbad, CA, USA) and incubated at room temperature for 10 min. Then 0.25 mL chloroform (Fisher Scientific) was added and samples were mixed by vigorous shaking for 30 s and incubated at room temperature for 5 min before being centrifuged for 15 min at 15,000xg. The aqueous layer was removed and mixed with an equal volume of isopropanol and incubated at room temperature for 10 min. RNA was precipitated by centrifugation for 15 min at 15,000xg. Pellets were washed with 70% (w/v) ethanol and re-suspended in 100 μL – 400 μL DEPC-treated ddH_2O (1 mL DEPC in 1 L ddH_2O, incubated overnight at 37°C, then autoclaved twice at 121°C) for an hour at room temperature. RNA was purified with the Qiagen RNeasy® Mini Kit (Qiagen, Valencia, CA, U.S.A.) following the instructions of the "RNA clean-up" protocol with on-column DNA digestion. Purified RNA was quantified on the Nanodrop spectrophotometer and stored at -80°C.

2.8 Polymerase chain reaction (PCR)

All PCR reactions were run for 30 or 35 cycles in Techne TC-512 FTC51F/H2D thermal cyclers (Bibby Scientific Lmtd, Stone, Staffordshire, U.K.). The following programmes were used when applying the Thermo Scientific (Sankt Leon-Rot, Germany) or Sigma RedTaq DNA Polymerases: denaturation at 94°C for 30 s or 1 min, annealing at 60°C for 1 min or 2 min and extension at 72°C for 2 min or 3 min. All DNA was denatured for 5 min at 94°C prior to the start of the programmes and extension was continued for another 5 min at 72°C after cycling had finished.

PCR reactions were run in a total volume of either 20μL or 50μL: 1 x Reaction Buffer (Thermo Scientific or Sigma), 200 μM of dNTPs, 0.2 μM per primer, 25mM $MgCl_2$ (for Thermo Scientific Polymerase only), between 2 ng/μL and 10 ng/μL template DNA and 0.05 U/μL of Red Taq Polymerase (Thermo Scientific and Sigma).

2.9 Agarose gels

Agarose gels were made to a final concentration of 1% (w/v) by dissolving agarose (Lonza, Rockland, Maine, U.S.A.) in 1 x TAE (stock 50 x: 242 g/L Tris base, 57.1 mL glacial acetic acid and 100 mL of 500 mM EDTA). To visualise DNA, ethidium bromide was added to a final concentration of 3×10^{-5} mg/mL. All gel pictures were taken with the Bio-Rad Chemidoc XRS+ using the Quantity One 4.6.6 programme (BioRad, Hemel Hempstead, Hertfordshire, U.K.).

2.10 Primer design

All gene sequences were downloaded from the *T. reesei* database version 2: *http://genome.jgi-psf.org/Trire2/Trire2.home.html*. Primers were designed using the Invitrogen Vector NTI Advance 10 programme (Invitrogen) and ordered from Sigma.

A list of genes studied during this project is given in Appendix A1 Table A.1.T.1.

2.11 Reverse-transcription

One μg of RNA was reverse transcribed to cDNA using the "SuperScript III® First-Strand Reverse Transcriptase" kit and oligo-dT (12-18) primer (all from Invitrogen) following the manufacturer's instructions.

Chapter 3: The effects of different carbon sources on spore development and mycelial growth

3.1 Introduction

3.1.1 Overview of filamentous growth

Understanding growth of filamentous fungi is a key component for improving strains for numerous industrial applications such as the development of bio-control agents for plant and crop diseases (Steyart *et al.*, 2010). The natural unit of propagation for most filamentous fungi are spores which develop, grow and expand over substrates and surfaces by forming mesh works of filaments known as mycelia. Once a spore has landed on a usable surface, it starts to swell and expand symmetrically. Germ tubes will then emerge from the swollen spore and grow into long tubes known as hyphae (see Figure 1.5 in Chapter 1). Apical growth (growth at the tip) of each hypha is the major driving force for growth through substrates (Breakspear *et al.*, 2007) and this is done by continuously locating cell wall components, vesicles and other cytoplasmic compounds to the hyphal tip. This allows extension of the cell wall and hence growth of the hyphae. Branching along the hypha supports further growth and expansion into a colony. On agar plates, fungal colonies grow out radially and in the old centre of the mycelial colony where nutrients are depleted, hyphae grow towards each other and fuse at the tips (Deacon, 2006; Brand and Gow, 2009).

3.1.2 Spore formation, dispersal and development

Most filamentous fungi can produce spores either asexually or through undergoing a sexual cycle (Seidl *et al.*, 2009). As asexually-formed conidia were used throughout this project, only asexual spore production will be discussed.

Spores are characterised as "dormant" structures with a low metabolic rate and a very thick, multi-layer cell wall which is coated with proteins. They also contain pigments which protect them from desiccation, osmotic lysis and UV radiation. The formation of conidia by *Trichoderma* spp. is dictated by several factors or a combination of any of these: UV or blue light, endogenous circadian rhythms, type of carbon (C) and nitrogen (N) source present, C: N ratio, low ambient pH, the presence of extracellular calcium, physical injury to

the hypha and resulting oxidative stress and the presence of volatile organic compounds such as esters, alcohols, ketones or furanes (Steyaert et al., 2010). After one of these environmental signals is detected, conidia formation is triggered. Specialised foot cells of hyphae develop into structures known as conidiophores (Beakspear et al., 2007). In Trichoderma spp., the tips of the conidiophores swell into vesicles which develop into flask-shaped phialides (Deacon, 2006). At the tip of each phialide, conidia are produced which then mature and acquire the green pigmentation typical for T. reesei. Phialides contain one nucleus which is duplicated and re- located into the daughter spore whereas the other nucleus remains within the phialide as more conidia are produced. The colony centre starts by producing conidia and sporulation then rapidly extends radially over the whole colony (Zhong et al., 2008). In this way, conidia are exposed to air and water and can be dispersed by wind and rain splash. Spores land through sedimentation, impact or are simply washed away onto a new area where they can develop into a colony if conditions are right; e.g. by sensing a potential nutrient source (Deacon, 2006). The development of a spore into hyphae consists of 4 major steps: 1) breaking of spore dormancy, 2) isotropic expansion, 3) establishment of cell polarity and germ tube emergence and 4) maintenance of polar growth (Wendland, 2001).

Detecting a suitable carbon source such as glucose is sufficient to increase protein synthesis, break spore dormancy and initiate changes to the thick cell wall. Spores and hyphae both contain GPCRs (G-protein coupled receptors, discussed in detail in Chapter 4) in their cell membranes which either trigger the breaking of spore dormancy by increasing intracellular protein synthesis, cell wall degradation and germination; or in the case of hyphae sustain apical growth and help respond to changes in the environment. Furthermore, cell wall-attached enzymes may cleave off inducing substrates from a complex carbon source which then trigger enzyme secretion required for carbohydrate breakdown and uptake. Internal carbon stores such as trehalose, mannitol, glucose or free amino acids are readily utilised and serve as fuel during the first steps of breaking spore dormancy. After breaking dormancy, the spore expands symmetrically into all directions and the materials and machinery needed for making new cell membrane and cell wall are dispersed throughout the cortex (Momany, 2002). Spores take up water and synthesise glycerol in order to increase turgor pressure and increase spore volume (Wendland, 2001). After this isotropic expansion, a polarised site for germ tube emergence is established and maintained. Several germ tubes can emerge from the enlarged spore and continuously grow into hyphae (Momany, 2002). Septation and branching at new, chosen spots along the hyphal filaments will occur, resulting in a new colony growing over the potential nutrient.

3.1.3 Hyphal structure

Fungal hyphae have both a plasma membrane and a cell wall. The cell wall has many functions: it provides a structural barrier for the organism and acts as an interface between the fungus and its environment. It also protects the fungus from osmotic lysis and degradative enzymes from other organisms and it regulates the secretion of enzymes into the environment (Deacon, 2006). Furthermore, the cell wall acts as a scaffold for the binding of enzymes required for nutrient degradation and uptake.

Fungal cell wall components can be divided into two types: the structural or fibrillar polymers and the matrix components (Deacon, 2006). Fibrillar polymers are straight chain molecules and consist mainly of chitin in ascomycetous fungi such as *T. reesei*. Chitin is composed of chains of β-1,4 linked N-acetylglucosamine residues. Matrix components are branched mannose and α-1,3-glucan polymers with β-1,3-linked backbones and β-1,6-linked side chains. The straight-chain chitin components are mainly located in the inner region of the cell wall and overlaid by the matrix components which cross-link, coat and embed the fibrillar polymers within the cell wall which further strengthens the structure (Deacon, 2006).

The cell membrane in fungi is located closely underneath the cell wall and contains ergosterol as the main sterol as opposed to cholesterol in animals and phytosterol in plants (Deacon, 2006). Transmembrane proteins and integral enzymes are anchored to the plasma membrane, which are required for cell wall remodelling or nutrient uptake. The plasma membrane also relays signal transduction (G-protein signalling) between environmental cues and the internal fungal compartments.

The structure of hyphae varies along the filament and shows a defined polarity in the arrangement of organelles from the apex back towards the sub-apical compartments. Apical or polarised growth takes place at the hyphal tip. Behind the apex, the cell wall thickens and rigidifies along the hyphal filament. Along each hypha, septa or cross-walls are found at regular intervals which structure it into compartments. Septa provide further structural support, enable differentiation and provide protection. When a hypha is damaged, a Woronin body localises to the nearest septum and seals of the rest of the hypha (Deacon, 2006). This protects the fungus from losing the hypha and it can start new growth from behind the damaged compartment. In the case of filamentous ascomycetes, compartments are multi-nucleated.

3.1.4 Apex structure

Apical growth in filamentous fungi takes place at the hyphal tip which differs from the rest of the filament in its structure. The tips of hyphae are characterised by a relatively weak cell wall which is proposed to be elastic and deformable. Hyphal tips can respond to environmental disturbances such as water or heat and it is the place where most enzyme secretion and nutrient uptake takes place. New materials and nutrients are shipped continuously back and forth between the hyphal compartments and the apex.

The hyphal tip consists of a large collection of vesicles, microtubules, actin filaments and polysomes and this structure is also known as the apical vesicle cluster (AVC) or Spitzenkörper. The AVC is thought to organise the delivery of secretory vesicles at the growing tip (Figure 3.1) (Leeder *et al.*, 2008). It can be viewed as a dark body at a growing hyphal tip under the light microscope (Harris *et al.*, 2005). The AVC is required for the directional growth of the hyphae and for hyphal branching as it is only seen at sites of germination and branching; it is also located at a position within the hyphal tip which correlates with the direction of growth (Figure 3.3; Harris *et al.*, 2005). The Spitzenkörper is only present in vegetative, growing hyphal tips. The subapical region behind the Spitzenkörper is rich in mitochondria, providing energy and the proton motive force for nutrient uptake, secretion and cell wall extension (Deacon, 2006).

In order to shift materials to the ever growing hyphal tip, a tubular vacuolar system, involved in the transport of materials, is thought to extend to the apex from older parts of the hyphae. Vacuole tubule dilation and contraction allows inflated elements to travel along them in a peristaltic manner (Deacon, 2006). Furthermore, the cytoskeleton also plays a crucial role in maintaining polarised growth and delivering materials to the apex. Microtubule and actin filaments are found in the periphery of hyphae and extend all the way to the membrane at the apex and sometimes through the AVC. The posterior end of the Spitzenkörper is therefore surrounded by a dense meshwork of microfilaments (Harris *et al.*, 2005). The vacuolar system together with the cytoskeleton components make up the morphogenetic machinery which initiates and maintains polarised growth. Myosin, dynein and kinesin motor proteins are associated with actin and microtubular filaments and they both help to transport vesicles back and forth between the Spitzenkörper and the Golgi of sub-apical compartments (Harris *et al.*, 2004). The general consensus is that microtubules are responsible for long-distance transport of vesicles whereas actin filaments control vesicle organisation within the AVC and

vesicle transport to the plasma membrane at the hyphal tip (Harris *et al.*, 2005).

The vesicles of the AVC can be roughly divided into two populations: the larger apical vesicles and the small microvesicles (Harris *et al.*, 2005). It is generally believed that all these vesicles are required for apical growth and enzyme secretion despite their unknown biochemical nature and composition. Separating and isolating different vesicles and organelles from fungal hyphae for characterisation is difficult and presents a number of challenges such as the presence of fungal proteases and the thick cell wall (Ferreira de Oliveira *et al.*, 2011)

One type of microvesicle which has been characterised is the chitosome. These are small spheroidal bodies of 40 nm to 70 nm thick and contain self-assembling aggregates of chitin synthase which is anchored to the plasma membrane (Deacon, 2006). Chitin synthase is an enzyme involved in the extension of the cell wall by forming chitin polymers and which is activated proteolytically at the apex.

Glucan synthase is another integral membrane enzyme involved in cell wall synthesis of Ascomycota and thought to be located to the apex via vesicles as well. Glucan synthase makes β-1,3-glucan chains from UDP-glucose and utilises GTP as energy for catalysis. Glucan chains will undergo further modifications and cross-linking will occur as the cell wall matures behind the apex. Many other enzymes, not involved in the cell wall extension, are also located to the hyphal tip where they are secreted into the environment.
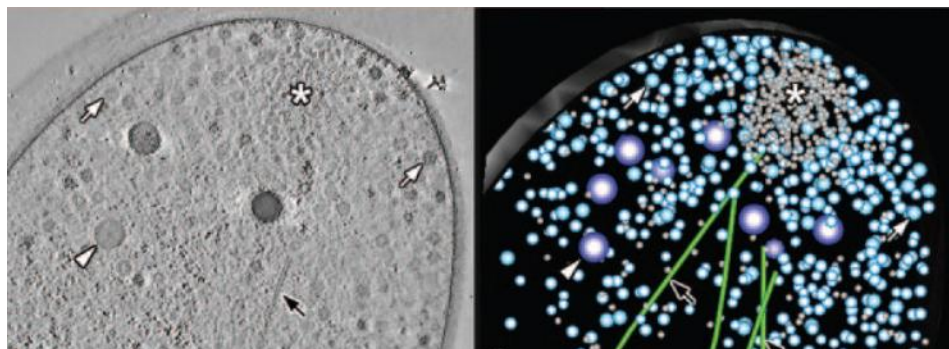
Figure 3.1: TEM of a hyphal tip in *A. nidulans* (left image) with a model of the left panel shown on the right. Microvesicles within the Spitzenkörper (asterisk), apical vesicles (white arrow), Woronin body (white arrow head) and microtubules (black arrow) are indicated in both panels. Taken from Harris *et al*. (2005).

3.1.5 Molecular mechanisms underlying polarised growth

The exact mechanisms for germ tube emergence, maintenance of polarised growth and hyphal branching in filamentous fungi remain largely unknown with only a few genes having been shown to be directly involved. It is clear though that there is a direct link between enzymatic secretion and growth at the hyphal tip and that a high-tip calcium gradient is important for polarity establishment and maintenance (Figure 3.2, Meyer *et al.*, 2008). The organism in which most molecular mechanisms for growth and development have been studied is *Saccharomyces cerevisiae* although major differences exist when compared to filamentous fungi (Momany, 2005).

Breaking dormancy and isotropic expansion in yeast and filamentous fungi is similar and has been characterised extensively (Wendland, 2001). This step will not be discussed in this Chapter due to irrelevancy to this project. Once spores have expanded symmetrically a bud site for germ tube emergence is chosen. In yeast cortical markers select the budding site through defining the axial budding pattern (Bud3p, Bud4p and Axl2p) and the bipolar budding pattern (Bud8p, Bud9p and Rax2p) (Harris *et al.*, 2004). These proteins are deposited to the cell wall throughout cell cycle progression and the previous round of budding (Momany, 2002, Wendland, 2001). These markers are either very poorly conserved or completely absent in filamentous fungi and the mechanisms of bud site selection in these organisms remains unknown (Momany, 2005). It is possible though that germ tube and branch emergence is subject to spatial regulation as has been demonstrated in *Aspergillus nidulans* and *Ashbya gossypii* (Harris and Momany, 2004). Furthermore it is postulated that filamentous fungi may use a unique set of cortical markers or induce receptor signalling and clustering to establish a bud site; even spontaneous polarisation through stochastic signal fluctuation has been proposed (Harris and Momany, 2004). Some part of the mechanisms underlying germ tube emergence and hyphal branching could also be associated to the proteins involved in cell cycle progression during mitosis (Wendland, 2001). In *A. nidulans* isotropic expansion stops and a germ tube emerges after the first nuclear division at 5 h to 6 h after initial spore inoculation (Breakspear *et al.*, 2007). It is also possible that the polarisome itself, the Spitzenkörper and cell end markers such as TeaA and TeaR mark sites for polar growth, as has been proposed in *A. nidulans* (Araujo-Palomares *et al.*, 2009).

Once a germination site has been chosen, polarity information is transduced by Rho GTPases such as Cdc42p to the cells morphogenetic machinery. In yeast, it has been shown that cortical markers recruit Rho GTPases as well as

their corresponding guanine-nucleotide exchange factors (Cdc24p), scaffold proteins (Bem1p) and GTPase activating proteins (GAPs) such as Rga1p, Bem2p and Bem3p to the site of polarised growth (Momany, 2002, 2005). The Rho GTPases are highly conserved between yeast and filamentous fungi and are also the main factors which relay the signalling required for polarised growth; they thus present a crucial link between the sensing of external environmental cues and the corresponding intracellular signalling cascades. *T. reesei* has a homologue of Cdc42p termed MOD1 as well as an additional Rho GTPase RAC1 which is absent in yeast. Cdc42p has various different downstream effectors such as Ste20p and Cla4p, two p21-activated kinases (PAKs) which control the assembly of the morphogenetic machinery. In *A. nidulans*, ModA (orthologue of Cdc42p) also interacts with components of the polarisome such as SpaA.

The organisation of the actin cytoskeleton as well as the Spitzenkörper at polarisation sites is controlled by multi-protein complexes such as the polarisome and the Arp2/3 complex in yeast and filamentous fungi through downstream signalling by PAKs (Harris *et al.*, 2004; Araujo-Palomares *et al.*, 2009). The polarisome and Arp2/3 components are conserved in both yeast and filamentous fungi. In *N. crassa*, polarisome components such as Spa2 are synthesised within the polysomes of the AVC (Araujo-Palomares *et al.*, 2009). The key protein of the polarisome is the formin protein Sep1 (Bni1p in yeast), recruited to the apex by the cell end marker Tea proteins (Brand and Gow, 2009). Sep1 binds to the barbed ends of the actin filaments and nucleates microfilament assembly at the apex and at branching sites (Figure 4, Harris *et al.*, 2004). The remaining polarisome components seem to regulate the timing and location of Sep1: SpaA (Spa2p in yeast) is a scaffold protein which interacts with other polarisome components but also with members of the MAPK cascades and the Rho GTPases (Araujo-Palomares, *et al.*, 2009). In filamentous fungi such as *A. nidulans* and *A. niger*, SpaA is present as a surface cap at the hyphal tip and it is thought to coordinate elongation of germ tubes and hyphae via stabilisation of hyphal polarity axes; whereas Bud6 (Bud6p in yeast), another polarisome component, also binds actin filaments (Meyer *et al.*, 2008).

The Arp2/3 complex on the other hand regulates the assembly of branched actin filaments which form fine networks and an exocyst complex underneath the cell surface; these structures are responsible for vesicle docking and fusion (Brand and Gow, 2009). It is thought that microtubules transport polarisome and Arp2/3 components to the apex and branching sites, hence playing a role in modulating actin dynamics. This has only been shown in *S. pombe* and in no filamentous fungus (Harris *et al.*, 2004). The relationship

between the polarisome and the Spitzenkörper has not been fully clarified yet. One protein which could be involved in this is *A. nidulans* BemA (Bem1p in yeast), a scaffold protein for ModA. BemA localises to a cap at the hyphal tips and has been shown to organise polarisome components such as SepA and also ensures proper vacuole fusion and protein secretion (Leeder *et al.*, 2008). Septins are proteins involved in septa formation in the hyphal filament and have also been shown to localise to sites of branching, germ tube emergence and conidiophore formation in *A. nidulans* and are thought to play a role in polarity establishment and growth (Harris *et al.*, 2004).

It is clear that mutations affecting vesicle transport, microfilament-based motor proteins, chitin deposition and components of the signalling cascade severely disturb the establishment and maintenance of polarised growth (Momany, 2002). Apical growth requires the expression of many genes (Breakspear *et al.*, 2007); as many as 45 different genes have been proved to be important for morphogenesis in *N. crassa* and *A. nidulans* (Harris *et al*., 2005). Furthermore, mutations in the two genes *swoF* and *swoA*, encoding an N-myristoyl transferase and a protein mannosyl transferase in *A. nidulans* were shown to be important for filamentous growth (Momany, 2005). These two enzymes add post-translational modifications to their target proteins, ensuring their proper localisation and function at the hyphal tip. Another candidate is AurA, an inositol phosphorylceramide (IPC) synthase involved in sphingolipid synthesis of eukaryotic membranes and which when inhibited, abolishes germ tube emergence (Momany, 2002). Many of the genes involved in polarity establishment and maintenance, encode proteins with novel functions (Breakspear *et al.*, 2007). Zhong *et al.* (2008) showed that by disrupting *trccd1* in *T. reesei*, a gene encoding a carotenoid cleavage dioxygenase, they not only impaired apacarotenoid synthesis but also found defects in morphogenesis and conidia formation hence linking secondary metabolite synthesis to polarised growth.

The understanding of the exact mechanisms underlying fungal growth is far from being complete; it is clear though that many pathways are involved either directly or indirectly in hyphal morphogenesis. The assembly/disassembly and composition of the Spitzenkörper remain largely unknown as do the steps connecting it to the morphogenetic machinery and the polarisome. So far none of this has been investigated thoroughly in any *Trichoderma* spp..
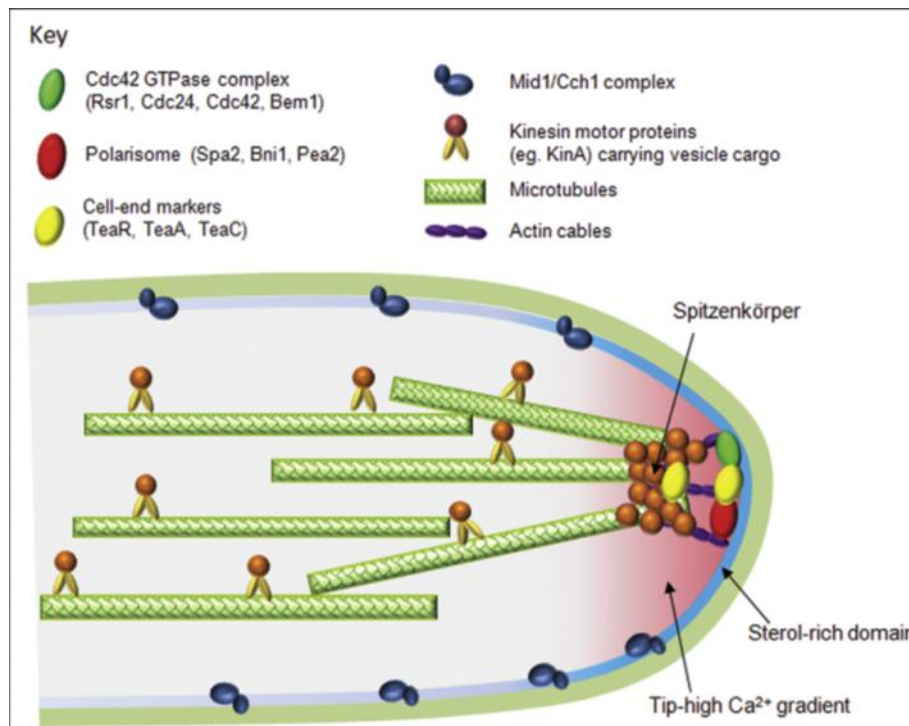
Figure 3.2: Diagrammatic representation of a hyphal tip of *A. nidulans* with components required for polarised growth. Taken from Brand and Gow (2009).
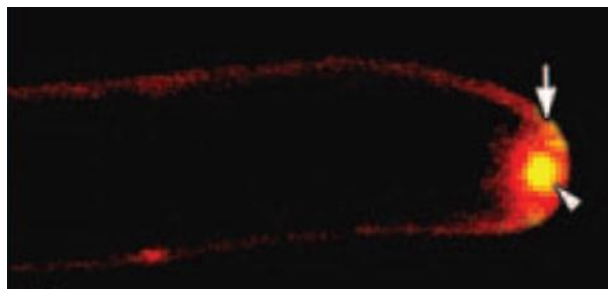
Figure 3.3: Confocal image showing a hyphal tip of *A. nidulans* stained with FM 4-64 (red) and SepA-GFP (yellow). The red dye delineates the cell wall and the vesicles found within the Spitzenkörper whereas the yellow pigment indicates the localisation of SepA. Taken from Harris *et al*. (2005).

3.1.6 Protein secretion

A second, important characteristic of filamentous fungi is that they have a very prominent secretion system which continuously ships material to the growing tip. Heterologous protein expression in *T. reesei* as well as high secretion of native proteins in naturally occurring and industrial strains requires a highly functioning and specialised secretory apparatus (Arvas *et al.*, 2006).

Proteins destined for secretion are synthesised from mRNA by ribosomes and contain an ER signal sequence which targets them to the ER. Proteins with an ER signal sequence are imported into the ER lumen either co-translationally (translation occurs simultaneously with import into the ER through ER associated ribosomes) or post-translationally (translation occurs first before the unfolded protein is imported into the ER). The ER is composed of flattened sacs and branching tubules which are all interconnected and form the ER lumen (Alberts *et al.*, 2002). Transport into the ER lumen happens via a translocator complex, termed the SEC61 complex. The SEC61 complex, a heterotrimeric doughnut shaped pore comprised of α, β and γ subunits, further associates with BIP1 and other SEC proteins (SEC63, SEC62, SEC72 and SEC66). This multi-protein complex then pulls the peptide through the ER membrane into the lumen where secretory proteins are folded and glycosylated. In filamentous fungi newly synthesised polypeptide chains are not only folded correctly within the ER but are also glycosylated at the same time by the N- or O-glycosylation of asparagine or serine and threonine residues. Sugar nucleotides are transported into the ER from the nucleus where enzymes such as glucosidases I and II as well as several ER-mannosidases add and cleave glycosyl residues (Geysens *et al.*, 2009). These enzymes as well as the ER chaperones and oxidoreductases also assist in the quality control of folded and glycosylated proteins (Arvas *et al.*, 2006). Once a protein has undergone all the checkpoints within the ER, it buds off in a vesicle and is transported to the Golgi apparatus where further glycosylation events, catalysed by other α-mannosidases, can take place before the protein is packed into Golgi vesicles and targeted for secretion at the apex (Geysens *et al.*, 2009). The Golgi of Ascomycota are sausage shaped with strings, beads and loops and hence differ structurally from animal and plant Golgi (Deacon, 2006).

Endocytosis is the process by which eukaryotic cells internalise and recycle plasma membrane lipids, membrane-associated proteins and take up nutrients. During endocytosis, molecules are ingested in vesicles derived from the plasma membrane and delivered to early endosomes and then to late

endosomes. Endocytosed molecules can be retrieved from early endosomes and returned to the plasma membrane for reuse; similarly some molecules can be returned from the late endosome to the Golgi apparatus and even back to the ER (Alberts *et al.*, 2002). Alternatively, molecules can be delivered from late endosomes to lysosomes where they are degraded. In *A. nidulans*, early endosomes show bi-directional long-distance movement along microtubules, using kinesin motors to move towards the tip and dynein-dependent transport to move away from the apex (Peñalva, 2010). It is thought that this bi-directional movement facilitates recycling of cargoes from endosomes to the plasma membrane. Vesicle transport during exocytosis and endocytosis is dependent on the morphogenetic machinery (actin patches, filaments and microtubules; see 3.1.5). Endocytosis has been shown to be important for polarised growth in filamentous fungi, as it occurs predominantly at the hyphal tip and is also spatially coupled to exocytosis. This was supported by the internalisation of the lipophilic dye FM4-64 by endocytic membranes and which also stained the Spitzenkörper (Peñalva, 2010). In *A. oryzae* suppression of *end4*, a gene which encodes a protein functioning as an adaptor that connects the invaginated plasma membrane and actin cytoskeleton, resulted in retarded apical growth, abnormal hyphal morphology, mis-localisation of a vesicle-SNARE (**S**oluble **N**-ethylmaleimide-sensitive factor **A**ttachement protein **RE**ceptor, membrane proteins facilitating membrane fusion during exocytosis) and aberrant accumulation of cell wall components at large invaginated structures (Higuchi *et al.*, 2009). Several other mutations of genes encoding proteins involved in endocytosis have been shown to prevent apical extension and can even lead to cell death: in *A. nidulans*, *slaB* (encodes a transmembrane actin-binding protein) knock-out is lethal whereas null mutants of *arfB* (encoding a GTPase regulating endocytic internalisation) displayed defects in maintaining polarised growth and disruption of *fimA* (encoding fimbrin, an actin-bundling protein) resulted in abnormally large and swollen conidia (Peñalva, 2010).

High throughput protein secretion induces stress on the secretory system. This stress activates a homeostatic response known as the Unfolded Protein Response (UPR). The UPR has been associated to several cellular processes such as the regulation of the ER lumenal environment, ER-associated protein degradation (ERAD), autophagy, apoptosis, nitrogen-sensing and meiosis (Geysens *et al.*, 2009). The UPR is initiated when the cell detects unfolded proteins within the ER. Hence, the ER not only serves as an environment for correct protein folding and glycosylation but it also houses the detection system to recognise problems in protein folding which could potentially be lethal to the organism. Unfolded proteins are sensed by a transmembrane riboendonuclease, IRE1, which oligomerises and is subsequently activated by

autophosphorylation. The substrate for IRE1 is *hac1* mRNA from which one intron is removed and the rest of the RNA is joined together by a tRNA ligase (Geysens *et al.*, 2009). In *T. reesei*, IRE1 recognises the loop-bulge structure formed by the 20 nucleotide intron and thus removes it from the *hac1* mRNA. This is thought to change the reading frame and bring an activation domain into the protein (Saloheimo *et al.*, 2003). Furthermore, a 250 bp fragment of the 5' flanking region of *hac1* mRNA is removed. The truncation of *hac1* mRNA removes an upstream ORF encoding 18 amino acids and it is thought that this produces a new transcription start site (Saloheimo *et al.*, 2003). The activated *hac1* mRNA is then translated to the HAC1 transcription factor that may bind to UPR elements (UPREs) within the promoter regions of its target genes such as *pdi1* and *bip1*, encoding other resident ER proteins.

The UPR affects transcription of a few hundred genes involved in processes such as protein translocation, protein glycosylation, lipid and inositol metabolism and ER homeostasis (Saloheimo *et al.*, 2003). In *T. reesei* down-regulation of genes encoding proteins destined for extracellular secretion (REpression under Secretion Stress, RESS) is also observed but it is unknown whether this response is directly linked to the UPR (Arvas *et al.*, 2006). Transcriptional up-regulation of genes encoding foldases, chaperones and members of the ERAD system is observed upon UPR activation. The UPR further increases levels of BIP1, a chaperone belonging to family HSP70 of chaperones. BIP1 binds to exposed, hydrophobic regions of the unfolded proteins through ATP hydrolysis. This binding holds the proteins in an aggregated state which provides a helpful folding environment (Geysens *et al.*, 2009). Furthermore, BIP1 also binds to the lumenal domain of IRE1 and releases IRE1 by binding to unfolded proteins which causes the aggregation and subsequent activation of IRE1. The UPR also increases levels of a protein disulphide isomerase, PDI1, which can catalyse the formation, reduction and isomerisation of disulphide bonds within proteins (Geysens *et al.*, 2009). PDI1 also possesses minor chaperone activity. PDI1 is made out of 5 domains (a, a', b, b' and c) of which a and a' possess a CGHC catalytic motif which can be oxidised by another protein named ERO1. Oxidised PDI1 is asymmetric in structure and this is responsible for the separate oxidase and isomerase activity of this protein (Geysens *et al.*, 2009). It is the pre-dominant disulphide isomerase within the ER, although others also exist.

3.1.7 Phases of hyphal growth

Growth of filamentous fungi can be divided into four different phases: 1) the lag phase, 2) the exponential growth or logarithmic phase, 3) the stationary phase and 4) the death phase or phase of autolysis (Deacon, 2006). During the lag phase when for example spores are transferred into a new medium, they adapt and sense a potential carbon source, germinate and grow into mycelia. During the logarithmic phase, growth conditions are optimal for the mycelia and the maximum specific growth rate, μ_{max}, can be obtained. In this case, filamentous fungi can have a duplication cycle; this means that averaged for a colony as a whole, they grow as hypothetical "units" where one produces two and two produce four (Deacon, 2006). The time it takes for the fungal biomass to double (termed doubling time) can therefore be deduced. After a while, the fungus reaches the stationary phase which is characterised by limiting food resources and no further expansion of the colony. If the organism is not provided with fresh nutrients, it will start to die and enter the phase of autolysis.

3.1.8 Aims

The general aim of this Chapter is to characterise fungal growth by:

- Investigating the influence of different carbon sources on spore development and subsequent hyphal extension
- Assessing whether genes which have been shown to be important for spore development in *Aspergillus* spp. and budding in yeast, are also expressed during germination in *T. reesei* and if their expression changes in the presence of different carbon sources
- Characterising *T. reesei* hyphal growth in the presence of different liquid and solid carbon sources with regards to what is known about the regulation of genes encoding plant cell wall-degrading enzymes in response to these simple and complex carbons sources

3.2 Materials and Methods

3.2.1 Spore preparation for microscopy

Pre-adapted conidia were obtained from PDA, PXA, PCA and PGA slopes (Chapter 2 section 2.3) and inoculated at a concentration of 10^8 spores/mL in 100 mL TMM (Chapter 2 section 2.4) complemented with either glucose, xylose, cellobiose or glycerol for 10 h at 150 rpm at 28°C. A 1 mL sample was taken every 2 h and spores were sedimented at 15,000xg for 2 min and washed in 1 mL 0.01 % (w/v) Tween 80. This was repeated twice and after the last wash spore pellets were re-suspended in 100 μL of 0.01 % (w/v) Tween 80.

5 μL of the spore suspension was analysed under the HBO50 microscope (Zeiss, Welwyn Garden, Hertfordshire, U.K.) with the 40 x objective magnification lens. For each set of conditions, pictures were taken of 100 spores and the spore area size was measured using the Axiovision 3.0 programme (Zeiss).

3.2.2 Observation of changes in spore population size using flow-cytometry

Using the same spore suspension as described in 3.2.1, spores were sedimented at 15,000xg for 2 min. They were re-suspended in 0.5 mL Tween 80 (0.01% v/v) and transferred to 5 mL round bottom BD falcon tubes (BD Biosciences, Bedford, Massachusetts,U.S.A.). The BD FACSCanto™ Flow Cytometer (Ref 337175, BD Biosciences, Oxford, Oxfordshire, U.K.) with the BD FACSDiva™ Software (BD Biosciences) was used to analyse the size of the spore population over 6 h and results were exported as PDF files.

3.2.3 RNA extraction from spores

The same inoculation details and method were used as described in 3.2.1. Spores from two 100 mL flasks were combined and sedimented at 15,000xg for 2 min. Pellets were re-suspended in 1 mL RNA extraction buffer (0.6 M NaCl, 0.2 M Sodium acetate, 100 mM EDTA, 4% SDS, heated at 65°C) and snap frozen in liquid N_2. The frozen material was added to a container (Sarstedt, Leicester, Leicestershire, England) together with 0.5 mL of 0.15 mm glass

beads (BDH Chemicals Limited, Poole, England) and dismembrated with the Mikro-Dismembrator U (Sartorius, Göttingen, Germany) at 2000 rpm for 4 min. Disrupted material was removed from the container and immediately added to 1 mL TriZol (Invitrogen) and mixed by inverting the tube. Glass beads were spun down and the liquid layer was removed into a new 2.0 mL Eppendorf tube. This was mixed for 30 s by rigorous shaking with 0.25 mL chloroform (Fisher Scientific) and incubated at room temperature for 5 min. The rest of this protocol is exactly the same as described in section 2.7 of "General materials and methods".

3.2.4 Real-time PCR

Twelve genes were chosen to be investigated by qRT-PCR during the first 10 h after inoculation of spores in TMM. PCR protocols and plate set up were compiled using the "7500 Fast Real-Time PCR system and software v.2.0.4 and v.2.0.6" by Applied Biosystems (Carlsbad, California, U.S.A.). Each gene and condition was run in 20 μL triplicates in a Fast Optical 96-well reaction plate with barcode (Applied Biosystems). Each 20 μL reaction contained 11.0 μl Fast SYBR green master mix (Applied Biosystems), 0.11 μL/primer (see Table 3.1 for primer details), 2.2 μL of sample DNA or standard curve DNA and 8.58 μL ddH$_2$O. Sample cDNA was prepared from RNA as described in section 2.11 of Chapter 2 and diluted to 50 ng/μL. Gene expression values were calculated by the software as a reference to the standard curve. For each condition, a template-free control was included.

Standard curve DNA was generated by amplifying each gene by PCR (Chapter 2 section 2.10) using the same primers as given in Table 3.1 and purified with the "Gel extraction and PCR purification" kit by Geneflow (Staffordshire, U.K.) according to the manufacturer's instructions. Purified genes were quantified by nanodrop spectrophotometry (see section 2.6 Chapter 2) and diluted to 0.1 ng/μL and then by a 1:10 serial dilution to generate 5 reference points for the standard curve.

PCR was run for 40 cycles with denaturation at 95°C for 30 s and annealing at 64°C for 30 s. DNA was denatured prior to cycle 1 at 95°C for 20 s and a melting temperature stage was also included at 95° for 15 s and 64°C for 1 min after the amplification stages.

Results were exported as an excel file and expression values converted from ng/μL to pg/g of total cDNA.

Table 3.1: Primer sequences, annealing temperatures and calculated product sizes of the 12 genes for QRT-PCR during spore development are shown

Name	Forward Sequence (5' to 3')	Annealing temperature (°C)	Reverse sequence (5' to 3')	Annealing temperature (°C)	Predicted product size (bp)	Predicted product size (bp) no intron
bud4	tggctggaattg gctcgcaa	59.7	cgccagggcttc cttctgct	60.3	264	264
axl2	acagacgatgg ggatcacggc	60.5	ctcggactgaat ggctcgtgga	60.9	350	350
rax2	tcatggcgaag gcttcgacg	60.6	tggcttcggcga ctaggct	60.7	428	428
rac1	ttgatggcaagc cgatcagcc	60.5	tgcgctgagtca gggcgg	60.4	446	373
mod1	cgttgtcgtcgg cgacgg	60.6	gagtcaacgcg ctgcattcga	60.8	737	467
sep1	gcacggaggga tttcttcacca	59.8	ctgcccagaggc tacacgaacc	59.6	306	306
spa1	acgagtccgcac cgcgtg	60.2	ccgtcgtagtgt tcctcagcg	60.6	469	469
bud1	ccagtcggagat ccagtcgtcc	59.4	ggtggcctgctc gacgagagt	60.2	471	471
pdi1	tccctctggctta catcttcgcc	60.4	tggcatcaacct tggcgatga	60.8	519	519
bip1	cgaggttgacg acatcgttctgg	60.3	gggctggttatc ggcagcagt	59.9	324	324
hac1	tcggtggagatg ccgctgtc	60.5	ggctgctggata gaatgtcgcg	60.5	406	406
sec63	gattgcccggac actgaaccag	60.2	ccattagcgcat ccagatcgtctt	60.3	497	497

3.2.5 Agar plates

TMM (see Chapter 2 section 2.4) containing 7.5 g/L of agar was complemented with 0.5% xylan, 1% cordiale wheat straw, 1% cellulose or 1% glucose. TMM agar media without any carbon source was used as a control. Media was poured into 90 mm, single vent plates (Bibby Sterilin, Stone, Staffordshire, England) and left to set. Plates were inoculated with 10 µL of spore suspension (section 2.3) and left to grow at 28°C. Plate growth was scanned (hpscanjet 64010) at 48 h, 72 h, 96 h and after one week.

3.2.6 Growth curves

To generate growth curves, 100 mL TCM or TMM complemented with the relevant carbon source was inoculated with 10^5 spores / mL (Chapter 2 sections 2.3 and 2.4) at 28°C at 150 rpm. Mycelia were harvested (Chapter 2 section 2.5) at the following time points: 24 h, 36 h, 48 h, 60 h, 72 h and 96 h. They were freeze-dried (SB4 model, HemLab England, U.K.) overnight and biomass was weighed out the following day. Each time point and condition was repeated in triplicate.

The average weight of biomass from the three replicates for each time point was plotted onto semi-log graph paper and a linear trend line added (time versus weight). The doubling time (dt) is deduced from the graph by deducing the time it takes for the fungal biomass to double in mass (e.g. from 1 g to 2 g). Maximal specific growth rates (μ_m) were then calculated using the following equation: $\mu_m = \ln2/dt$.

3.3 Results

3.3.1 Microscopic analysis of spore development in different carbon sources

Microscopic analysis of spores showed that spores developed and germinated the first 10 hours after inoculation in minimal medium and that the timing of development differed between carbon sources (Appendix A0, Figure A.0.F.1). Changes in spore size occurred earlier in glucose (Figure 3.3.1) and cellobiose than in xylose and glycerol (Appendix A0 Figure A.0.F.1). In glucose and cellobiose, germ tubes began to emerge between 4 h and 6 h after initial inoculation whereas in xylose and glycerol germ tube emergence started between 8 h and 10 h (Appendix A0 Figure A.0.F.1). These results were also supported when measuring spore area during the first 6 hours after initial inoculation: spores increased in size in the presence of glucose and cellobiose but not in the presence of the other two carbon sources (Figure 3.3.2). Spores at time point 0 were around 20 μm^2 in area; only those grown on PCA slopes were smaller by about 2 μm^2.

0 h   2 h   4 h

6 h   8 h
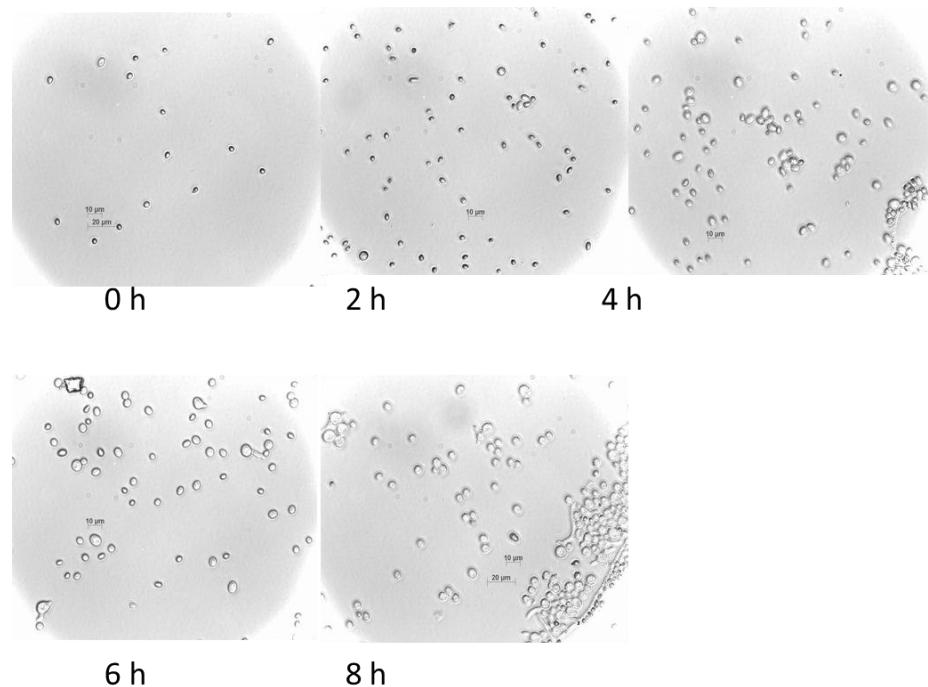
Figure 3.3.1: Microscopic analysis of spore development in minimal media complemented with glucose. Pictures were taken at 2 h time intervals; scale bars (10 μm and 20 μm) indicate size.
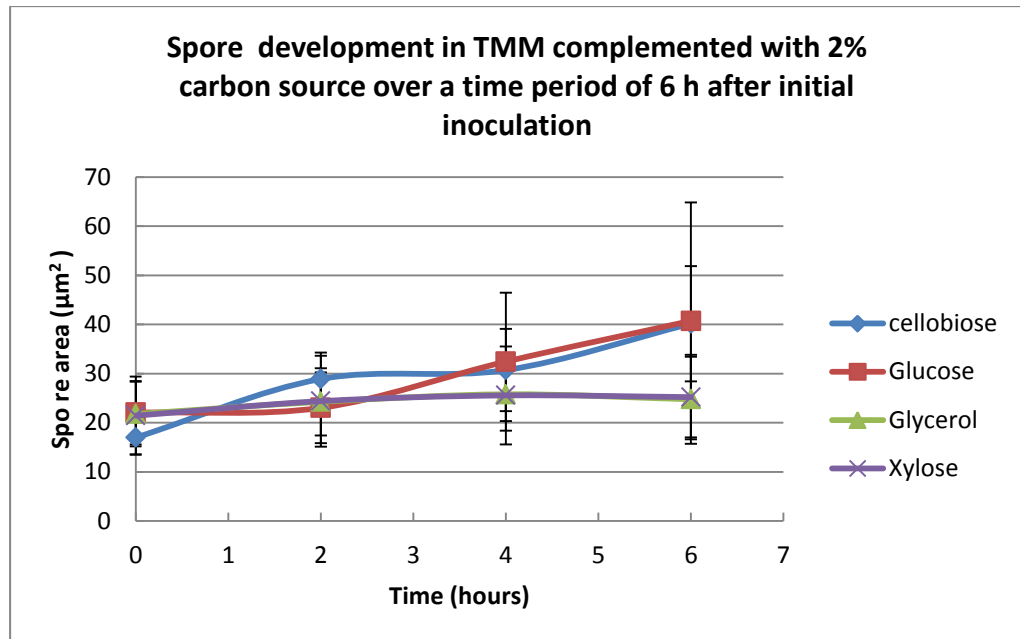
Figure 3.3.2: Recorded area size of 200 spores for each time point and condition. Spores were inoculated in minimal media with different carbon sources over a time period of 6 h. Error bars indicate standard deviation for the 200 measurements.

3.3.2 Flow-cytometry analysis of spore population size

Flow-cytometry analysis showed that spore population size changed during the first 8 h after inoculation in TMM (Figures 3.3.3 and 3.3.4). Again, the changes in population size were different between carbon sources and were more rapid in glucose and cellobiose than in glycerol and xylose. Flow cytometer analysis showed that changes in the spore population size started after 6 h of incubation in glycerol and xylose-based media and at 2 h in the presence of glucose and xylose (Figure 3.3.4).

A large proportion of spores (about 20% to 30% of the total spore population) inoculated in glucose and cellobiose did not manifest any changes in size. This became clear as two distinct sizes in population appeared during flow-cytometry analysis after 8 h (Figure 3.3.3). This was also the case for growth in xylose and glycerol (as supported by microscopic observations) but happened at a much later time point.
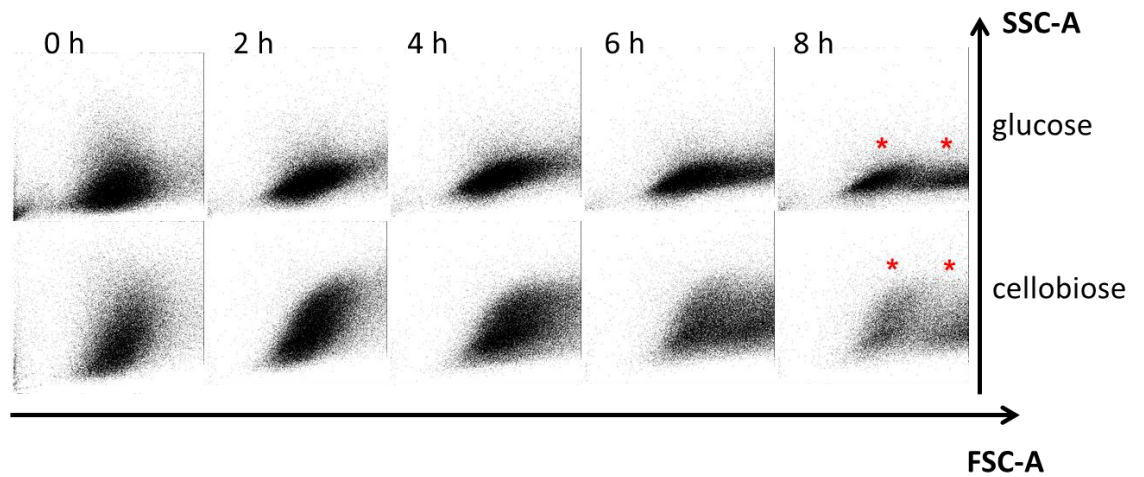
Figure 3.3.3: Recorded population size by flow-cytometry of *T. reesei* spores grown for 8 h in minimal media complemented with different carbon sources (FSC-A = forward scatter-A gives population size of 100,000 spores; SSC-A = side scatter-A gives the roughness of the surface of 100,000 spores; * = indicates two distinct spore population sizes).
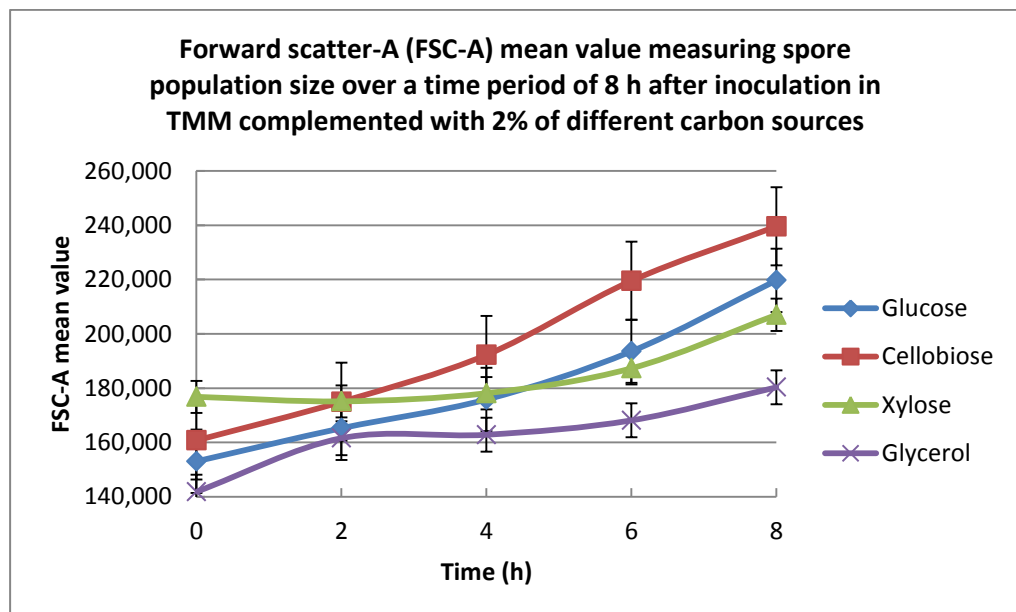
Figure 3.3.4: Mean spore population size (FSC-A) of 100,000 *T. reesei* spores in different carbon sources over a time period of 6 h. Standard error for the value of the total population size at each time point is also shown.

3.3.3 Expression of genes involved in spore development

Several genes were chosen to assess their expression and potential role in spore development in *T. reesei*. These genes were 3 potential cortical markers (*bud4*, *axl2*, *rax2*), 2 GTPases (*rac1*, *mod1*), 3 polarisome components (*sep1*, *spa1*, *bud1*) and 4 genes encoding components of the UPR (*hac1*, *pdi1*, *bip1*, *sec63*). Gene comparisons between yeast and filamentous fungi have generally focused on *Aspergillus spp*. and *S. cerevisiae* and many proteins of *Aspergillus* have homologues in yeast. That is why proteins and their corresponding genes from *A. nidulans* were searched for in *T. reesei* (Table 3.2). There is high sequence identity between GTPases, polarisome and UPR proteins of *A. nidulans* and *T. reesei*, the only exception for this being HAC1. On the other hand, there is low similarity between the cortical markers, Bud4p, Axl2p and Rax2p of *S. cerevisiae* compared to the corresponding proteins in the filamentous fungi. Expression of all 12 genes was assessed through qRT-PCR in all conditions and time points.

Table 3.2: Names and sequence identities of homologous proteins found in *S. cerevisiae* (http://www.yeastgenome.org) and *A. nidulans* (http://www.aspergillusgenome.org) for genes investigated during spore development in *T.reesei*. A description and I.D. of the corresponding gene in *T. reesei* is also shown (http://genome.jgi.doe.gov/Trire2/Trire2.home.html).

T. reesei protein Name	*S. cervisiae* protein name	*A. nidulans* protein name	Sequence identity of *A. nidulans* to *S. cerevisiae* entire protein	Transcript I.D. *T. reesei*	Description in *T. reesei*	Sequence identity of *T. reesei* to *A. nidulans* entire protein
BUD4	Bud4p	An6150	22%	76029	Hypothetical protein	37%
AXL2	Axl2p	An1359	32.3%	110910	Transmembrane glycoprotein	30%
RAX2	Rax2p	An6658	23.2%	76515	DNA binding (HMG1/2 box), protein transporter involved in protein secretion, amine oxidase	60%
RAC1	Cdc42p	RacA An4743	66.1%	47055	Ras GTPase	77%
MOD1	Cdc42p	ModA An7487	82.6%	50335	Cdc42, Rho GTPase	94%
SEP1	Bni1p	SepA An6523	33%	77031	Actin binding FH2, cytokinesis protein	55%
SPA1	Spa2p	SpaA An3815	46.8%	108829	Class II aminotransferase protein	36%
BUD1	Bud6p	BudA An1324	31%	35386	Hypothetical protein, actin binding in *S. cerevisiae*	48%
PDI1	Pdi1p	PdiA An7436	38.6%	122415	Protein disulphide isomerase	65%
BIP1	Kar2p	BipA An2062	69.3%	122920	HSP70 family ER chaperone	78%
HAC1	Hac1p	HacA An9397	14%	46902	bZIP transcription factor	43.5%
SEC63	Sec63p	An0834	44.3%	121754	ER protein translocation complex subunit	58%

3.3.3.1 Transcript levels of genes encoding potential cortical markers

Out of the three genes studied, the level of *bud4* mRNA was about twice as high as the other two potential cortical markers (Figure 3.3.7). The levels of *axl2* and *rax2* mRNA, were not very high in all conditions studied. In cellobiose and xylose the level of mRNA from all 3 genes seemed to stagnate during the first 4 hours after inoculation in TMM whereas in glycerol there was a low level increase during that same time. At 6 h, there was a big increase in transcript levels of *bud4* and *rax2* whereas *axl2* levels remained low. In glucose mRNA from all 3 genes peaked at 4 h with a major decrease in levels at 6 h (Figure 3.3.5).
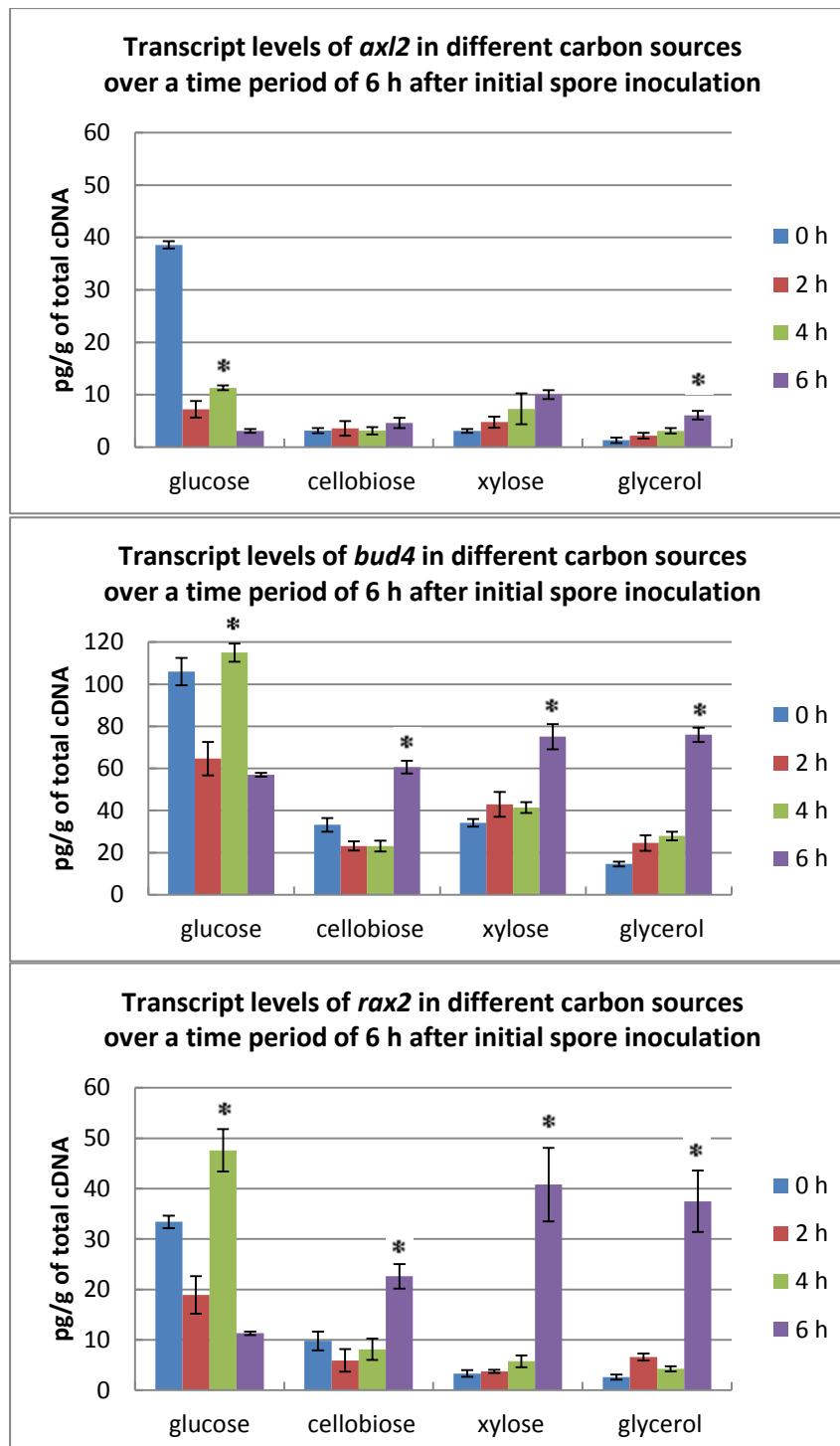
Figure 3.3.5: Levels of mRNA from *bud4*, *axl2* and *rax2* in *T. reesei* grown from conidia in different carbon sources over a time period of 6 h after initial inoculation. Error bars are the standard deviation for the 3 qRT-PCR replicates. * indicates significant difference (a p-value of <0.01 in an equal variance, one-tailed T test) between 4 h glucose and 2 h, 6 h glucose expression levels and between 6 h and 4 h cellobiose, xylose or glycerol expression levels.

3.3.3.2 Transcript levels of genes encoding GTPases required for signal transduction

Levels of mRNA from *rac1* and *mod1* were generally 10- to 100- fold higher than for the genes encoding potential cortical markers. Out of the two GTPases, *mod1* transcript levels were the highest. Levels of *mod1* and *rac1* transcripts increased largely at 6 h in cellobiose, xylose and glycerol, whereas in glucose, levels were highest at 4 h with a decrease in transcript levels at 6 h (Figure 3.3.6).
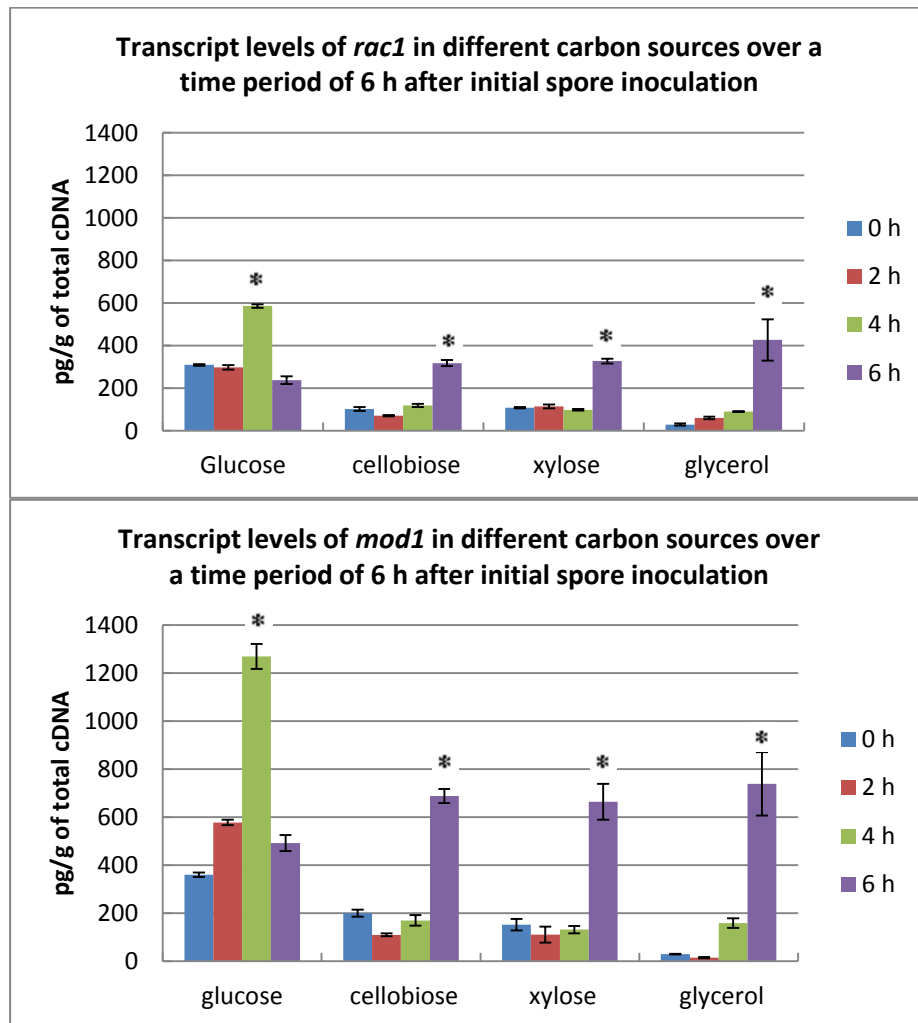
Figure 3.3.6: Levels of mRNA from *rac1* and *mod1* in *T. reesei* grown from conidia in different carbon sources over a time period of 6 h after initial inoculation. Error bars are the standard deviation for the 3 qRT-PCR replicates. * indicates significant difference (a p-value of <0.01 in an equal variance, one-tailed T test) between 4 h glucose and 2 h, 6 h glucose expression levels and between 6 h and 4 h cellobiose, xylose or glycerol expression levels.

3.3.3.3 Transcript levels of genes encoding potential polarisome components

Transcript levels of *sep1*, *spa1* and *bud1* were higher than those for genes encoding potential cortical markers but not as high as for genes encoding GTPases. Out of the three genes studied, *sep1* transcript levels were the highest, followed by transcript levels of *spa1* whereas *bud1* barely seemed to be transcribed in any of the conditions studied (Figure 3.3.7). All gene transcript levels were highest at 6 h in cellobiose, xylose and glycerol and at 4 h in glucose. A decrease in transcript levels at 6 h in glucose was observed whereas mRNA levels in cellobiose and xylose stagnated over the first 4 h after inoculation before increasing at 6 h. In glycerol the same slow rise in transcript levels was observed again over the first 4 h after spore inoculation in TMM.
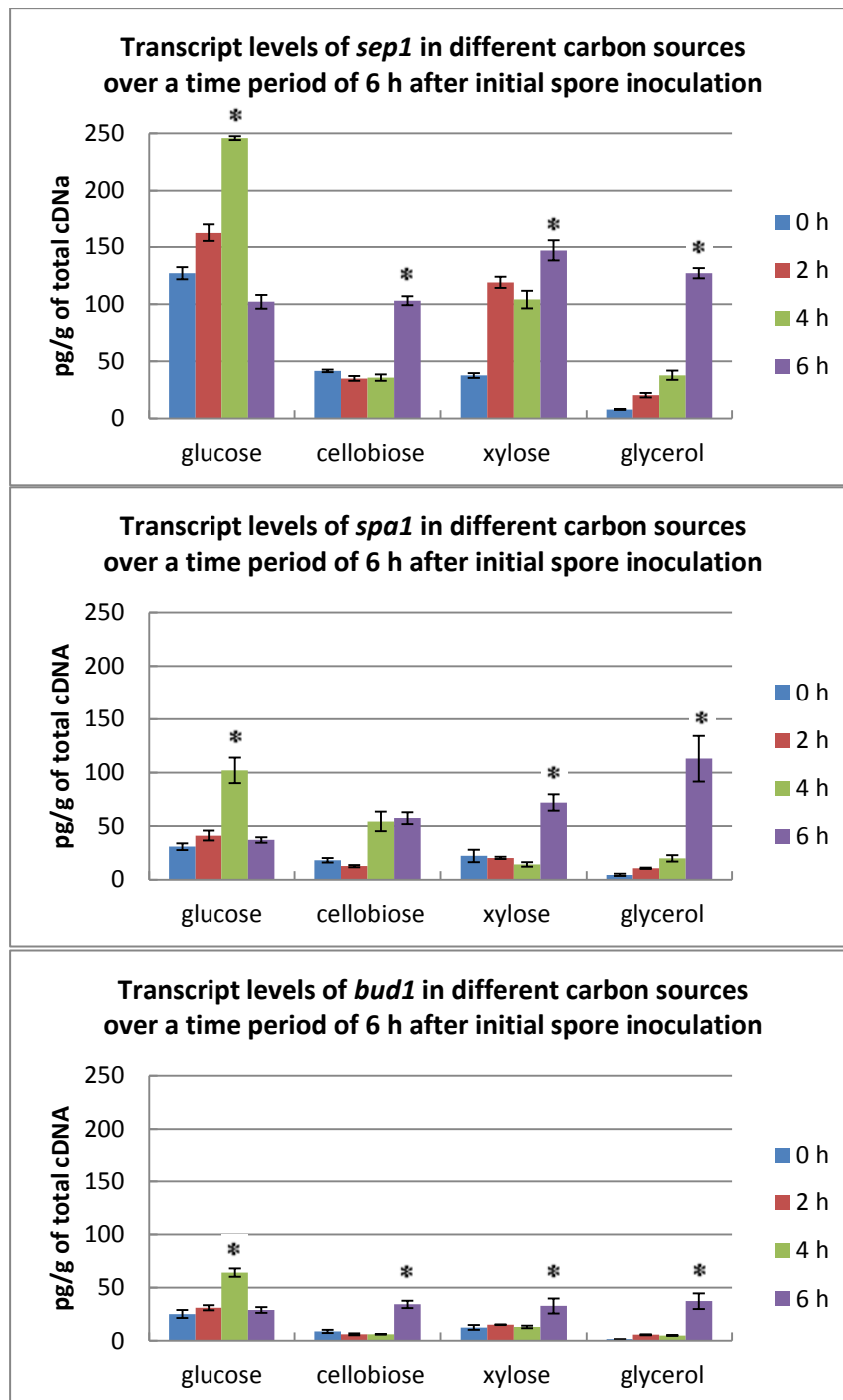
Figure 3.3.7: Levels of mRNA from *sep1*, *spa1* and *bud1* in *T. reesei* grown from conidia in different carbon sources over a time period of 6 h after initial inoculation. Error bars represent the standard deviation for the 3 qRT-PCR replicates. * indicates significant difference (a p-value of <0.01 in an equal variance, one-tailed T test) between 4 h glucose and 2 h, 6 h glucose expression levels and between 6 h and 4 h cellobiose, xylose or glycerol expression levels.

3.3.3.4 Transcript levels of genes encoding components of the UPR

Transcript levels from these genes were relatively high compared to gene categories studied during spore development. Transcript levels of *pdi1* (Figure 3.3.8) were very high, followed by levels of *bip1* and then *hac1* (Figure 3.3.9). Transcript levels from *sec63* were not very high throughout all conditions studied (compare all figures to Figure 3.3.8 and 3.3.9). As for the transcript levels of the other gene categories, they increased continuously over the first 6 h in xylose and glycerol with the highest transcript levels observed at 6 h. Transcript levels stagnated in cellobiose and then increased markedly at 6 h. In glucose gene mRNA levels were again highest at 4 h; *hac1* and *pdi1* transcript levels increased up to that time point whereas mRNA levels for the other two genes seemed to be relatively constant. At 6 h, transcript levels from all genes decreased again.
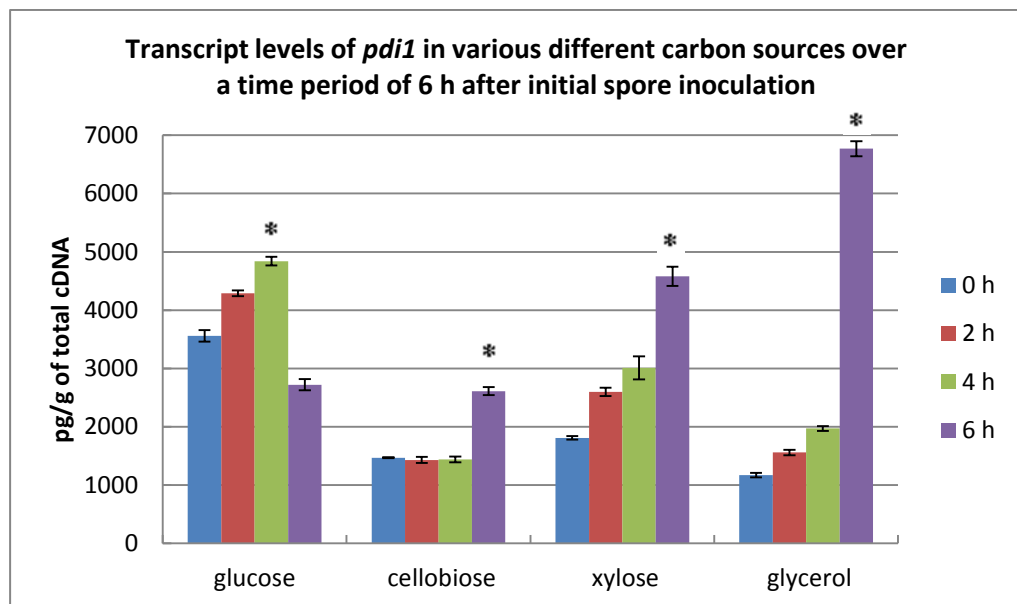
Figure 3.3.8: Levels of mRNA from *pdi1* in *T. reesei* grown from conidia in different carbon sources over a time period of 6 h after initial inoculation. Error bars represent the standard deviation for the 3 qRT-PCR replicates. * indicates significant difference (a p-value of <0.01 in an equal variance, one-tailed T test) between 4 h glucose and 2 h, 6 h glucose expression levels and between 6 h and 4 h cellobiose, xylose or glycerol expression levels.
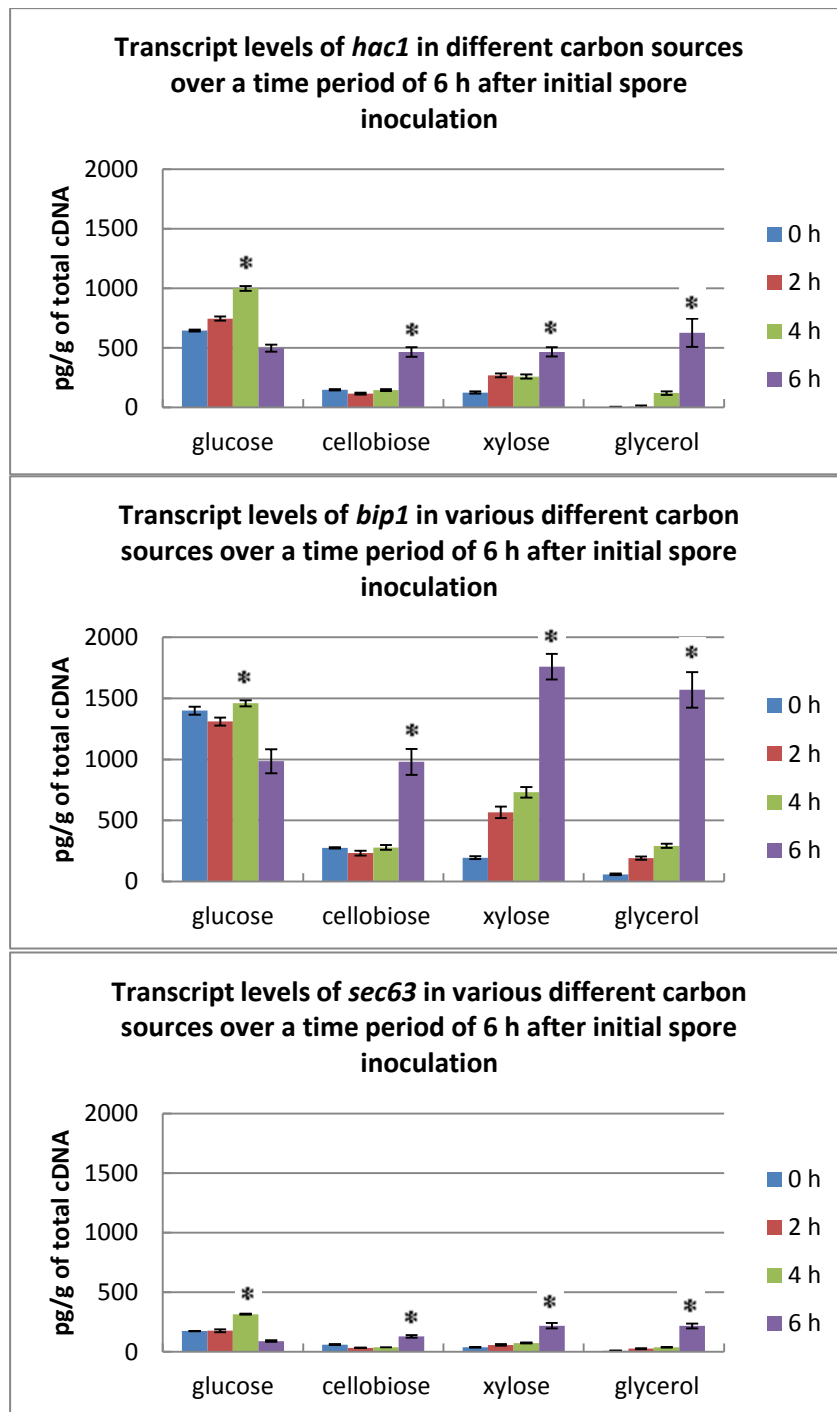
Figure 3.3.9: Levels of mRNA from *bip1*, *hac1* and *sec63* in *T. reesei* grown from conidia in different carbon sources over a time period of 6 h after initial inoculation. Error bars represent the standard deviation for the 3 qRT-PCR replicates. * indicates significant difference (a p-value of <0.01 in an equal variance, one-tailed T test) between 4 h glucose and 2 h, 6 h glucose expression levels and between 6 h and 4 h cellobiose, xylose or glycerol expression levels.

To summarise, mRNA levels from genes encoding proteins of the UPR and proteins involved in cell signalling reached very high levels when compared to polarisome components or potential cortical markers. For all categories studied, transcript levels followed a different pattern in glucose than in the other carbon sources: in glucose gene mRNA levels peaked at 4 h and decreased at 6 h, whereas in cellobiose, xylose and glycerol transcript levels were highest at 6 h. The mRNA levels from *bud1*, *sec63* and from the genes encoding potential cortical markers were very low throughout all conditions studied. On the other hand transcript levels of *pdi1* and *bip1* were very high at 0 h and 2 h already.

3.3.4 Growth on solid media

To further study the influence of different carbon sources on spore development, *T. reesei* QM6a and RutC30 spores were inoculated on solid media supplemented with different carbon sources (Figure 3.3.10). After 3 days of incubation, signs of mycelia growth were clearly visible on all carbon sources. Growth was best on xylan, glucose and xylose and slowest on cellulose and straw for both strains.

There were also differences between the two strains: RutC30 did not grow as well on glucose, xylose and cellulose as QM6a. Growth was similar on xylan and both strains produced spores after 6 days of incubation (green colouration). This was also the case when grown on cellulose, straw and glucose.

It seems that different carbon sources influence the rate of growth of both strains; in general, the more complex the sugar, the slower the growth. The lack of a carbon source does not sustain growth (control plates, Figure 3.3.10). Some carbon sources also seem to trigger more sporulation than others: the agar plate containing xylan as a carbon source was covered in dark green conidia and contained most spores when compared to the other carbon sources. After 6 days of incubation, green patches of conidia were also visible on agar plates containing cellulose, straw and glucose although this was more pronounced for the wild-type strain QM6a than for RutC30. Very few conidia were present when xylose was the main carbon source for both strains.
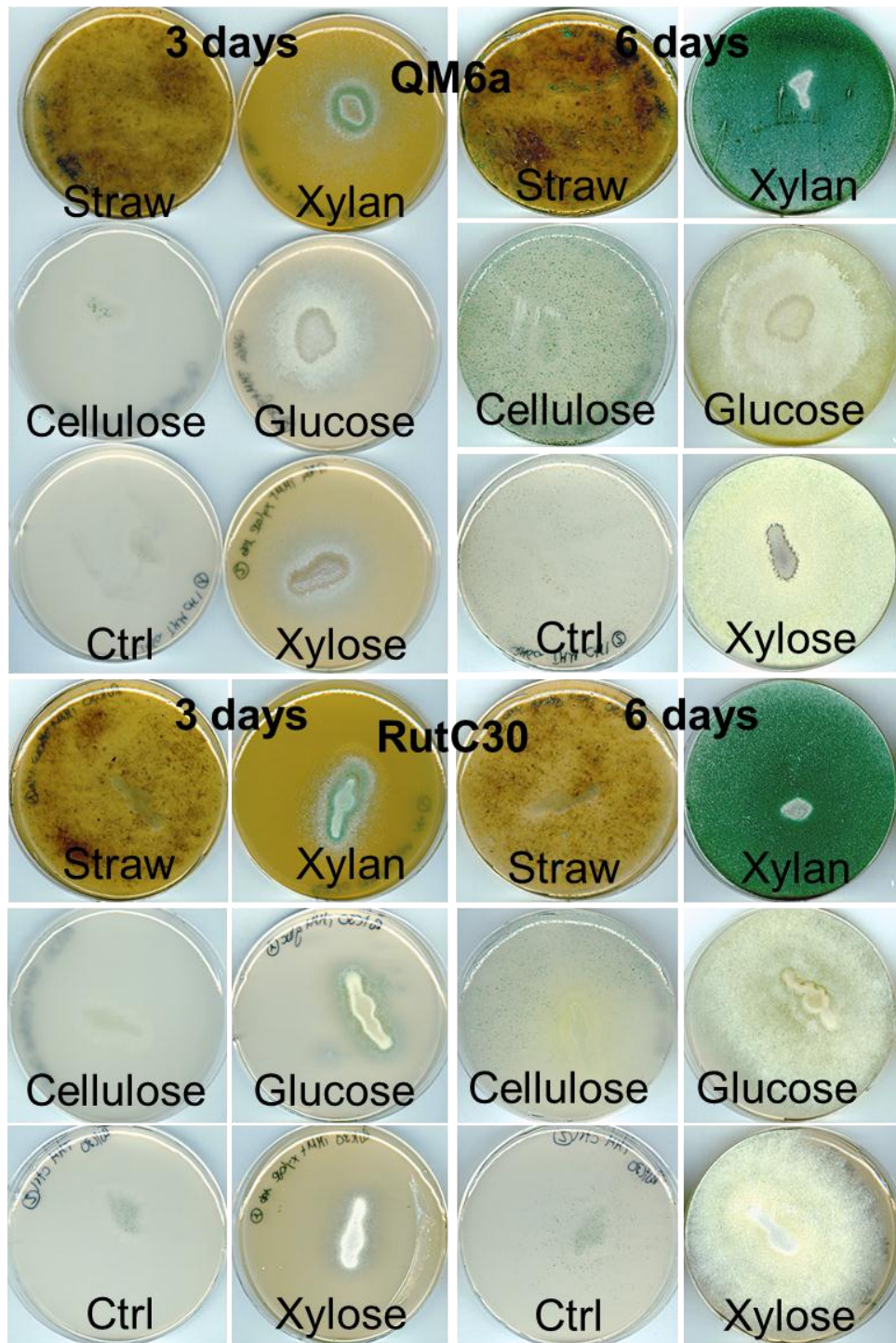
Figure 3.3.10: Growth after 3 and 6 days of *T. reesei* QM6a and RutC30 on agar plates containing 0.5% xylan and 1% of the other carbon sources. The control (ctrl) plates contain no carbon source.

3.3.5 Mycelial growth

Fungal growth was also studied in liquid media containing different carbon sources and the total mycelial biomass was assessed at different time points (Figure 3.3.11). The times and weights of biomass shown in Figure 3.3.11 were plotted onto semi-log paper (See Appendix A0 Figure A.0.F.2 for an example) and specific growth rates and doubling times were calculated as described in section 3.2.7.

The mycelial biomass reached a level that could be accurately weighed at 15 h for TCM supplemented with glucose and 20 h for mycelia grown in TCM supplemented with the other carbon sources. In TMM, biomass could only be weighed after 24 h for mycelia grown in glucose and cellobiose and 30 h when grown in xylose, glycerol and sophorose.

In TCM, QM6a grew best in glucose and glycerol, followed by lactose (autoclaved and non-autoclaved). RutC30 grew best in glucose and almost similar amounts of biomass were reached in lactose (autoclaved and non-autoclaved) but less RutC30 biomass was recorded for growth in glycerol. QM6a grew better in glucose and glycerol than RutC30 whereas RutC30 grew better in lactose and soluble cellulose when compared to QM6a. The least amount of dry biomass recorded for both strains was in soluble cellulose (sodium carboxymethylcellulose; Sigma Aldrich) and mycelia started to die after 44 h (QM6a) and 56 h (RutC30) (Figure 3.3.11).

In TMM, the highest amount of biomass was obtained when QM6a was grown in cellobiose followed by glucose and glycerol; the least amount of biomass was recorded in xylose and sophorose (Figure 3.3.12). Less biomass was present when mycelia were grown in TMM than in TCM for glucose and glycerol and growth rates and doubling times also differed between conditions (compare Tables 3.3 and 3.4). QM6a produced around 6.3 g/L and 5.5 g/L mycelia in TCM with glycerol and glucose and around 2.8 g/L and 4.5 g/L in TMM with glycerol and glucose. To summarise, different carbon sources influence the growth rate and the time it takes for the fungal biomass to double when cultured in TCM and TMM (Tables 3.3 and 3.4). The composition of the media (TCM or TMM) also seems to affect growth rate and doubling time.
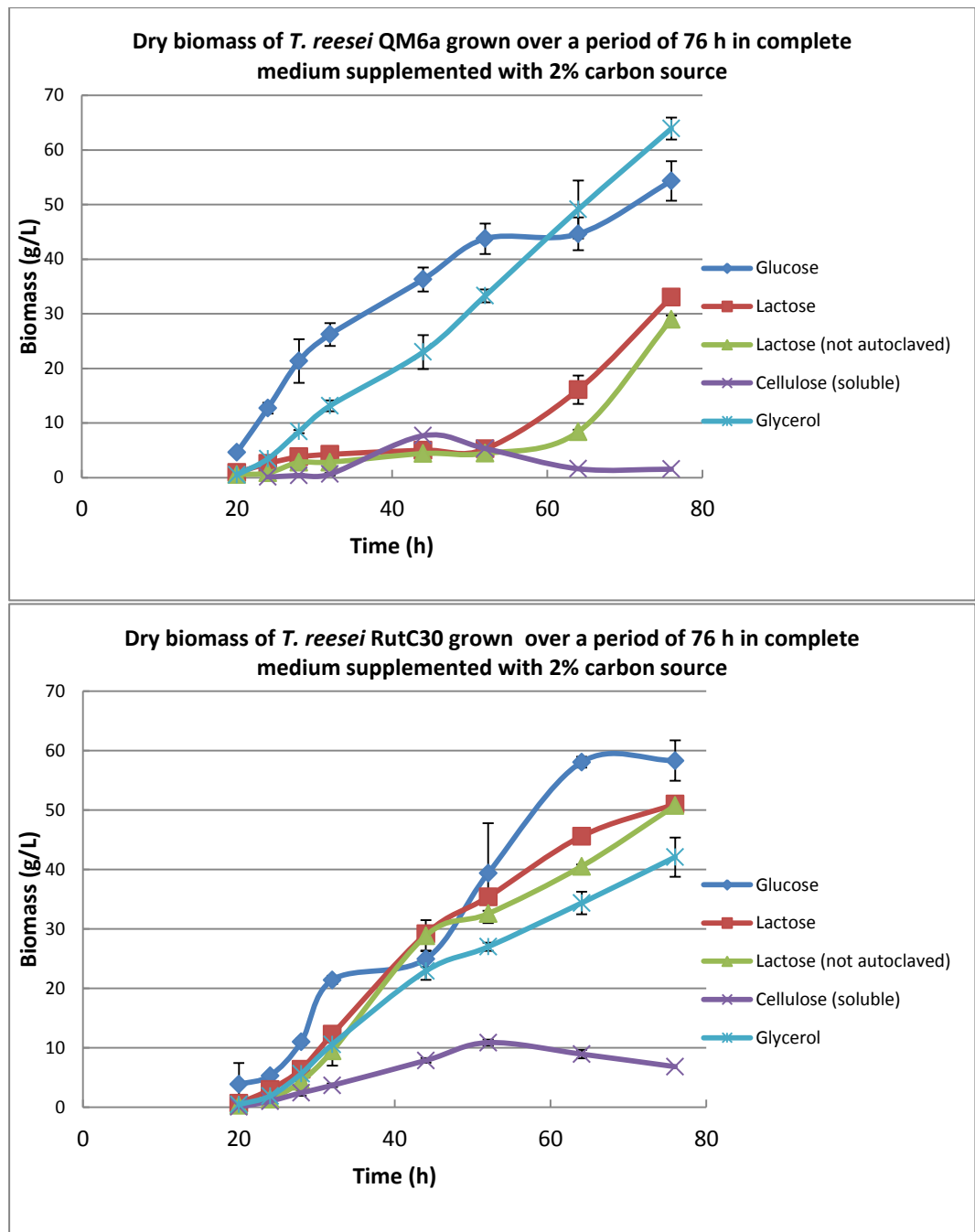
Figure 3.3.11: Growth curves of *T. reesei* QM6a and RutC30 in TCM supplemented with 2% of various different carbon sources over a time period of 76 h. Soluble cellulose corresponds to sodium carboxymethylcellulose. Standard deviation of 3 independent replicates is shown as vertical error bars.

Table 3.3: Specific growth rates and doubling times of QM6a and RutC30 when grown in TCM supplemented with 2% (w/v) of the respective carbon source.

	QM6a		RUTC30	
C-source	Doubling time (hours)	Specific growth rate μ_m (h^{-1})	Doubling time (hours)	Specific growth rate μ_m (h^{-1})
Glucose	2.70	0.26	5.0	0.14
Lactose	10.2	0.07	4.3	0.16
Lactose (not autoclaved)	12.1	0.06	4.5	0.15
Cellulose	31.0	0.02	7.0	0.10
Glycerol	3.80	0.18	4.1	0.17
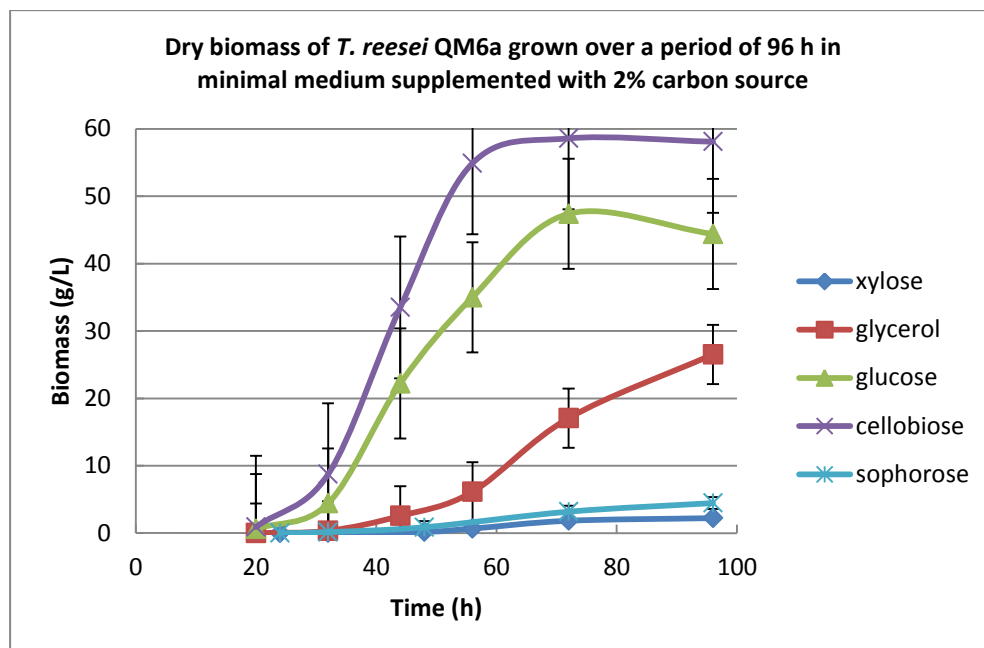
Figure 3.3.12: Growth curves of *T. reesei* QM6a grown in TMM supplemented with 2% of various different carbon sources over a time period of 96 h. Standard deviation of 3 independent replicates is shown as vertical error bars.

Table 3.4: Specific growth rate and doubling times of QM6a when grown in TMM supplemented with 2% (w/v) of 5 different carbon sources.

C-source	Doubling Time (dt) (hours)	Specific growth rate μ_M (h^{-1})
Glucose	4.3	0.16
Cellobiose	4.0	0.17
Xylose	5.0	0.14
Glycerol	6.0	0.12
Sophorose	5.4	0.13

3.4 Discussion

The impact of different carbon sources (glucose, cellobiose, xylose and glycerol) on spore development and hyphal growth of *T. reesei* was studied. In all carbon sources, spores underwent isotropic expansion followed by germ tube emergence and hyphal extension. The timing of these steps differed between conditions: in glucose (Figure 3.3.1) and cellobiose, symmetrical expansion of the spore cell happened between 2 and 4 hours after initial inoculation with germ tube emergence at around 6 h (Appendix A0 Figure A.0.F.1). At 8 h and 10 h, filamentous hyphae were clearly visible under the microscope (Figure 3.3.1 and Appendix A0 Figure A.0.F.1). In xylose and glycerol, on the other hand, isotropic expansion and germ tube emergence were much slower and happened at later time points: spore swelling was observed between 4 h and 6 h after initial inoculation with germ tube emergence at around 8 h (Appendix A0 Figure A.0F.1). These data were also supported by flow cytometry analysis of the spore population sizes (Figure 3.3.3): changes in the population size of 100,000 spores were not apparent after 8 h in xylose and glycerol whereas in glucose changes in spore size became evident at 6 h. After 12 h of incubation, the majority of spores had produced hyphal filaments in all 4 carbon sources. More hyphae were visible in glucose and cellobiose than in xylose and glycerol due to spores germinating earlier in those carbon sources. This difference in the timing of germination between carbon sources also had an effect on subsequent mycelial biomass production. Most fungal biomass was produced in glucose and cellobiose followed by glycerol and xylose when grown over a time period of 96 h (Figure 3.3.12). The growth phase where the maximal specific growth rate can be achieved was reached much earlier in glucose and cellobiose at around 30 h, whereas in xylose and glycerol the exponential phase was attained after 50 h of incubation. Spores were obtained from agar slopes supplemented with the respective carbon source (see Materials and Methods 3.2.1) and thus conidia were adapted to the respective carbon source before being inoculated in liquid cultures supplemented with the same carbon source. The maximum specific growth rate was higher in glucose and cellobiose (Table 3.4) and more mycelia were produced during the same period of time than when compared to growth in xylose and glycerol.

Conidia of *T. reesei* mainly contain CBH2 as well as CBH1, EGL2 and BGL1 on the cell surface, which synergistically act together and cleave off glucose and/or other cellooligosaccharides from cellulose and this enables the fungus to grow on this substrate (Suto *et al.,* 2001). Deletion of *cbh2* and *egl2* but also of *cbh1* resulted in reduced growth and cellulase transcript levels when conidia were inoculated in media containing cellulose as the sole carbon

source (Seiboth *et al.*, 1997). It is thought that BGL1 either cleaves cellobiose into individual glucose monomers or converts it by transglucosylation activity to sophorose. Once these simpler sugars are cleaved off by the enzymes associated to the spore cell wall, they can be taken up by the cells and induce cellulolytic and xylanolytic genes required for carbon utilisation whilst germinating and growing over the substrate. Glucose is a preferred carbon source, as it can readily be taken up by the cells and is of high nutritional value for the organism. Glucose uptake was shown in germinating *A. nidulans* spores, where, during isotropic expansion, it occurred with low affinity; once spores had started germinating glucose was taken up with higher affinity (MacCabe *et al.*, 2003). In contrast to *S. cerevisiae* which takes up glucose by facilitated diffusion, glucose uptake in filamentous fungi, such as *Trichoderma* spp. and *Aspergillus* spp., requires energy. Furthermore, cellobiose, a disaccharide which is cleaved by BGL1 into individual glucose monomers as well as xylose and glycerol require energy-dependent transport in order to be taken up into the cell. The genome of *T. reesei* contains 161 genes encoding annotated major facilitator superfamily (MFS) proteins such as pentose permeases (e.g. xylose), hexose permeases (glucose, fucose and maltose), transporters, monocarboxylate transporters and other sugar transporters (see Appendix A4 Table A.4.T.1 for a few examples). In *A. niger*, proteins of the major facilitator superfamily are also extremely abundant with 461 genes predicted to encode MFS transporters in this organism (Pel *et al.*, 2007).The MFS superfamily of transporters can be divided into a further 17 families of which families 1, 5 and 7 organise monosaccharide (hexoses, pentoses) and disaccharide transport into the cell by coupling it to proton symport or antiport (Pao *et al.*, 1998). These MSF transporters are very likely to organise the import of sugars into spores and hyphae of both species of filamentous fungi (*T. reesei* and *A. niger*) and some of them are likely to be involved in nutrient sensing as well. It is not known though how many of these transporters are present in the cell membrane of conidia in *T. reesei*. The difference in the timing of germination between the different carbon sources may therefore be explained, at least in part, by the preferred uptake of glucose in spores. Germination in cellobiose was also quicker than in xylose and glycerol as the cleavage of this disaccharide provides more glucose for uptake. Extracellular and intracellular BGLs cleave cellobiose, which can be taken up by the cell, into its glucose components (Suto *et al.*, 2001; Kubicek *et al.*, 2009). Glucose can enter glycolysis directly upon entering the fungal cell wall where it is used to generate ATP. Xylose on the other hand needs to be converted to D-xylulose-5-phosphate first, a process which requires ATP hydrolysis, before it can enter the pentose phosphate pathway and generate energy (Mach-Aigner *et al.*, 2011). Thus, xylose utilisation is a less exergonic

metabolic process and could be the reason for later spore germination. Glycerol is a polyol which is used by many fungi as a regulator of osmotic pressure within the cell. In *A. nidulans*, osmoadaptation requires the expression of genes encoding enzymes such as glyceraldehyde-3-phosphate dehydrogenase involved in glycerol biosynthesis (Ferreira de Oliveira *et al*, 2011). Glyceraldehyde-3-phosphate is also involved in glycolysis and thus glucose uptake and conversion into ATP goes hand in hand with regulating osmotic pressure within the cell. It was shown that glycerol is synthesised during the first hour of germination within spores and that it is also taken up during starvation conditions supporting a further role for this substance as a carbon storage compound (Wititeveen *et al.*, 1995). Furthermore, glycerol cannot diffuse through the cell membrane and needs to be transported into the cell. Taken together, the late germination and accumulation of biomass in glycerol-rich minimal media when compared to glucose-rich TMM could be due to osmotic pressure regulation and the expression of a transport system for glycerol uptake.

It was observed by microscopy that during the first 8 hours of spore development about a quarter (20% to 30%) of spores did not germinate at all nor did so at a much later time point. This was also supported by flow cytometry analysis where in glucose and cellobiose, two distinct population sizes became apparent at the later time points (6 h – 8 h). One out of the two spore populations had increased in size whereas the other had the same size as the spore population at time zero (Figure 3.3.3). The reason for this is unknown but it can be partially linked to the media used and not to the age of spores as slopes were re-grown each time for microscopy and flow cytometry analysis. In minimal media the overall number of spores which germinated was lower than in complete media where up to 90% of all spores could germinate (microscopy analysis only, data not shown due to TCM media interfering with flow cytometry analysis). The reason for this is not known and there has been no research carried out on this.

The same trend which was observed for spore germination in different carbon sources was also recorded at the mycelial level. Over a period of 96 h, growth in cellobiose and glucose was much faster and resulted in a higher accumulation of fungal biomass than in glycerol and xylose (Table 3.4, Figure 3.3.12). High internal ATP concentrations have been shown to coincide with high growth rate in *T. reesei* QM9414; i.e. a faster accumulation of ATP supports faster growth (Fuji *et al.*, 2010). As mentioned before, glucose is the preferred carbon source for most eukaryotes as it is metabolised readily and converted into energy. *T. reesei* uses aerobic respiration to metabolise glucose and it has been shown that genes encoding enzymes of the glycolysis

pathway, the TCA cycle and the mitochondrial respiratory chain are highly up-regulated in the presence of 100 mM glucose concentrations (Chambergo *et al.*, 2002). Furthermore, high glucose concentrations also trigger the formation of acetate from acetaldehyde. Upon glucose depletion in *T. reesei*, acetate can re-enter the TCA cycle and more protons and ATP molecules are formed. Growth was also recorded in the presence of sophorose, a known inducer of cellulases. Spores were inoculated in glycerol rich medium supplemented with sophorose but the growth rate in sophorose was much slower when compared to glycerol and yielded less biomass (Figure 3.3.12). The most likely explanation for this is that the induction and secretion of high levels of cellulases uses ATP which otherwise, e.g. in the presence of glucose and cellobiose, would be exclusively used for mycelial growth. Sophorose uptake has been shown to be inhibited in the presence of glucose (Kubicek *et al.*, 2009).

A difference between complete and minimal media was also observed from the growth curves in glucose and glycerol with higher growth rates and doubling times achieved in complete media when compared to minimal media (see 3.3.5). This was due to the presence of amino acids in complete media (Materials and Methods section 2.4). Amino acids can also be used as potential carbon sources and generate energy through various metabolic pathways. QM6a and RUTC30 both grew very quickly in glucose and glycerol and specific growth rates were similar in the presence of these substrates (Table 3.3). It has been reported that lactose is a good inducer for genes encoding cellulases and hemicellulases in *T. reesei* (Stricker *et al.*, 2007) but this was not the case in this project when compared to growth in the presence of glucose or glycerol (Figure 3.3.11). On the other hand, the growth rate of RutC30 was twice as high as QM6a in lactose (Table 3.3) and it produced 4 times the amount of biomass (40 g/L) after 60 h of growth than QM6a (Figure 3.3.11). In the presence of soluble cellulose neither strain grew well and both started to decline after 50 h. Soluble cellulose (carboxymethyl cellulose) has carboxymethyl groups added to some of the hydroxyl groups of the glucose units. This may limit the access of cellulases to the substrate thus decreasing efficient cellulose degradation and lower the levels of glucose release. Xu *et al.* (2009) showed that oxidation by cupric ion ($CuSO_4$) of the reducing ends of cellulose micro-fibrils severely decreased cellulose hydrolysis by different cellulases. The authors speculated that the oxidation of the cellulose reducing ends to negatively charged aldonic acids would interfere with the binding of the otherwise neutral un-oxidised cellulose chain ends in the active site tunnel of CBH1. Furthermore, addition of a *T. reesei* enzyme mixture (CBH, EGL and BGL) to oxidised cellulose resulted in substrate hydrolysis similar to non-oxidised cellulose at the early stages; however after

prolonged incubation, cellulose hydrolysis decreased when compared to un-oxidised cellulose hydrolysis, probably due to an accumulation of oxidised cellulose chain ends (Xu *et al.*, 2009).

The type of carbon source also influenced mycelial growth on solid media. Agar plates supplemented with xylose, xylan and glucose sustained faster radial outgrowth of the colony whereas growth on cellulose and straw was much slower and less mycelia were present on these plates (Figure 3.3.10). This is because straw and cellulose are complex carbohydrates which need the expression and secretion of many enzymes for degradation (discussed in Chapters 4 and 5), hence slowing growth down when compared to simpler polysaccharides such as xylan or the monosaccharides xylose and glucose. Furthermore *T. reesei* possesses an array of hemicellulases which are highly induced by xylan and are able to degrade it [e.g. XYN1 (xylanase), XYN2 and BXL1 (β-xylosidase)]. As mentioned before simpler sugars are taken up and metabolised more readily, fuelling a quicker growth.

As observed in liquid cultures, there was also a difference between QM6a and RutC30 growth on solid substrates (Figure 3.3.10). The reason for the difference between the two strains most likely relates to the fact that RutC30 has many mutations within its genome (Le Crom *et al.*, 2009). RutC30 is a strain which was obtained from 3 mutational steps: first UV mutagenesis which generated strain M7, then chemical treatment by N-nitrosoguanidine (NTG) which generated strain NG14 and finally a last round of UV mutagenesis. After each of these steps, colonies were selected which were able to hydrolyse cellulose during carbon catabolite repressive conditions. In total, RutC30 possesses 223 single nucleotide variants, 15 small deletions or insertions and 18 larger deletions leading to the loss of more than 100 kb of genomic DNA (Le Crom *et al.*, 2009). The 15 small deletions or insertions led to frameshift mutations in 2 genes including in a previously reported β-glucosidase subunit expressed in the ER and involved in quality control of proteins (Seidl *et al.*, 2008). The large deletions included the already identified 85 kb deletion from scaffold 15 and the truncation of the *cre1* gene (Seidl *et al.*, 2008). Furthermore, mutations were also generated in 43 genes encoding functionalities that include nuclear transport, mRNA stability, secretion/vacuolar targeting, metabolism and several fungal specific transcription factors (Le Crom *et al.*, 2009). That is why the growth profiles on various different carbon sources differ when compared to QM6a due to the many mutations in genes crucial for metabolism and various other processes.

The expression of 12 selected genes was also monitored to assess their potential role in governing spore development in *T. reesei*. The 12 genes were

chosen based on their relevancy in spore development in *Aspergillus* spp. and homologues for all genes were found in *T. reesei* (Table 3.2). These genes encoded 3 potential cortical markers (*rax1*, *axl2*, *bud4*), 2 GTPases (*rac1*, *mod1*), 3 polarisome components (*sep1*, *spa1*, *bud1*) and 4 proteins involved in the UPR (*pdi1*, *hac1*, *bip1*, *sec63*). Homology between *A. nidulans* and *T. reesei* proteins encoded by these genes was fairly high and the only exception for this was the UPR protein HAC1, a transcription factor for which homology was largely restricted to the Zn finger region and which has been characterised in *T. reesei* previously (Saloheimo *et al.*, 2003). The highest sequence identity between the two organisms was for the GTPases and BIP1, a molecular chaperone, supporting previous findings that these proteins are highly conserved across organisms due to their essential role in signalling pathways and protein folding (Momany, 2005).

The expression of these genes was studied via qRT-PCR in 4 different carbon sources at time points 0, 2, 4 and 6 hours after initial inoculation. The genes with the highest expression levels were those encoding components of the UPR such as *pdi1* and *bip1* (Figures 3.3.8 and 3.3.9). Upon breaking spore dormancy and during isotropic expansion, metabolism and protein synthesis are increased (Deacon, 2006). Furthermore, protein synthesis is required to sustain apical growth in nutrient rich conditions and secretion and synthesis of proteins at the hyphal tip is continuous. This increase in protein synthesis during spore development, germination and hyphal growth puts stress on the cell and activates the UPR. The UPR is a homeostatic response and essential for correct folding of secreted proteins: accumulation of mis-folded proteins in the endoplasmic reticulum would otherwise prove lethal to the organism (Geysens *et al.* 2009); thus high levels of expression of the main components of the UPR are required for spore germination and would be expected to increase during the early stages of spore development.

The second category of genes with high expression levels were those encoding the Rho GTPases MOD1 and RAC1. MOD1 expression levels were higher than RAC1 and this GTPase may be more essential for signal transduction during spore development than RAC1 (Figure 3.3.6). *S. cerevisiae* possesses one Rho GTPase termed Cdc42p which is essential for bud site selection whereas filamentous fungi have an expansion of the number of these enzymes. This highlights one of the differences in growth between yeast and filamentous fungi and underlines the complexity associated with hyphal polarised growth. Based on the high sequence similarity to other organisms, it is very likely that MOD1 and RAC1 also relay signals between the extracellular space and the morphogenetic machinery in *T. reesei*. RAS1, a small GTPase is repressed by CRE1, the global carbon catabolite regulator, at high growth

rates (Portnoy *et al*., 2011). In *S. pombe*, Ras1 regulates two distinct pathways: one controls mating through the Byr2-MAPK cascade and the second one signals through Scd1-Cdc42 to maintain elongated cell morphology (Portnoy *et al*., 2011). Interplay between CRE1 and RAS1 at different growth rates may affect subsequent signalling of MOD1 and RAC1 and therefore be important for conidial development and maintencance of polarised growth in *T. reesei*.

The expression of genes encoding polarisome components was also shown to increase during the first few hours after inoculation into a medium with a usable carbon source. SEP1 is the key polarisome component which was also reflected here by higher gene expression levels when compared to *spa1* and *bud4* (Figure 3.3.7). The latter two genes encode proteins which regulate the timing and location of SEP1 (see introduction). The expression of these genes was relatively low when compared to those encoding UPR components and the Rho GTPases and this may reflect the importance of these genes and the proteins they encode. The UPR is a homeostatic response keeping a check on unfolded proteins, which could otherwise be lethal to the cell, whereas Rho GTPases such as RAC1 and MOD1 are involved in many different signalling pathways. Alternatively, low transcript levels of *spa1* and *bud4* are sufficient due to high impact these proteins may have in regulating SEP1 timing and location.

The last category of genes studied were those encoding putative "cortical markers". The transcript levels from *axl2*, *rax2* and *bud4* were about half of the transcript levels from genes encoding polarisome components. This is especially true for the expression levels of *axl2* which remained very low throughout all conditions studied and only increased slightly at 6 h after initial inoculation in cellobiose, xylose and glycerol and after 4 h in glucose before decreasing again (Figure 3.3.5). It is questionable whether the protein encoded by this gene has a role in marking a site for germ tube emergence due to the very low levels of transcripts present in all the conditions studied here. On the other hand *axl2* encodes a calcium-binding trans-membrane protein which interacts with other proteins and which, in animals, mediates cell adhesion and cell polarity (IPR006644). Furthermore, Axl2p is involved in bud site selection in *S. cerevisiae*. Transcript levels of *bud4* also increased after 4 h of initial inoculation in glucose and after 6 h in cellobiose, xylose and glycerol, and is the gene with the highest expression levels out of all three cortical marker genes studied (Figure 3.3.5). The Bud4p is involved in bud site selection in *S. cerevisiae* and the corresponding genes in *A. nidulans* and *T. reesei* encode a putative GTP binding protein. BUD4 contains a PH (pleckstrin homology) domain which can bind the βγ subunit of G-proteins or to protein

kinases (IPR001849). The exact role of BUD4 is unknown and further functional studies are required to determine whether this could be a potential cortical marker in *T. reesei* or whether it is involved in the spore germination signalling pathway through recruiting G-proteins or other protein kinases. It is possible though that this protein is involved somehow in spore germination and development in *T. reesei*. The transcript levels of *rax2* also increased at the later time points (4 h and 6 h) in the different carbon sources studied here. The protein encoded by this gene contains an amine oxidase protein domain and a Sec-1 like protein domain (IPR002937, IPR001619) with the former involved in the oxidation of amines and the latter involved in vesicle transport and exocytosis. This gene may be a good candidate to be involved in the process of germ tube emergence as this requires localisation of vesicles to the chosen site for subsequent protein secretion and cell wall synthesis. Again, further studies are required to determine the exact function and localisation of RAX2.

Another observation made during the analysis of the qRT-PCR results was that the transcript levels of all genes within the categories examined generally differed between glucose and the other carbon sources. In glucose, expression of all genes peaked at 4 h and had already decreased at 6 h after initial inoculation. In cellobiose, xylose and glycerol, gene transcript levels either increased or stagnated during the first 4 hours but increased noticeably at 6 h. This may be due to different ways in which *T. reesei* takes up and metabolises sugars with glucose being the preferred one and thus eliciting an earlier genetic response when compared to the other carbon sources.

3.5 Conclusion

It is very clear from the results obtained here and from many previous reports that the type of carbon source influences the rate of growth and biomass accumulation. Simple, glucose-based saccharides are by far the preferred carbon source as they are metabolised very quickly and provides high energy yields. Carbon metabolism, energy generation and growth are closely linked processes and like most eukaryotic organisms, fungi have evolved means to transport sugars into the cell which provide the organism with high energy. Furthermore, these easily metabolised sugars provide the energy for enhanced hyphal growth and thus a correlation between the type of sugar and fungal biomass accumulation can be observed. This process already starts at the spore level where different carbon sources induce different rates of germination. This can be counteracted by adding transcriptional inducers such as sophorose to the cell. Once sophorose is taken up into the cell, it elicits a high transcriptional response of cellulase-encoding genes and ATP is mainly used for the making and secretion of these proteins rather than utilised to fuel fungal growth. In the presence of glucose, sophorose uptake is inhibited, further highlighting the preference of the fungus for glucose (Kubicek *et al*., 2009). Furthermore, mutations in genes encoding enzymes for various metabolic processes (e.g. in RutC30) can also alter carbon utilisation and biomass accumulation. The process by which a carbon source signals its presence to conidia, causing spore germination and polarised hyphal growth also remains largely unknown. Enzymes attached to the spore cell wall are thought to cleave off inducers which break spore dormancy. Protein synthesis in germinating spores is increased and this activates the UPR as was also shown by an up-regulation of transcript levels of the genes which encode the main components of the UPR in this project. Furthermore detection of a carbon source is likely to activate signalling components such as small Rho GTPases which then relay the signal for spore germination and hyphal extension in *T. reesei* and interact with cell-wide master carbon regulators such as CRE1. CRE1 was shown to repress RAS1 (GTPase) at high growth rates (Portnoy *et al*., 2011). Choosing a site for germ tube emergence in filamentous fungi is a largely unknown process and differs in many aspects from the known process in *S. cerevisiae* (see introduction). The three genes, *rax2*, *axl2* and *bud4*, could be potential candidates involved in the process of germ tube emergence although a thorough investigation into the roles of these genes and the proteins they encode needs to be carried out. The protein encoded by the *rax2* gene is putatively involved in vesicle transport and exocytosis, *axl2* encodes a trans-membrane protein involved in establishing cell polarity and bud site selection in animals and yeast

respectively and the protein encoded by *bud4* is likely to interact with G-proteins and protein kinases. As many different genes are proposed to be involved in site selection for germ tube emergence and maintenance of polarised growth, these three genes may only play a minor role and their encoded proteins are likely to interact with other components involved in signalling and protein secretion.

Chapter 4: Genome-wide transcriptome studies of *T. reesei* strains in the presence of different carbon sources and with comparison to *Aspergillus niger*

4.1 Introduction

4.1.1 Composition and sequencing of transcriptomes

The transcriptome of eukaryotes consists of many different types of RNA molecules; some are involved in protein synthesis (mRNA) or in the process of it (ribosomal RNA, transfer RNA and signal recognition particle RNA) whereas others are non-protein coding, functional RNAs. One type of non-protein coding RNAs are regulatory RNAs which are essential in establishing organismal complexity and provide an additional mechanism of gene transcriptional and translational regulation. These RNAs are non-coding transcripts from genic and intergenic locations and are transcribed throughout the genome. Based on a 200 nucleotide cut-off, these transcripts are classified into short RNAs and long non-protein-coding RNAs (Werner and Swan 2010). Short RNAs comprise micro RNAs (miRNAs, absent in filamentous fungi), silencing/short interfering RNAs (siRNAs, see section 4.1.5), piwi-interacting RNAs (piRNAs; interact with the piwi domain of mammalian Argonaute proteins only), PASRs (promoter-associated short RNAs) and TASRs (termini associated short RNAs). RNAs can also be associated to gene boundary regions such as promoters; these transcript species, which can be up to 600 nucleotides long, are known as PROMPTs (promoter upstream transcripts) and control gene transcription (Werner and Swan, 2010). Another category of long regulatory RNAs are natural antisense transcripts (NATs) which are non-protein coding, fully processed mRNAs from the non-coding strand and partially overlap the protein-coding transcripts (Faghihi and Wahlestedt, 2009). NATs have many regulatory roles, including generating siRNAs, and these will be described in section 4.1.5.

Analysis of the pool of RNAs present within a cell or a collection of cells under certain conditions has been made available through high-throughput sequencing technologies. These technologies with which cDNA (reverse-transcribed from RNA) is sequenced, require low nucleotide input, provide deep coverage and base-scale resolution and are more accurate and faster than bacterial cloning of cDNA, micro- and tiling arrays. The resulting sequence reads are individually mapped to the source genome and counted

to obtain the number and density of reads corresponding to RNA from each known exon, splice event or new candidate gene (Mortazavi *et al.*, 2008). SOLiD-based (see materials and methods) RNA-sequencing was applied in this chapter to access information about the expression of genes involved in the process of lignocellulose (i.e. wheat straw) deconstruction and to screen the genome of *T. reesei* for NATs.

4.1.2 Accessory proteins involved in plant cell wall deconstruction

4.1.2.1 Expansins

The degradation of complex plant polysaccharides not only involves the synergistic action of glycoside hydrolases such as cellulases and hemicellulases (see Chapter 1, section 1.4), but also involves additional plant expansin-related proteins such as swollenins (so called because of its swelling activity on cotton fibres) which help in cell wall degradation. These enzymes are thought to play a role in the loosening of the plant cell wall by disrupting hydrogen bonds between plant polysaccharides, causing a displacement of cellulose fibers and increasing water uptake. As a result, the cell wall area is enlarged which allows access of hydrolytic enzymes to the underlying polymers. Expansin-related proteins are non-hydrolytic enzymes as no sugars are produced during their activity (Saloheimo *et al.*, 2002). In other *Trichoderma* spp. such as *T. asperellum*, swollenins significantly enhance the plant colonising capability of the fungus by disrupting the cell wall (Schuster and Schmoll, 2010). The genome of *T. reesei* encodes 5 plant expansin-related proteins (*swo1*, *swo2*, *eel1*, *eel2* and *eel3*).

The SWO1 protein has a signal sequence and a typical, fungal carbohydrate binding module (CBM1) at the N-terminal end connected by a linker region of serine and threonine residues to the catalytic domain (CD). The C-terminal end of SWO1 shows high similarity (25%) to plant expansins. Furthermore, SWO1 shows sequence similarity to fibronectin III (FnIII) type repeats of mammalian titin proteins (Saloheimo *et al.*, 2002). These FnIII repeats form β-sandwich domains which fold and unfold easily and are thus able to stretch the protein. The *swo1* gene is located adjacent to the *cbh1* gene on scaffold 29 in the *T. reesei* genome. Like the other cellulases, *swo1* is up-regulated in the presence of cellulose, sophorose, lactose and cellobiose, but production levels of SWO1 are less than those of the main cellulases (Saloheimo *et al.*, 2002). The *swo1* gene is also de-repressed after long cultivation in glucose-rich conditions and low levels of expression are recorded in the presence of glycerol and L-sorbose. SWO1 shows activity against the cellulose components

of algal cell walls, filter paper and cotton fibers without releasing any sugar. When this protein is incubated with cellulases such as CBH1 or EGL1, the total amount of sugar released from algal cell walls, filter paper and cotton fibers is increased (Saloheimo *et al.*, 2002). The regulation of *swo1* is similar to the regulation of cellulase-encoding genes and the promoter of *swo1* contains a putative XYR1 binding motif (GGCTAT) and another GGTAAA motif which could bind a yet unknown regulatory protein in the presence of sophorose (Verbeke *et al.*, 2009). So far, no CRE1 binding site has been identified within the *swo1* promoter region and further studies are required to fully elucidate *swo1* gene regulation.

The EEL (expansin/family 45 endoglucanase-like) proteins show significant similarity to plant expansins (25% - 30%) but do not possess a fungal or plant expansin cellulose binding domain. Furthermore EEL3 has 88% identity with MRSP1 from *T. virens*, a protein which is possibly involved in cell wall modification in germinating conidia (Verbeke *et al.*, 2009). The genes *eel1* and *eel3* are located within genome clusters enriched in genes encoding CAZymes but their promoters do not contain any XYR1 binding motifs. On the contrary, the *eel2* promoter has two perfectly matching XYR1 binding sites and was shown to be constitutively expressed in the presence of lactose, sophorose, cellulose, sorbose and cellobiose (Verbeke *et al.*, 2009). Expression of the other two *eel* genes on any of these substrates was not recorded and it remains to be determined under which conditions these genes are induced and how they are regulated. Furthermore, a CEL12A endoglucanase from *T. reesei* has been shown to disintegrate cellulose fibres by being capable of initiating plant cell wall extension thus weakening xyloglucan-cellulose interactions (Hilden and Johansson, 2004).

4.1.2.2 Glucose-ribitol dehydrogenase

An additional protein which is thought to enhance cellulase activity in the presence of sophorose and cellulose is GRD1 (glucose-ribitol dehydrogenase). Enzymes of this group catalyse the oxidation of D-glucose to D-β-gluconolactone. GRD1 was shown to have enzymatic activity in the presence of cellobiose. The *grd1* gene does not encode a signal sequence indicating that the derived protein is probably not secreted into the extracellular environment but rather stays in the cytoplasm. Furthermore, *grd1* is found within a putative secondary metabolism cluster in *T.reesei* which underlines a connection between carbohydrate metabolism and secondary metabolism (Schuster *et al.*, 2011). The genome of *T. reesei* encodes another 9 putative GRD proteins and orthologues are found in several *Aspergillus* spp. and in *T. virens* but not in *A. niger* (Schuster *et al.*, 2011). Transcription of *grd1*

correlates with cellulase gene induction such as *cbh1*; lactose and especially sophorose induce high levels of *grd1* expression. Deletion of *grd1* severely reduced transcription levels and secretion of cellulases supporting a role for this enzyme in cellulase gene regulation. It is known that decreased BGL activity leads to higher cellulase activities by maintaining a larger pool of inducing substrates (Seiboth *et al.*, 2007) and it is speculated that GRD functions as an intracellular sensor of extracellular cellulolytic activity by checking the activity of intracellular BGL (mechanism is unknown) in order to prevent inhibitory glucose build up (Schuster *et al.*, 2011). This also allows signalling of the presence of cellulose outside the cell; taking up cellobiose via a specialised permease could activate GRD1 which then in turn could indirectly control cellulase expression levels.

4.1.2.3 Hydrophobins

A further class of proteins, ubiquitous to fungi and thought to mediate attachment of the fungus to plant carbohydrates are hydrophobins. Hydrophobins can adsorb to hydrophobic surfaces and to interfaces between hydrophobic (air, oil) and hydrophilic phases (water, cell wall), thus mediating the interaction of the fungus with its environment (Ohtaki *et al.*, 2006). The most important role of hydrophobins is to interact with surfaces, coating the surfaces and lowering surface tension (Linder *et al.*, 2005). The hydrophobic part of the amphiphilic protein faces hydrophobic surfaces such as oil and air, whereas the hydrophilic part interacts with water or cell walls and membranes (hydrophilic).

Hydrophobins are small (between 75 and 400 amino acids) secreted, mainly hydrophobic proteins with eight conserved cysteine residues (Seidl-Seiboth *et al.*, 2011). These proteins can be grouped into either class I or class II of hydrophobins according to their solubility in solvents, hydropathy profiles and spacing between the conserved cysteines. Class I hydrophobins only dissolve in strong acids whereas class II hydrophobins can be dissolved using aqueous dilutions of organic solvents. Furthermore, class I hydrophobins have a longer putative first β hairpin loop and incorporate more hydrophobic residues, resulting in a larger hydrophobic patch on the surface of this protein (Sunde *et al.*, 2008). Sequence similarity between the two classes and within class I is usually very low whereas primary amino acid sequence is more conserved in class II hydrophobins. The genome of *T. reesei* encodes 6 class II hydrophobins (HFBI, HFBII, HFBIII, HFBIV, HFBV and HFBVI) of which the 3D structure of HFBII has been resolved. HFBII has one N-terminal β hairpin and one C-terminal β hairpin structures, both of which connect and interlock to form a barrel-like structure, exposing a large hydrophobic patch at the surface. About

half of the hydrophobic aliphatic residues of HFBII are located in this hydrophobic patch. The structure of the protein is stabilised by four disulphide bonds formed between the 8 cysteine residues which are located inside the structure and are clustered into two parts of the protein. The amino termini of hydrophobin proteins are variable in length and composition and are thought to confer structural and binding differences to individual proteins from the same class. Furthermore, the export of *T. reesei* hydrophobins requires processing (signal peptide removal and pro-peptide cleavage) and maturation (disulphide bond formation and correct folding): HFBI is processed and disulphide bonds are formed to give rise to a 7.533 kDa protein (original mass is 9.874 kDa) and HFBII has a size of 7.188 kDa (from 8.766 kDa) after processing and disulphide bond formation (Neuhof *et al.*, 2007).

The *T. reesei* HFBII hydrophobin is found on conidial cell walls and is also secreted by hyphae into the medium; HFBI and HFBIII on the other hand are largely retained in the mycelium in vegetative cultures (Linder *et al.*, 2005). Mycelia in differing physiological states express different hydrophobin-encoding genes in relation to the morphological state (Neuhof *et al.*, 2007). This has also been observed for *Hypocrea atroviridis* whose genome encodes 11 class II hydrophobins: one gene encoding HFB-2B (closely related to HFBI and HFBII of *T. reesei*), is expressed in light, dark and under carbon starvation; HFB-1B, HFB-2A and HFB-5A are up-regulated when switching from darkness into light and HFB-6A, HFB-6B and HFB-6C are only expressed in the presence of light (Mikus *et al.*, 2009). Hydrophobins such as HFBII, coat the outer cell walls of conidia, protecting against desiccation and wetting, and helping in the water and air mediated dispersal of spores. In solution, *T. reesei* HFBI and HFBII self-assemble into crystalline rodlet-like structures or fibrils, consisting of protein chains of 2-3 µm in diameter and 15-25 µm in length on the outer cell walls of mycelia and conidia where they reduce water tension and aid the formation of aerial structures such as hyphae and conidia (Sunde *et al.*, 2008). How these proteins oligomerise is unknown but hydrophobin assembly at interfaces and on fungal cell walls is a reversible process. Hydrophobins can also adsorb to solid surfaces, helping to attach mycelia to the substrate and enhancing growth. In the case of *T. reesei* HFBII, the hydrophobic part of the protein faces and coats the hydrophobic surface of the substrate whereas the hydrophilic part would face inwards and away from the surface. The surface of the hyphal wall when growing in a liquid is hydrophilic whereas the cell wall of aerial hyphae is hydrophobic (Sunde *et al.*, 2008). Thus, *T. reesei* hydrophobins help the interaction with dead plant biomass during the saprophytic lifestyle of this organism in both the soil, water and the air.

A specific example of the role of a hydrophobin involves the *A. oryzae* hydrophobin RolA which adsorbs to the surface of the hydrophobic, biodegradable plastic PBSA (polybutylene succinate-*co*-adipate) where it recruits the cutinase CUTL1 (Ohtaki *et al.*, 2006). Cutinases are carbohydrate esterases and they cleave cutin into individual monomers. Cutin is a plant and insect polyester component consisting of omega hydroxy acids interlinked by ester bonds. Recently, a protein was discovered in *A. oryzae*, termed HsbA (hydrophobic surface binding protein) which was also shown to recruit CUTL1 to hydrophobic surfaces such as PBSA and promote substrate degradation (Ohtaki *et al.*, 2006). HsbA is not a hydrophobin as it does not possess the conserved eight cysteine residues and also does not associate to the cell wall after secretion. Adsorption of HsbA to PBSA was enhanced in the presence of NaCl or $CaCl_2$, which is again different to the mechanism of attachment of conventional hydrophobins. Thus *A. oryzae* seems to use several types of surface interacting proteins required for carbohydrate degradation.

4.1.3 Sensing and signalling in *T. reesei*

It is generally accepted that filamentous fungi sense and respond to extracellular signals such as pheromones, lipids, nucleotides, peptides, nitrogen, glucose and other nutrients by G-protein signalling (Kulkarni *et al.*, 2005). G-proteins are transmembrane proteins, consisting of a G-protein coupled receptor (GPCR) and an associated heterotrimeric G-protein (α, β and γ) and are divided into 9 different classes. The GPCRs contain 7 transmembrane helices connected by intracellular and extracellular loops, with an extracellular amino terminus and the carboxy terminus extending into the cytoplasm (Li *et al.*, 2007). The extracellular amino terminus interacts in most cases with the ligand and the carboxyl terminus with the G-proteins. Attachment of the fatty acids myristate and/or palmitate at the N-terminus of Gα proteins allows correct membrane targeting and interaction with adenylyl cyclase, the Gβγ subunit and GPCR proteins (McCudden *et al.*, 2005). Similarly, lipid modifications (farnesyl or geranylgeranyl) of the Gγ C-terminus are important for the membrane localisation of the Gβγ dimer. In filamentous fungi such as *T. reesei* three genes encode Gα subunits (GNA1, GNA2 and GNA3) and one gene encodes a Gβ and Gγ subunit respectively. One of each of the three Gα subunit proteins in *T. reesei* belongs to group I, group II and group III. Group I Gα proteins are involved in many pathways and conserved amongst filamentous fungi; group II Gα subunits are less well conserved and their function is less obvious; group III Gα members are also well conserved and generally positively influence cAMP levels (Li *et al.*, 2007). Gβ proteins are

mainly involved in the asexual and sexual cycles of filamentous fungi and also stabilise the heterotrimeric G-protein complex. Gγ subunits contain a coiled-coil N-terminal domain necessary for interaction with the Gβ subunit and they seem to have similar roles to the Gβ subunits. G-protein mediated signalling is transmitted via one or more of the following pathways (Yu, 2006): 1) activated protein kinase A (PKA) through elevated cAMP levels, 2) mitogen-activated protein (MAP) kinase pathways and 3) IP$_3$(inositol triphosphate)-Ca^{2+}-DAG (diacyl-glycerol)-dependent protein kinase C (PKC).

T. reesei is different from other filamentous fungi as it possesses fewer PTH11-type GPCRs (35 in total), no class IX GPCRs (opsins) and no *S. cerevisiae* Gpr1-like glucose sensing GPCR (Schmoll, 2008). The genome of *T. reesei* encodes GPCRs from the other 8 classes including 3 putative carbon sensors (Brunner *et al.*, 2008). On the other hand, groups of cAMP-receptor like proteins, *H. sapiens* mPR-like proteins (steroid receptors) and *Aspergillus* GPRK (G-protein coupled recepor kinase)-like receptors are expanded in this fungus (Brunner *et al.*, 2008). Signalling by G-proteins needs to be regulated as defects in the transmission of signals can be detrimental to the cell; usually the rates of GTP hydrolysis determine the intensity of the signal. Phosducin-like proteins (PhLPs) positively regulate Gβγ subunit signalling by acting as molecular chaperones required for the assembly and stabilisation of both G-protein subunits; the genome of *T. reesei* encodes two PhLPs. GPCRs can also be negatively regulated and fine-tuned by RGS (regulation of G-protein signalling) domain proteins of which *T. reesei* has 4 *A. nidulans* RGS orthologues (RGSA – RGSD), and 3 class VI RGS domain-containing GPCRs (GPRK). GPRK-type GPCRs contain an RGS domain which is part of the same protein and located in the cytoplasm. This is the highest number of these RGS proteins found in any fungus screened to date (Schmoll, 2008). RGS proteins regulate G-protein signalling by enhancing GTP hydrolysis of the Gα subunit, thus increasing the rate of re-assembly with the Gβγ dimer and causing GPCR desensitisation (De Vries, L. and Farquhar, M. G., 1999).

The PTH11-type GPCRs are a category of G-protein coupled receptors which were originally discovered in *Magnaporthe grisea* (pathogenic fungus infecting crab grasses) and which are crucial for pathogenicity in this organism. Knock-out or mutation of *pth11* caused *M. grisea* to be non-pathogenic due to a defect in appressorium (narrow peg which exerts hydrostatic pressure and invades plant host cuticle and epidermal layer) differentiation (DeZwaan *et al.*, 1999). PTH11 is a 7 trans-membrane domain protein which is thought to sense surface cues such as soluble plant cutin monomers and induce appressorium differentiation through cAMP and diacylglycerol (discussed in section 4.1.4) signalling in the presence of the

right plant hosts. The PTH11 amino-terminal domain contains an EGF (epidermal growth factor)-like cysteine rich CFEM (conserved fungal specific extracellular membrane-spanning) domain which is predicted to be extracellular (Dean *et al.*, 2005). CFEM-domain-containing proteins are thought to interact with these receptors and are secreted and attached to the cell membrane through their GPI (glycosylphosphatidylinositol) anchor. PTH11-like receptors are unique to the fungal subphylum of Pezizomycotina within the Ascomycota and conserved domains include the seven transmembrane domains and the region towards the amino-terminal. Not all PTH11 receptors contain a CFEM domain: the genome of *M. grisea* encodes 61 PTH11-like receptors of which 12 have a CFEM domain and in *A. nidulans* there is 1 CFEM-PTH11 receptor out of 70 in total (Wilson and Talbot, 2009). The genome of *T. reesei* encodes 8 proteins containing a CFEM domain. Furthermore, the *M. grisea* fungal hydrophobin MPG1p is thought to mediate attachment of germlings to hydrophobic plant surface, acting as an additional cue for PTH11-mediated signalling and subsequent appressorium formation. Taken together, plant invasion by *M. grisea* is triggered by a CFEM-Pth11 GPCR which senses plant cutin monomer and interacts with MPG1 to mediate attachment of a spore to the hydrophobic plant surface.

4.1.4 G-protein signalling

Ligand binding to the GPCR induces conformational changes which allows binding of GTP to the Gα subunit. GTP binding alters the conformation of three "switch" regions in the Gα subunit that are the primary contact sites with Gβγ, favouring subunit dissociation (Dohlman and Thorners, 1997).

The Gα subunit phosphorylates and activates adenylyl cyclase which catalyses the formation of cAMP. Cyclic AMP activates a variety of glucose repressed functions such as catabolism of carbon sources by controlling the utilisation of endogenous and exogenous carbon sources and the regulation of glycoside hydrolase expression and secretion (Brunner *et al.*, 2008). It does so by activating protein kinase A (PKA) which phosphorylates serines and threonines of many target proteins. The other cAMP targets are MAP (mitogen activated protein) kinase pathways which regulate pheromone response, biocontrol, stress response, protein degradation and osmosensing through phosphorylation of target proteins. For example, in *T. reesei* MAP kinases target transcription factors such as CRE1 and HAP2 (of the Hap2/3/5 complex) which when phosphorylated, are able to bind to the promoters of *cbh1* and *cbh2* respectively (see Chapter 5). Another example of G-protein signalling involves the gene product encoded by *gph1*. A glycogen phosphorylase

(GH35) is expressed from this gene which is clustered on scaffold 2 with *gna3*, encoding a Gα subunit, and *tmk3*, encoding a MAP kinase. GPH1 is phosphorylated by PKA and its' expression is regulated by the TMK3 MAP kinase pathway (Schmoll *et al.*, 2009).

Furthermore, the Gα protein dimer activates the membrane bound phospholipase C (PLC) which catalyses the formation of diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3) from phosphatidylinositol 4,5-bisphosphate. IP_3 then diffuses through the cytosol and binds to calcium channels in the ER, increasing intracellular Ca^{2+} concentration. DAG activates protein kinase C which goes on to phosphorylate many other proteins. *T. reesei* possesses 5 PLCs which thus act as a regulator of growth, metabolism and protein secretion.

The main role of the Gβγ protein dimer is to inactivate the Gα subunit but it can also be involved in the regulation of various ion channels such as K^+ and Ca^{2+} channels. Ca^{2+} levels within the cell are important for signalling and they can be regulated by calcium channels, calcium pumps, calcium transporters and signalling proteins as a response to various external signals (e.g. polarised growth, see Chapter 3). One factor which responds to increased calcium concentrations within the cell is calmodulin which is activated through Ca^{2+} binding. This activated protein can now bind to other enzymes such as protein kinases and phosphoprotein phosphatases, changing their activity. In *T. reesei* calmodulin antagonists interfere with xylanase formation (Schmoll, 2008).

One of the best studied examples involving G-protein signalling in *T. reesei* and subsequent carbohydrate degradation is that light perception modulates cellulase gene expression in the presence of an inducer such as cellulose. The protein complex BLR1/BLR2 senses light and enhances transcription of target genes such as *env1*, encoding ENVOY, a PAS (Per, Arnt, Sim)/LOV (light, oxygen, voltage) domain protein which modulates and adapts the response to light through protein-protein interactions but also regulates various genes during growth in darkness (Schmoll, 2008; Castellanos *et al.*, 2010). In the presence of a carbon source, *gna3*, encoding a Gα subunit protein, is transcribed and activated upon ligand (polysaccharide derived inducer) binding to the GPCR. GNA3 and ENVOY increase cAMP levels within the cell through different mechanisms: GNA3 causes the activation of adenylyl cyclase whereas ENVOY inhibits a phosphodiesterase involved in cAMP degradation (Tisch *et al.*, 2011). The elevated levels of cAMP then activate a signalling cascade which results in cellulase, chitinase, glucanase, lipase and acid phosphatase gene expression (Do Nascimento Silva *et al.*, 2009). The exact details and intermediates of this cascade are not known yet but additional

supporting data exist: the expression of *cbh1* has been shown to increase in the presence of ENVOY and the envoy upstream motif 1 (EUM1) is found in the *cbh1* and *cbh2* promoters (Schmoll *et al.*, 2005).

4.1.5 Natural antisense transcripts (NATs)

NATs have been shown to occur in organisms from all three kingdoms of life (bacteria, archaea and eukaryotes) and can be associated to between 0.5% and 70% of all genes in an organism; their prevalence varies on the condition studied and the method employed to characterise them (Lapidot and Pilpel, 2006). NATs can be categorised according to their orientation and degree of overlap: head to head (5' to 5'), tail to tail (3' to 3') or fully overlapping (Georg, and Hess, 2011). Sense-antisense gene pairs are co-expressed or inversely expressed but can also have anti-correlated expression patterns, and they show high conservation throughout evolution. The antisense transcripts are usually present at lower levels than the sense mRNA in the cell (Ni *et al.*, 2010; Xu *et al.*, 2011). Furthermore, one of the transcripts can also be delayed in its transcription when compared to the transcript on the opposite strand; as a result there is a delayed shut down of the gene or the dampening of stochastic fluctuations in the level of the sense transcripts (Lapidot and Pilpel, 2006).

In *S. pombe*, transcriptional read-through and transcription of bidirectional promoters are two mechanisms by which antisense transcripts are generated and which spread to regulate adjacent genes. In *S. pombe*, ncRNAs (non-coding RNAs) have higher antisense readings than protein coding RNAs; this may hint at a role of these transcripts in chromatin remodelling and/or transcriptional gene silencing (Ni *et al.*, 2010). Furthermore, 47.5% of all *S. pombe* protein-encoding genes, and some *S. cerevisiae* genes, have associated antisense transcripts and they are pre-dominantly found at meiotic gene loci and stress response pathway loci. In *S. cerevisiae* antisense expression is associated with genes of larger expression variability, especially at the lower end of their expression scale (Xu *et al.*, 2011). This hints at a mechanism where antisense transcripts switch off low level sense expression and upon gene activation, sense RNA transcripts are increased and mediated antisense RNA inhibition is relaxed. In that way, RNAs may prevent leaky promoters during the vegetative stages of the yeast cell, fine tune gene expression or allow a quick response to stress. This hypothesis is supported by the *S. cerevisiae sur7* gene (unknown function), which is highly expressed in galactose and repressed upon α-factor pheromone stimulation (Xu *et al.*, 2011). The antisense transcript of *sur7*, named *sut719*, is highly expressed in

both conditions and disruption of this transcript increased *sur7* expression in the presence of the pheromone. An example of this is a bidirectional promoter in *S. cerevisiae* which is located between *sur7* and *gal80*; activation of the promoter by GAL4p and subsequent transcription of *gal80* in galactose also increases *sut719* transcript abundance. Thus it looks like gene regulation through antisense expression is, at least in yeast, condition-specific and affects non-housekeeping genes.

NATs operate through a variety of mechanisms and are a heterogenous group of regulatory RNAs. The mechanisms for NAT-mediated regulation of sense mRNAs can be broadly divided into four categories: 1) transcription-related, 2) RNA-DNA interactions, 3) RNA-RNA interactions in the nucleus and 4) RNA-RNA interactions in the cytoplasm and which can also activate the RNA interference (RNAi) pathway (Faghihi and Wahlestedt, 2010). The RNAi pathway is a much conserved mechanism amongst filamentous fungi serving as genomic defense, heterochromatin formation and gene regulation. The RNAi pathway in filamentous fungi has diverse functions and different pathways exist which generate short dsRNAs. In *S. cerevisiae* most of the RNAi components have been lost whereas in *S. pombe* RNAi mediates heterochromatin formation (transcriptional gene silencing), a phenomenon which has not been observed in filamentous fungi yet (mainly post-transcriptional gene silencing). It is noteworthy that siRNAs can be generated either endogenously (NATs, described above) or exogenously by genomic transposon insertions and viral infections. In both cases, the RNAi pathway is activated and results in the degradation of transposon RNA, viral RNA or any other mRNA (Figure 4.1). The RNAi machinery consists of three core components: Dicer, Argonaute and RNA-dependent RNA polymerase (RdRPs). RdRPs generate ds RNA from ssRNAs or they amplify ssRNA signals. Dicer, an RNA polymerase III domain containing enzyme, recognises double stranded RNA and cleaves it into siRNAs of 25 nucleotides. One of the RNA strands is then incorporated into RISC (RNA-induced silencing complex) whose core component is an Argonaute protein and which is guided to the target RNA sequence to degrade it. One of the strands of the dsRNA, loaded onto RISC, needs to be removed in order to activate the Argonaute protein subunit of this complex. This is usually carried out by the activity of another exonuclease. Experimental evidence elucidating RNAi pathways and the generation of siRNAs is practically non-existent in *Trichoderma* spp., although its genome encodes all the necessary components to activate the RNAi pathway: two Dicer-like proteins (*dcl1* and *dcl2*), one RdRP (*qde1* = quelling-deficient), one Argonaute protein (*qde2*), one RecQ helicase (*qde3*) and an Argonaute siRNA chaperone (ARC) complex subunit.

In the first NAT mediated mechanism, two RNA polymerase complexes, operating on the sense and anti-sense DNA strands at divergent or tandem transcribed gene regions, can "collide" (long-distance electrostatic interaction) in the overlapping region thus blocking further transcription (Georg and Hess, 2011). Transcriptional collision has been recorded in *S. cerevisiae* during the elongation step of the gene pair *gal10* and *gal7*. By manipulating the direction of these two genes (from tandem to convergent) and eliminating the intergenic region, Prescott and Proudfoot (2002) showed transcriptional collision at overlapping regions between the two genes, resulting in no mRNA accumulation and repression of both genes (the exact mechanism of the collision was not investigated).

The second role of NATs is to modify a genomic locus through chromatin modifications by increasing DNA methylation or by recruiting histone modifying enzymes thus causing gene silencing at the transcriptional level. In *S. pombe*, heterochromatin assembly at periocentromeric regions during cell cycle progression, is catalysed by CLR4 (methylase)-mediated histone 3 lysine 9 methylation (H3K9me); this provides binding sites for CHP1, CHP2 and SWI6 chromodomain proteins. CHP3 and SWI6 recruit various histone deacetylases (HDACs) which are associated with heterochromatin formation (discussed in Chapter 5). A RITS (RNA-induced initiation of transcriptional gene-silencing) complex is also recruited to genomic loci targeted for heterochromatin formation. The RITS complex is composed of CHP1, AGO1 (Argonaute protein with associated single-stranded siRNA) and RDP1 (RNA-dependent RNA polymerase). Presumably, transcription of repeat DNA sequences triggers the RNAi machinery; siRNAs generated by Dicer provide a platform for recruiting the RITS complex to a specific target site. RITS then recruits and interacts with CLR4 which catalyses H3K9me. This post-translational modification is recognised by CHP1 which binds to H3K9 and increases heterochromatin formation through hypoacetylation (Grewal, 2010).

In the nucleus, RNA duplex formation can result in several outcomes, all of which modulate sense mRNA expression; this can occur through either alternative splicing by masking splice sites and altering splice variants, or by providing a substrate for editing enzymes (e.g. deamination of adenosines to inosines) which ultimately alter the protein-coding mRNA localisation, transport and stability.

In the last mechanism, RNA duplex formation in the cytoplasm, affects mRNA stability, mRNA translation and serves as a template for endogenous siRNA formation. Binding of the antisense RNA to the mRNA reduces mRNA decay

and increases its stability or it can simply inhibit translation by preventing the ribosomal units to bind to the coding RNA strand (Georg and Hess, 2011).

Not all of the above mentioned mechanisms have an equally strong experimental support: convergent promoter activity and engagement of multiple RNAPII complexes in the opposite orientation of active promoters provide evidence against the transcriptional collision model (at least in higher eukaryotes) and increased levels of DNA methylation in chromatin modifications are also not well documented; it is more logical to assume that RNAs can recruit histone modifying enzymes as portrayed in the *S. pombe* example. RNA duplex formation with the above mentioned various outcomes have all been proved experimentally but they are likely to not be the predominant basis of NAT regulation. It is still unclear how co-expressed sense and antisense transcripts escape the RNAi pathway, what the criteria are for entering one of different regulatory pathways and with which other proteins they interact within the nucleus or the cytoplasm.
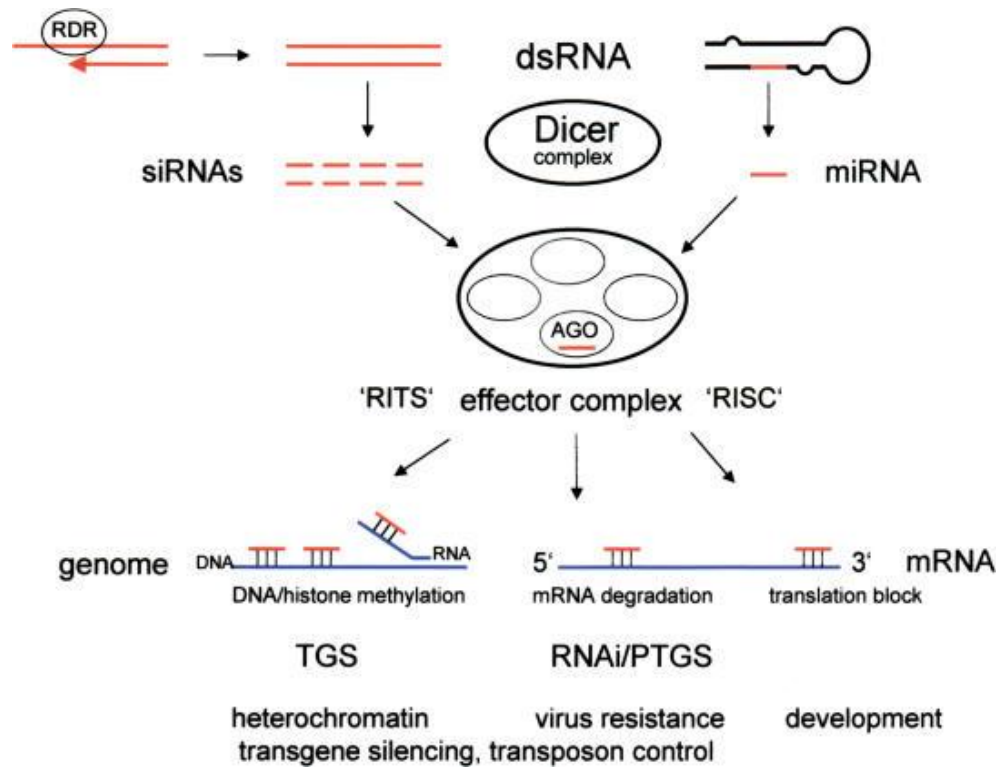
Figure 4.1: Diagrammatic representation of the core mechanism of RNAi. Endogenous or exogenous dsRNA and hairpin RNA of which some are generated by RDR (RNA-dependent RNA polymerase), are cleaved by dicer into 25 basepair RNA molecules, termed siRNAs and miRNAs and incorporated into RISC (RNA-induced silencing complex) or RITS (RNA-induced trancriptional silencing) which has an argonaute (AGO) protein as the catalytic unit. One of the RNA strands is removed and RISC or RITS is guided to its target mRNA, resulting in the degradation of the mRNA or in an increase of DNA/histone methylation (TGS = transcriptional gene silencing) or in the blocking of mRNA translation (PTGS = post-transcriptional gene silencing). Taken from Matzke and Matzke (2004).

4.1.6 Aims

The general aim of this chapter is to analyse the sequenced transcriptomes of *T. reesei* mycelia grown in a simple carbon source (glucose) for 48 h, transferred into a complex lignocellulosic substrate (wheat straw) for 24 h and with the addition of glucose to straw for 5 h, in order to:

- Find all the genes which are significantly differently expressed between all 3 conditions and gain information about them by categorising them into functional categories based on the proteins they encode

- Find genes, which are highly induced in the presence of straw and repressed in both glucose conditions, encoding CAZy (GHs, CEs and PLs) enzymes, accessory enzymes as well as non-CAZy enzymes. These genes encode proteins which may be involved in complex lignocellulose substrate deconstruction

- Describe the expression patterns of a few of these CAZy and non-CAZy-encoding genes and characterise the regulatory basis of these genes by investigating their expression patterns in a carbon catabolite-deficient strain

- Propose a model (based on the previous two aims) for lignocellulose degradation in *T. reesei* and compare it to the previously proposed model of carbohydrate degradation in *A. niger* in order to provide information on the conserved and diverged approaches used by saprobic fungi for lignocellulose degradation

- Prove the existence of NATs, a type of regulatory RNAs, in *T. reesei* and which have previously been described in *A. niger* and *N. crassa*

4.2 Materials and Methods

4.2.1 Experimental design

T. reesei QM6a conidia (see Chapter 2 section 2.3) were inoculated at a concentration of 10^5 spores/mL in 3 x 100 mL TMM complemented with 1% (w/v) glucose for 48 h. Glucose-grown mycelia from two flasks were rinsed 2 x with 100 mL ddH$_2$O and transferred into 2 x 100 mL TMM supplemented with 1% (w/v) cordiale straw (see Chapter 2 section 2.4) or no carbon source for 24 h. Then 1% (v/v) glucose was added for 5 h to one flask. All growth was carried out at 28°C, 150 rpm. After each step, mycelia were harvested, snap frozen in liquid N$_2$ and stored at -80°C. Each growth step was carried out in triplicate. See Chapter 2 sections 2.4 and 2.5 for growth details. RNA was extracted (as described in Chapter 2 section 2.7) from mycelia grown in triplicate from each of the three conditions described above and reverse transcribed into cDNA (Chapter 2 section 2.11).

4.2.2 RNA-sequencing of 48 h glucose, 24 h straw and 5 h glucose-grown cultures

Ribosomal RNA was degraded in 10 µg of total RNA using the Ribominus Eukaryotic kit (Invitrogen, Cat. No. A10837-08). SOLiD whole transcriptome libraries were made according to the SOLiD Whole transcriptome kit protocol (Applied Biosystems, Cat. No. 4425680) and the library concentrations were measured with the Quant-ut HS dsDNA assay kit (Invitrogen, Cat No. Q32851). Libraries were pooled to equimolar amounts (Invitrogen, Cat. No. Q32851) and gel purified using 2% size-select E-gels to 200-300 bp (Invitrogen). Emulsion PCR (0.5 M final concentration of pooled libraries) and bead-based enrichment was done according to the SOLiD 3+ template bead preparation guide containing library. Sequencing was carried out on a SOLiD 3+ ABi sequencer platform according to manufacturer's instructions to generate 50 bp reads in colour space.

4.2.3 Analysis of sequencing results

RNA-seq reads from SOLiD 4 from each of the experimental samples were mapped to the *T. reesei* genome (JGI version 2.0/FilteredModels16) using the BioScope 1.3.1 Transcriptome Pipeline (LifeTechnologies). Reads were first

filtered against published *T. reesei* rRNA sequences. Bioscope then mapped each read against the complete genome sequence and exon spanning junctions using available gene coordinate information, hence providing the primary read alignment position. Mapping results were recorded in a BAM (binary alignment/map) format.

The amount of reads per gene for each sample was calculated with Htseq-count (http://www.huber.embl.de/users/anders/HTSeq) using the BAM files and the genome annotation. Strand-specific RNA-seq reads, as generated by SOLiD, can be specified for when executing Htseq-count in determining accurate read counts per gene. These counts were used to calculate normalised expression values (reads per kilobase of exon model per million mapped reads = RPKM) for each gene (Mortazavi *et al.*, 2008). Antisense transcription was detected by either excluding or including strand-specificity in the calculations when comparing the Htseq-count generated counts. 3 independent statistical tests were carried on the number of reads counted per gene for each condition. These tests were the Likelihood Ratio Test (Marioni *et al.*, 2008), Fishers's Exact Test (Bloom *et al.*, 2009) and an MA-plot-based method with the Random Sampling method (Wang *et al.*, 2009). These tests allowed to find all the genes which are significantly differently expressed for all three conditions studied and all those genes scored a p-value of >0.001.

4.2.3.1 Acquiring individual gene information

Files containing the RPKM value per gene and the annotated genome from JGI were used to map gene I.D., annotation and expression value (using the average from triplicates) to each other using the Excel "lookup and reference" function. All genes were individually looked up in the JGI database to make sure that they were correctly annotated or to fill in missing gene information. Only genes which passed the three statistical tests were looked up, hence each gene now had an ID number, description and RPKM expression value between all three conditions studied. This permitted further analysis of individual genes. Each gene was also assigned to a functional KEGG group based on the classification used by JGI (http://genome.jgi-psf.org/cgi-bin/metapathways?db=Trire2).

4.2.3.2 Finding NATs

In order to visualise the sense and antisense reads, BAM files were put into the Integrative Genome Viewer (IGV 1.5) programme. IGV 1.5 (Broad Institute, Cambridge, Massachusetts, U.S.A.) and Python 2.7.2 (Python Software Foundation) programmes were downloaded together with manufacturer's instructions. The *T. reesei* filtered models genome (JGI) and

the annotated whole genome sequence (JGI) were down-loaded as .gzip files. Then the combined, counted raw reads (BAM file) from the triplicates per condition were mapped to the filtered genome. This was done on both the sense and antisense raw reads. The direction of all the small RNAs is pictured in IGV. The whole *T. reesei* genome was checked for genes with small RNAs which change in direction (sense/anti-sense or antisense/sense) between all three conditions studied.

4.2.4 Strand-specific PCR (ssPCR)

Four genes with NATs were chosen to be confirmed by reverse-transcriptase PCR. Primers were designed for the genes (sequence downloaded from JGI) and are given in Table 4.1. RNA was extracted (Chapter 2 section 2.7) from mycelia grown as described in 4.2.1 and reverse transcribed to cDNA (Chapter 2 section 2.11).

First of all, a PCR reaction (Table 4.1 for primer pairs) was run on cDNA from all three conditions and using gDNA (Chapter 2 section 2.6) as template.

Then single stranded PCR reactions (ssPCR) were carried out at an annealing temperature of 64°C. During the first reaction, each strand was amplified with one primer (Table 4.1). The second PCR reaction was run on the product obtained during the first reaction using the same primer pairs (Table 4.1). 30 cycles were run for each PCR reaction with Red Taq Polymerase (Chapter 2 section 2.8).

All PCR products were run on a 1% (w/v) agarose gel at 100 V for 1 h.

Table 4.1: Primer sequences, annealing temperatures and calculated product sizes of 4 genes analysed by RT-PCR.

Name	Forward sequence (5' to 3')	Annealing temperature (°C)	Reverse sequence (5' to 3')	Annealing temperature (°C)	Predicted Product size (bp)	Predicted Product size (bp) no intron
aa int	cccggctgattc cgacgaa	60.3	tggtgaggaca atggcccaga	59.7	326	263
βgluc int	aacggagccaa ggtcgggag	59.9	ctggacacctca tcatcggcaag	60.3	670	469
gh16 int	cgaggcggttcc ggtgg	59.4	gaaggtgcccgt cttggacg	59.0	344	259
tf1 int	gagcggcaacg gcaggg	59.7	gcattggtggct acgtcttcgag	59.6	299	240

4.2.5 RT-PCR and qRT-PCR

Mycelia were cultured as described in 4.2.1. For each condition, RNA was extracted from mycelia and reverse transcribed to cDNA (Chapter 2 section 2.11). RT-PCR (30 cycles) was performed on cDNA for several genes using the Red Taq Polymerase from Sigma (Chapter 2 section 2.8). Primer details for cellulase and hemicellulase-encoding genes are the same as described in Table 5.1 (Chapter 5) and primer details for other glycoside hydrolase-encoding genes are given in Table 4.2. PCR products were run on a 1% (w/v) agarose gel at 100 V for 1 h (Chapter 2 section 2.9).

QRT-PCR was performed on the same cDNA as described in the previous paragraph and *cbh1*, *xyn1*, *xyn2*, *gh47* and *gh67* genes were probed for in the three above described conditions. Reaction set-up, PCR details and analysis were exactly the same as described in Chapter 3 under 3.2.4.

4.2.6 RT-PCR and qRT-PCR of other genes of interest

Mycelia were grown in TMM with 1% glucose for 48 h and then transferred into TMM complemented with 1% cordiale wheat straw or no carbon source for 0.5 h, 1 h, 2 h, 3 h, 6 h, 9 h, 12 h and 24 h at 28°C, 150 rpm. Growth was carried out in duplicates. At each time point, mycelia were harvested, snap frozen and ground to a fine powder in liquid Nitrogen. RNA was extracted and reverse transcribed to cDNA (Chapter 2 section 2.11).

PCR reactions (30 cycles) were run on cDNA using the Red Taq Polymerase from Sigma (Chapter 2 section 2.8). Primer pair details are given in Table 4.2. PCR products were run on a 1% (w/v) agarose gel at 100 V for 1 h (Chapter 2 2.9).

QRT-PCR was performed on the same cDNA as described in the first paragraph and the following genes were probed for: *pth11*, *gh61*, *gh61a* and *gh61b*. QRT-PCR reaction set-up, reaction details and analysis of results were exactly the same as described in section 3.2.4 in Chapter 3.

Table 4.2: Primer sequences, annealing temperatures and calculated product sizes of 12 genes analysed by RT-PCR.

Name	Forward sequence (5' to 3')	Annealing temperature (°C)	Reverse sequence (5' to 3')	Annealing temperature (°C)	Predicted product size (bp)	Predicted product size (bp) no intron
hfb3	agttcctcgccgtc gccg	60.7	tgcagagaaga gccacgccg	61.0	423	282
swo1	cggcgagtgcatt gagctgat	59.8	cgagtcgggatg acgccatt	60.3	355	282
pth11	ccaggacactgca tcaacgtgaat	59.8	cttcaatggtag accaccaggcg	59.4	355	283
gprk	cgcctcctatgttc ccgcct	59.9	gcttgacttcgct ggacgacaagt	60.6	624	552
gh61	acggcaaccaacc cggtcaa	60.5	cgtcttggagtc gtagccgtcct	60.6	348	276
gh61a	agtagtgggctgg acggctgc	60.1	ccgagcctgag acggcaatg	60.2	521	467
gh61b	cctgcgctgggtca gattcag	60.5	cgagcctgtgac ggcgatgt	60.6	294	226
gh2	aaccatcactacc cgcctggga	60.2	cctgtggtgacg ttgaataccgg	60.0	380	380
gh47	gagaccatgtcga ggacgttcctg	60.7	aggagatatcg cgggtccttgg	60.1	437	437
gh67	actggatcgagac gtcgtttcgac	60.3	gcaccgtcgcag caccggt	60.7	327	327
gh76	ccagacaggtcgc tgccgtc	60.2	gatgcccgagcc gtcgtagac	60.4	273	273
gh92	cggcggcgacgac gccatgat	60.3	tcatacgacgca tcgaacccg	60.3	279	279

4.3 Results

4.3.1 Comparison of transcriptomes from 48 h glucose, 24 h straw and 5 h glucose

All genes which were significantly differently expressed between 48 h glucose, 24 h straw and 5 h glucose, were classed into categories according to the KEGG function of the proteins they encode (Appendix A2 Figures A.2.F.1 to A.2.F.3). There were 427 genes significantly differently regulated between 48 h glucose and 24 h straw; 725 genes between 24 h straw and 5 h glucose; and 165 genes between 48 h glucose and 5 h glucose. The biggest difference in gene expression occurred when switching from glucose to straw and straw to glucose and the highest number of genes which were differently regulated was between 24 h straw and 5 h glucose. There is not much difference between the two glucose conditions.

Most genes with a significant different expression between 48 h glucose and 24 h straw were genes encoding proteins of unknown function or general function prediction only; other genes encode proteins involved in carbohydrate and amino acid transport and metabolism, secondary metabolite synthesis and catabolism and energy production and conversion.

The addition of glucose to straw grown mycelia caused an increase in the number of genes involved in transcription, posttranslational modifications, protein turnover, signal transduction mechanisms, lipid and carbohydrate metabolism and transport and secondary metabolite biosynthesis and catabolism (Appendix A2 Figures A.2.F.1 and A.2.F.2). Again, the number of genes encoding proteins of unknown function or general function prediction remained high. Additional KEGG functions were also recorded between 24 h straw and 5 h glucose: genes encoding parts of the cytoskeleton and extracellular structures and genes involved in nucleolar transport and metabolism were all significantly differently expressed between 24 h straw and 5 h glucose but not between 48 h glucose and 24 h straw (Appendix A2 Figure A.2.F.2).

Most genes which were significantly differently regulated between 48 h and 5 h glucose were also genes encoding proteins of unknown function and general function prediction (Appendix A2 Figure A.2.F.3). Genes involved in energy production and conversion, transcription, secondary metabolite biosynthesis and catabolism, posttranslation modifications, protein turnover and carbohydrate transport and metabolism were also expressed significantly

differently. There was a lack of genes, which were differentially regulated, involved in cell cycle control and cell division, cytoskeleton and extracellular structures, nucleolar transport and metabolism when compared to the other 2 conditions.

4.3.2 Expression of GHs, CEs and PLs

One of the most numerous categories of genes which were significantly differently expressed between all three conditions studied were those encoding proteins involved in carbohydrate deconstruction. Figure 4.3.1 shows the percentage of reads corresponding to GHs, CEs and PLs for each of the three conditions studied here (see Appendix A3 Table A.3.T.1 for details). In straw there was a large increase in transcripts corresponding to CAZy-encoding genes, whereas this expression level was reduced in both glucose conditions. CAZyme transcripts made up 8.25% of the total RNA in *T. reesei* mycelia when incubated in straw-rich media.

79 GH-encoding genes and 7 CE-encoding genes were significantly differently expressed between 48 h glucose, 24 h straw and 5 h glucose. A list of GH-encoding genes together with their transcript I.Ds, annotations, expression values for all 3 conditions, the presence of a secretion signal and GH protein family classification is found in Appendix A3 Table A.3.T.2. These 79 GHs can be classed into 37 different families (Appendix A3 Table A.3.T.2). Members from glycoside hydrolase families 2, 3, 27 and 92 were the most numerous when grown on straw (Figure 4.3.2). CE-encoding genes from families 3, 4, 5, 15 and 16 were significantly differently expressed between the 3 conditions studied here.

Expression values varied between each gene in each condition: some genes were expressed very highly in glucose and their expression increased even more in straw whereas other genes were only expressed very highly in straw and some genes had a very low expression across all 3 conditions (Appendix A3 Tables A.3.T.1 and A.3.T.2). Expression of GH and CE-encoding genes followed a general trend: expression was always increased in straw and decreased in glucose. An example of this is shown in Figure 4.3.3.
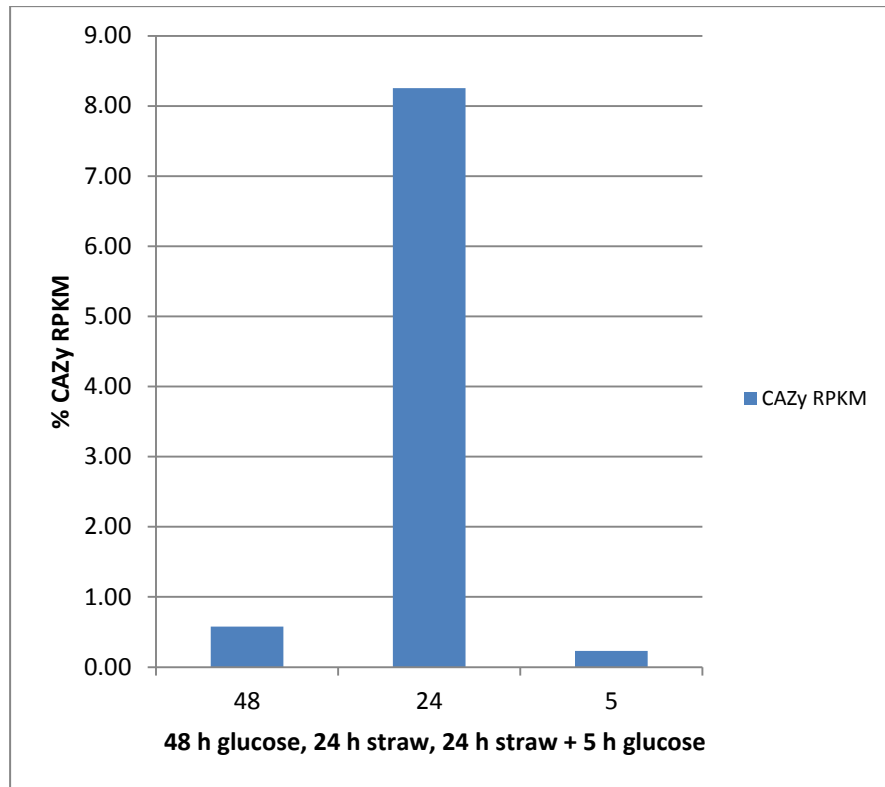
Figure 4.3.1: Percentage of transcripts corresponding to CAZy-encoding genes out of total cellular RNA in *T. reesei* in all three conditions studied.
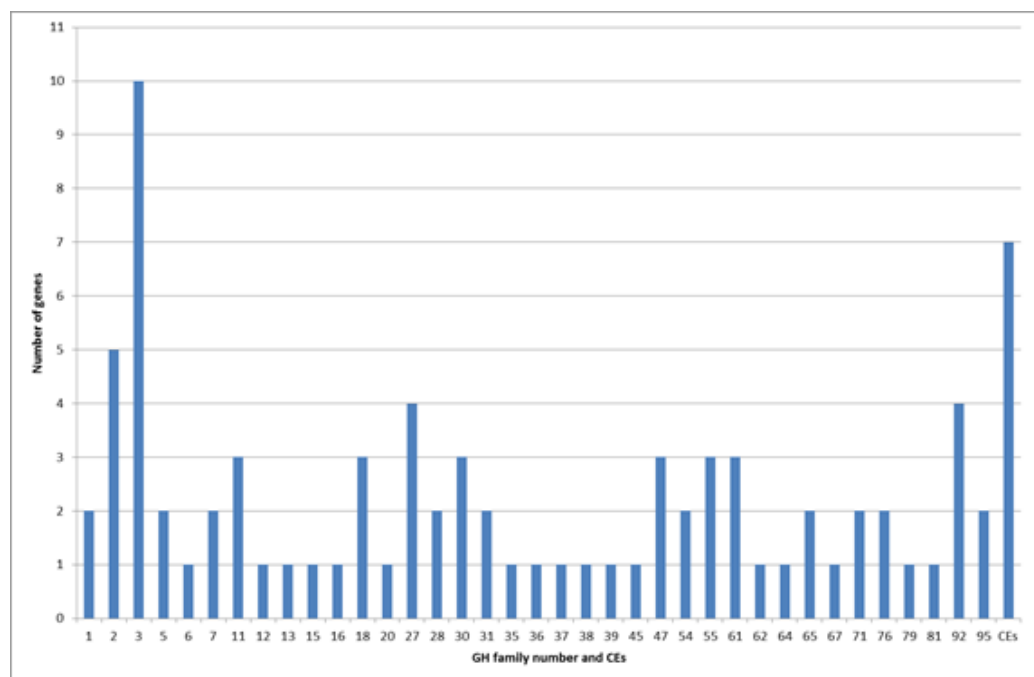
Figure 4.3.2: Number of CE-encoding genes (families 3, 4, 5, 15 and 16) and number of genes per GH family which were significantly differently regulated between 48 h glucose, 24 h straw and 24 h straw + 5 h glucose.
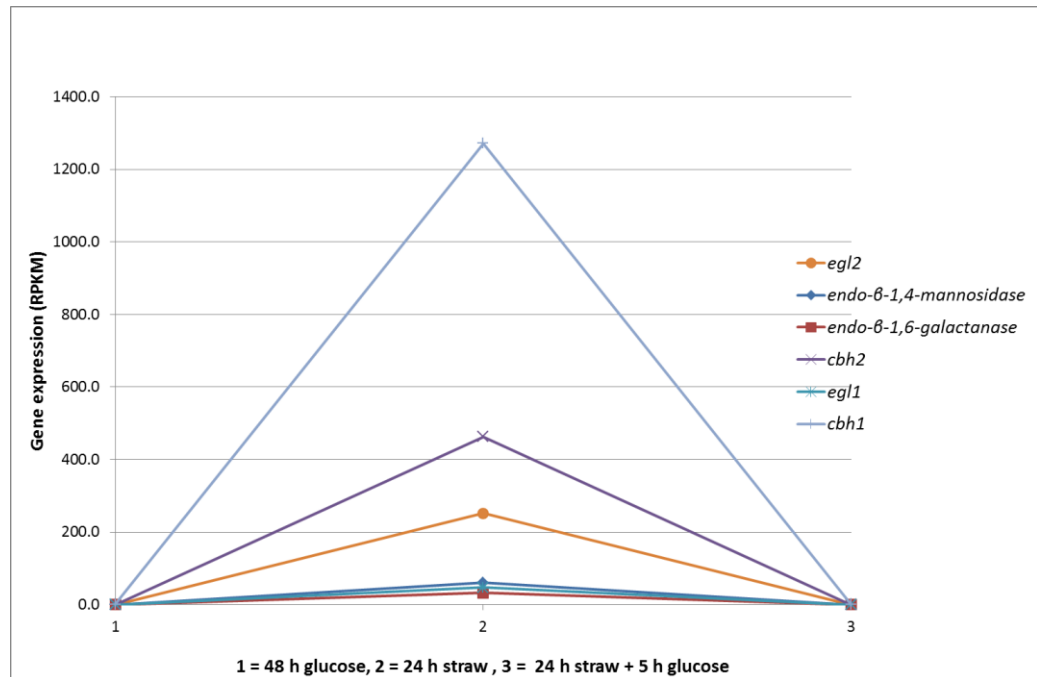
Figure 4.3.3: Expression patterns of genes encoding GHs from families 5 (*egl2*, endo-*β-1,4-mannosidase*, *endo-β-1,6-galactanase*), 6 (*cbh2*) and 7 (*egl1* and *cbh1*) between 48 h glucose, 24 straw and 24 h straw + 5 h glucose.

4.3.3 CAZY and non-CAZY gene expression in *T. reesei* QM6a

T. reesei also highly induced the transcription of genes encoding non-hydrolytic CAZymes involved in enhancing cellulose degradation. One such enzyme is the non-hydrolytic, expansin-like swollenin (*swo1*; out of five expansin-like-encoding genes; Appendix A5, Table A.5.T.2), thought to play a role in the loosening of the plant cell wall by disrupting hydrogen bonds between plant polysaccharides, thus increasing cell wall area and access of hydrolytic enzymes (such as cellulases) to the underlying polymers (Saloheimo *et al.*, 2002). CIP1, which contains a cellulose binding module (CBM) belonging to family 1 (Appendix A5, Table A.5.T.2), is also thought to enhance cellulose hydrolysis (Schuster and Schmoll, 2010). Transcript levels of *cip1* were detected in the presence of sophorose and regulation of this gene is the same as for other cellulases (e.g. *cbh1*), indicating a potential role for this protein in cellulose degradation (Foreman *et al.*, 2003).

Twenty-seven genes, which do not encode CAZymes, were also induced highly when switching from glucose to straw (Appendix A4 Table A.4.T.1). Some of these genes, such as surface interacting proteins (e.g. hydrophobins),

113

transporters (Appendix A4 Table A.4.T.1) and GPCRs may possibly be involved in carbohydrate deconstruction. Out of the 6 hydrophobin-encoding genes, HFB2, HFB3 and HFB5 were expressed significantly between glucose and straw; whereas HFB1 was down-regulated in straw and no changes occurred in the expression of the genes encoding the other two hydrophobins (Appendix A5 Table A.5.T.2). Furthermore, the genome of *T. reesei* encodes 35 Pth11-type GPCRs of which 7 increased in expression more than 10-fold when switching from 48 h glucose to 24 h straw (Appendix A5 Table A.5.T.3). The expression of some non-CAZyme and some CAZyme-encoding genes was further confirmed via RT-PCR and qRT-PCR.

4.3.3.1 Expression of glycoside hydrolase-encoding genes in *T.reesei* QM6a when grown in glucose for 48 h, transferred into straw or no carbon source (NC) media for 24 h and with the addition of glucose for 5 h

The expression levels of a few GHs were confirmed by RT-PCR (Figure 4.3.4) and qRT-PCR (Figure 4.3.5 and Figure 4.3.6) after 24 h incubation in straw-rich medium or under carbon starvation conditions (no carbon source, NC, -for 24 h). The main cellulase (*cbh1*, *cbh2*, *egl1* and *bgl1*) and hemicellulase (*xyn1*, *xyn2* and *bxl1*)-encoding genes were transcribed to high levels in straw but not in the absence of a carbon source. Some low levels of transcripts for these genes were present after 24 h incubation in NC but not to the same levels when incubated in straw (confirmed by qRT-PCR). This was also the case for genes encoding two GHs belonging to families 67 and 92.

A GH2 protein-encoding gene followed the same trend in straw as described above but in contrast to other GH encoding genes, the transcription pattern of *gh2* remained unchanged under carbon starvation conditions. The transcription pattern of two other genes, encoding GHs from families 47 and 76 did not change between 48 h glucose, 24 h straw and 5 h glucose. The transcription pattern of *gh47* remained the same under NC conditions whereas *gh76* transcript abundance was decreased in NC medium.

The expression patterns of some of the GH encoding genes were also confirmed by qRT-PCR: cellulase (*cbh1*) and hemicellulase (*xyn1* and *xyn2*) expression was highly induced in the presence of straw but not under carbon starvation conditions; this was also true for the gene encoding a GH67 protein. Transcript levels of a GH47 encoding gene were similar in straw-rich conditions when compared to carbon starvation conditions and a slight increase in transcript abundance at 24 h straw and 24 h NC was observed when compared to both glucose conditions. Transcript levels were also different between genes during 24 h incubation in straw: *xyn1* was the gene with the highest transcript levels, followed by *cbh1*, *xyn2*, *gh67* and then

114

gh47. The qRT-PCR data was qualitatively in agreement with the RT-PCR results but not with the RNA-seq results: RPKM data indicated that *xyn2* has the highest expression values followed by *cbh1* and then *xyn1*.
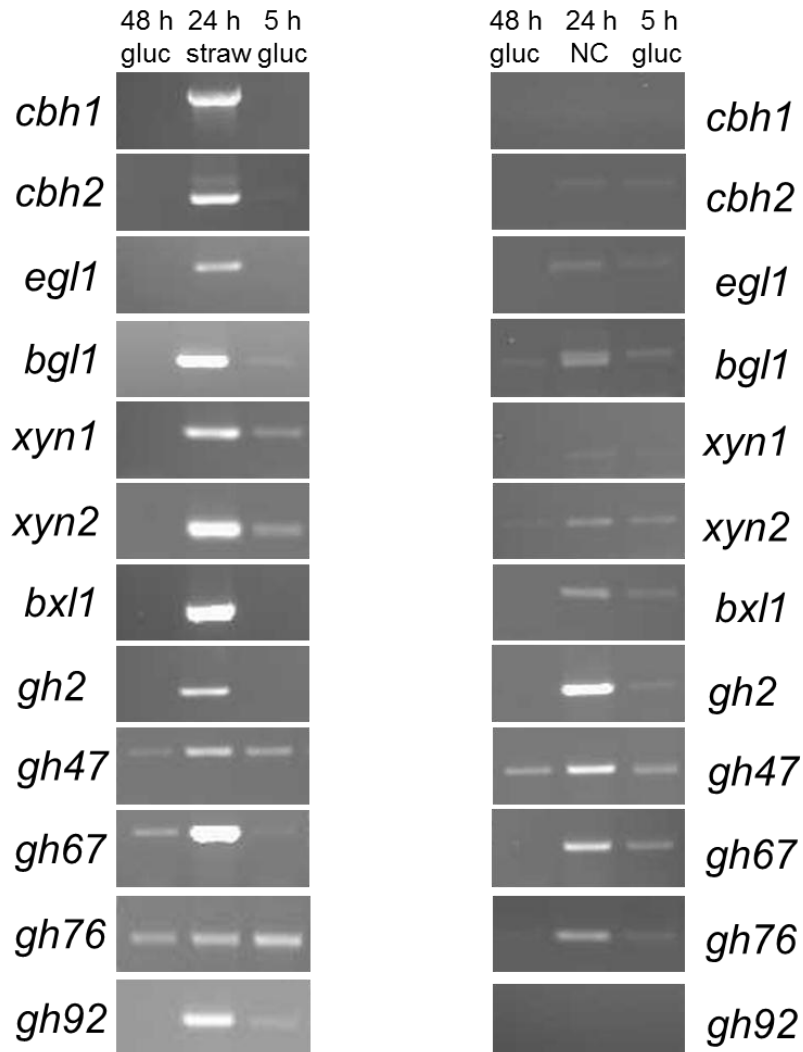
Figure 4.3.4: Products of RT-PCRs run on cDNA obtained from RNA extracted from mycelia grown for 48 h in glucose then transferred for 24 h into straw-rich or NC (no carbon source) media and with the addition of glucose to straw or NC media for 5 h.
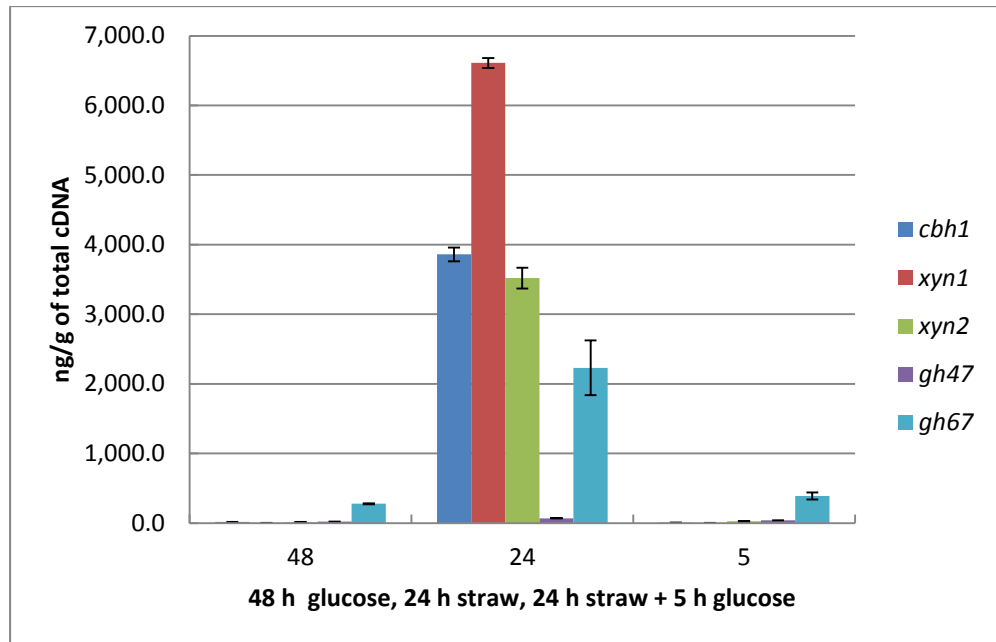
Figure 4.3.5: Expression of several glycoside hydrolase-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for 24 h into straw-rich media and with the addition of glucose to straw for 5 h. Error bars indicate the standard deviation for three replicates.
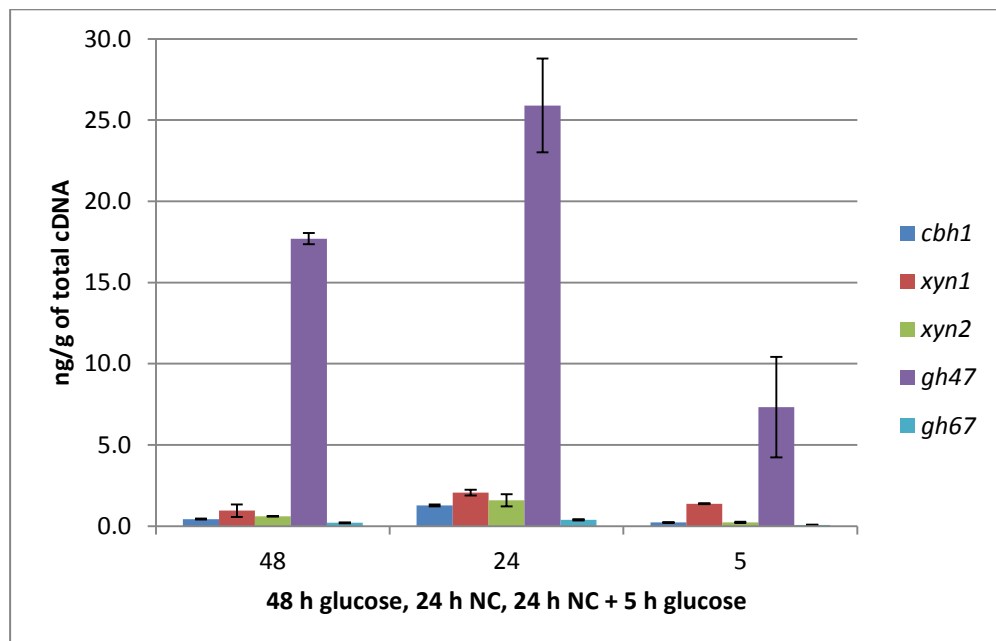
Figure 4.3.6: Expression of several glycoside hydrolase-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for 24 h into NC (no carbon source) media and with the addition of glucose to NC media for 5 h. Error bars indicate the standard deviation for three replicates.

4.3.3.2 Expression of CAZy and non-CAZy-encoding genes in mycelia grown for 48 h in glucose before being transferred for a 24 h time course in straw-rich or NC media

The expression of genes encoding some glycoside hydrolases and other proteins, thought to be involved in carbohydrate deconstruction [*hfb3*, *pth11*, *swo1* and *gprk* (G-protein coupled receptor kinase)], was monitored at different time points during a 24 h incubation period in straw and NC media (Figure 4.3.7). The expression of cellulase (*cbh1* and *bgl1*) and xylanase (*xyn1* and *xyn2*)-encoding genes started after 3 h to 6 h and they were fully induced after 24 h when incubated in straw. Transcript levels of *bgl1* were present at all time points including in glucose medium. Under carbon starvation, cellulolytic and xylanolytic genes were barely induced and weak transcript bands became apparent at around 9 h. Again, *bgl1* transcripts were present at all time points tested, although they were weaker than when incubated in straw. Expression patterns of three GH61 encoding genes were also different between straw and NC conditions. Transcripts of *gh61* were detected at all time points and increased in intensity at 3 h (intensity was the same during all subsequent time points); *gh61a* expression was detected at 6 h and increased in intensity all the way through to time point 24 h, whereas *gh61b* transcript levels were detected at all time points during incubation in straw and their intensity did not seem to change much. Under carbon starvation conditions, expression pattern of *gh61* was similar in straw whereas *gh61a* transcript levels were detected at all time points and *gh61b* transcripts were only detected at 6 h, 9 h, 12 h and 24 h with an increase in intensity at 24 h (Figure 4.3.7).

Transcript levels of a hydrophobin encoding gene, *hfb3*, were detected after 2 h incubation in straw and at 0.5 h to 1 h under carbon starvation conditions. The swollenin-encoding gene, *swo1*, was transcribed during all conditions studied here although the intensities of the transcript bands were lower under NC conditions than when compared to incubation in straw media. Expression patterns of the two GPCR encoding genes, *gprk* and *pth11* were similar in straw and NC media: *pth11* transcripts were detected during all time points studied here whereas *gprk* transcription occurred after 3 h in straw and 2 h under carbon starvation. The intensity of the *pth11* transcript bands was higher in straw than in media with no carbon source (also confirmed by qRT-PCR) whereas the intensity of the *gprk* bands did not change between conditions and time points.

The expression of some of these genes was also monitored by qRT-PCR (Figures 4.3.8 and 4.3.9) and was qualitatively in agreement with the RT-PCR

and the RNA-seq data: *pth11* transcripts were detected in straw and under carbon starvation, with levels being higher in straw. Levels of *gh61* transcripts were similar in both conditions whereas those of *gh61a* increased in the presence of straw and were similar in media with no carbon source. Transcript levels of *gh61b* were generally higher in straw than in NC media and they increased at the later time points in both conditions. Some of the RT-PCR products showed multiple bands for one condition (e.g. *pth11*, *swo1*) and this could be due to different splice variants for this gene: the mapped reads for this gene, when put into IGV (see materials and methods), showed potential splicing sites (Appendix A7 Figure A.7.F.1).
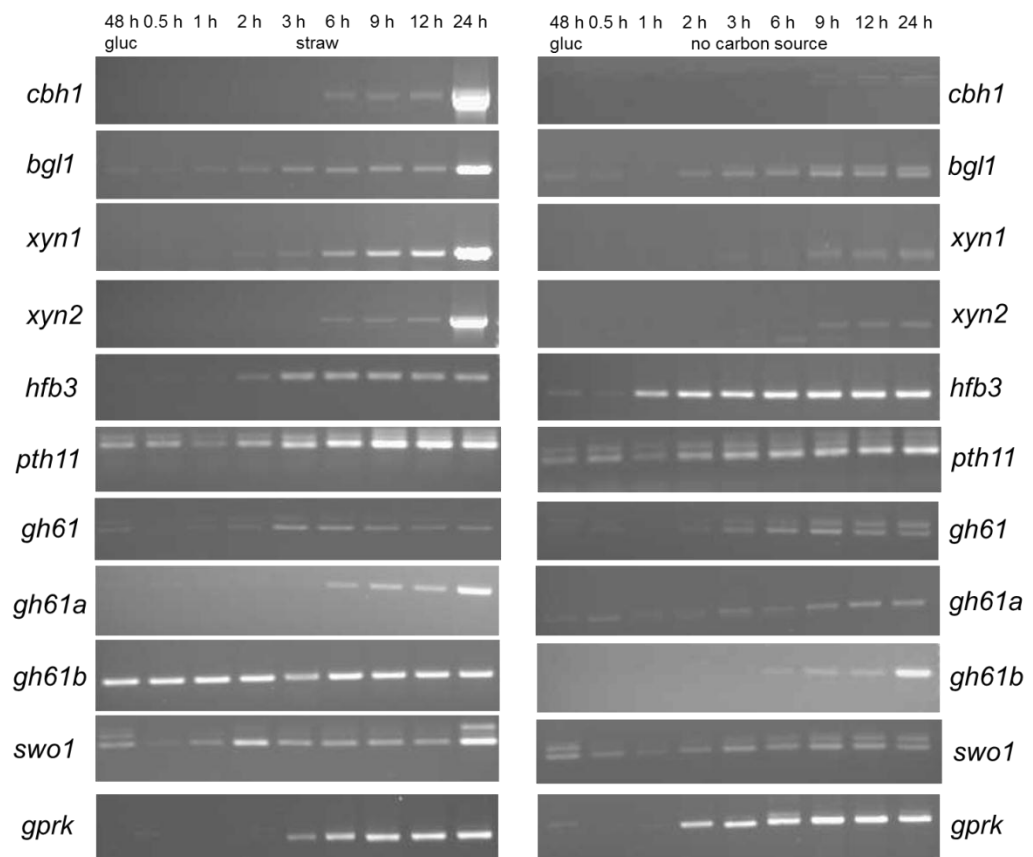
Figure 4.3.7: Products of RT-PCRs run on cDNA obtained from RNA extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into straw-rich or NC (no carbon source) media.
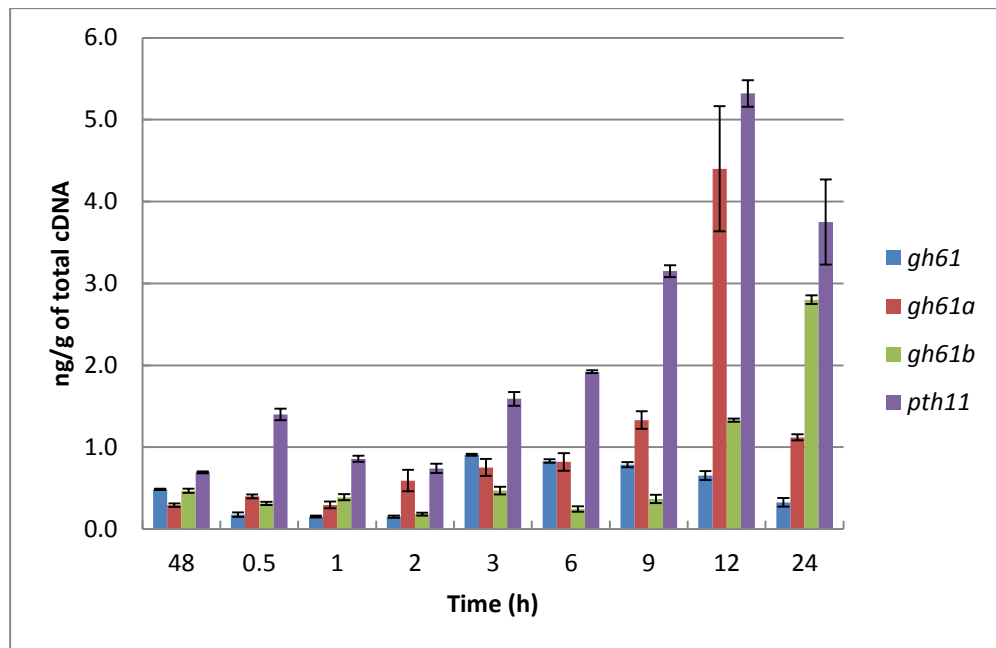
Figure 4.3.8: Expression of a few CAZy and non-CAZy-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into straw-rich media. Error bars indicate the standard deviation for three replicates.
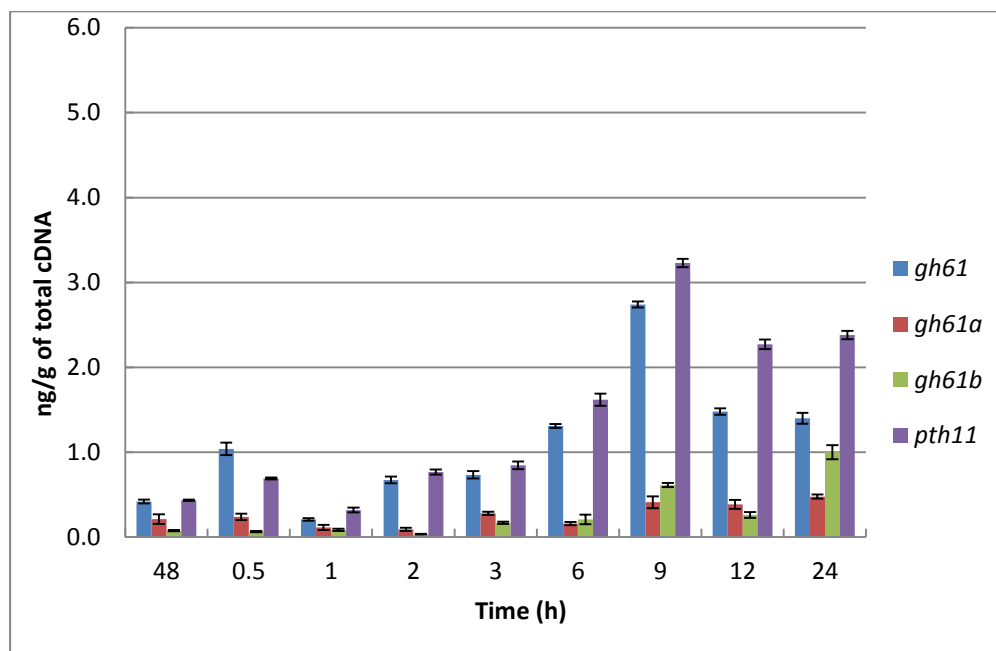
Figure 4.3.9: Expression of a few CAZy and non-CAZy-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into NC (no carbon source) media. Error bars indicate the standard deviation for three replicates.

4.3.4 CAZy and non-CAZy gene expression in *T. reesei Δcre1*

4.3.4.1 Expression of glycoside hydrolase-encoding genes in *T.reesei Δcre1* when grown in glucose for 48 h, transferred into straw or NC media for 24 h and with the addition of glucose for 5 h

In order to investigate the regulatory basis of the same glycoside hydrolase-encoding genes described in section 4.3.3, gene expression was also monitored in the same conditions in a mutant strain carrying a *cre1* gene deletion (Figure 4.3.10). The pattern of gene expression in this strain was essentially the same as that observed in the wild-type strain (compare Figures 4.3.10 and 4.3.4): cellulase and hemicellulase-encoding genes were expressed in straw and repressed in glucose whereas under carbon starvation conditions they were not induced. Repression by glucose was also stronger for some genes (e.g. *cbh1*) than others (e.g. *xyn2*). Some transcript abundance was observed for *bgl1*, *bxl1* and *xyn2* after 24 h in NC media but the levels were much lower than those observed in straw and resembled transcript abundance seen for glucose-rich conditions. The same trend was observed for the gene encoding a *gh67*: transcript abundance was high in straw and low in glucose whereas under carbon starvation conditions expression of this gene was much lower than when compared to straw.

The expression of two genes encoding a GH2 family protein and a GH47 family protein was observed in straw and NC conditions and hence differed from the one observed for the cellulase and hemicellulase-encoding genes. Glucose-mediated repression also seemed to occur for *gh2* and only slightly for *gh47*; this was also observed in the wild-type strain.

Transcript abundance of a gene encoding a GH76 was lower in 24 h straw medium than under NC conditions whereas transcript abundance was similar in all glucose conditions studied here. Transcript abundance of a GH92-encoding gene was not observed in straw conditions, carbon starvation conditions or any glucose conditions in the Δ*cre1* strain. These two genes (*gh76* and *gh92*) differed in their expression pattern between the wild-type and the Δ*cre1* strains: *gh92* was induced in straw conditions in QM6a but not in the *cre1* knock-out strain whereas transcript levels of *gh76* were much lower in the Δ*cre1* strain than in the wild-type strain in the presence of straw.

The expression patterns of some of these genes were also confirmed by qRT-PCR (Figure 4.3.11). As in the wild-type strain, transcript levels of *cbh1*, *xyn1*, *xyn2* and *gh67* increased in straw conditions and decreased in both glucose conditions; transcript levels of these genes were much lower in QM6a Δ*cre1* when compared to wild-type QM6a when these strains were cultivated in

straw conditions (compare Figures 4.3.11 and 4.3.5). This was also confirmed by the RT-PCR data where the intensity of the transcript bands was less in the mutant strain than in the wild-type strain (Figures 4.3.10 and 4.3.4 for *xyn1* and *xyn2*, Figures 4.3.16 and 4.3.10 for *cbh1*). Under carbon starvation, expression levels of these glycoside hydrolase-encoding genes were very low when compared to the ones in straw conditions, and were similar to those observed in the wild type strain, although *xyn2* transcript levels were higher in the mutant strain (compare Figures 4.3.12 and 4.3.6). Transcript levels of *gh47* were similar in straw conditions when compared to carbon starvation conditions. This was also observed in the wild-type strain and *gh47* transcript levels were similar under carbon starvation conditions in both strains. QRT-PCR and RT-PCR data were qualitatively in agreement with each other.
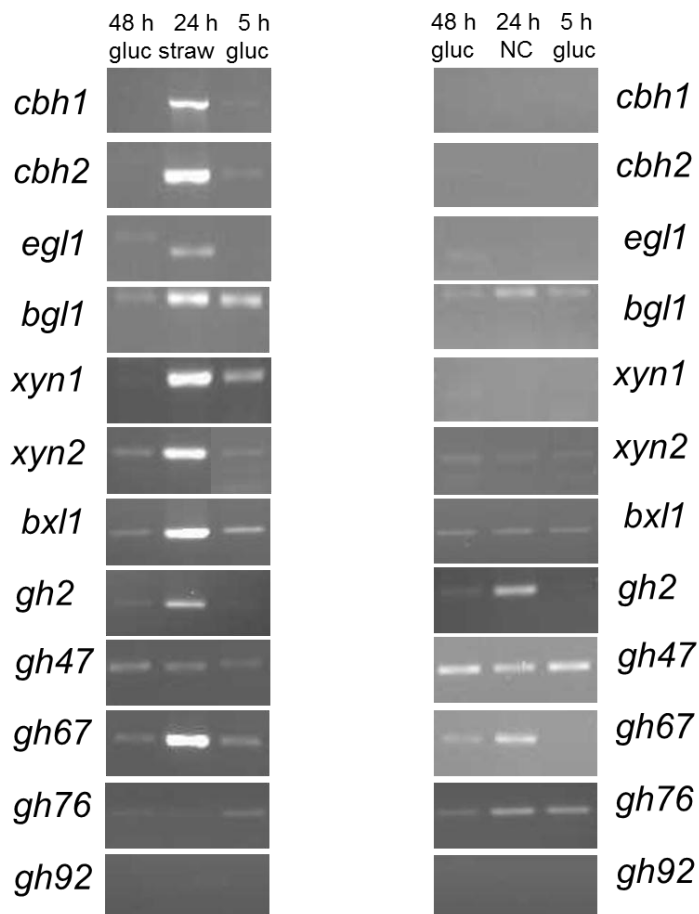
Figure 4.3.10: Products of RT-PCRs run on cDNA obtained from RNA extracted from mycelia grown for 48 h in glucose then transferred for 24 h into straw-rich or NC (no carbon source) media and with the addition of glucose to straw or NC media for 5 h.
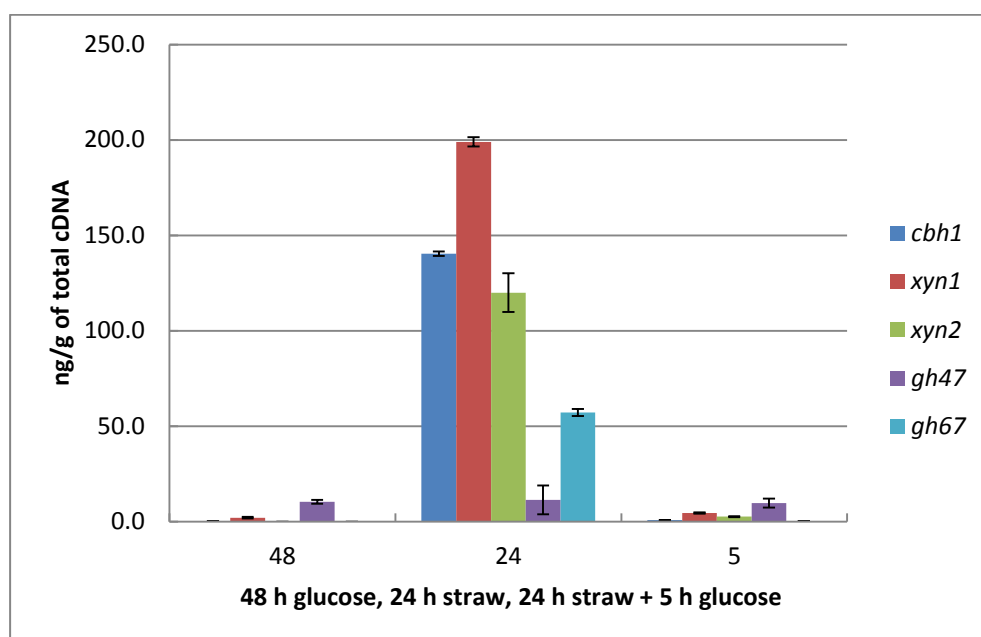
Figure 4.3.11: Expression of several glycoside hydrolase-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for 24 h into straw-rich media and with the addition of glucose to straw for 5 h. Error bars indicate the standard deviation for three replicates.
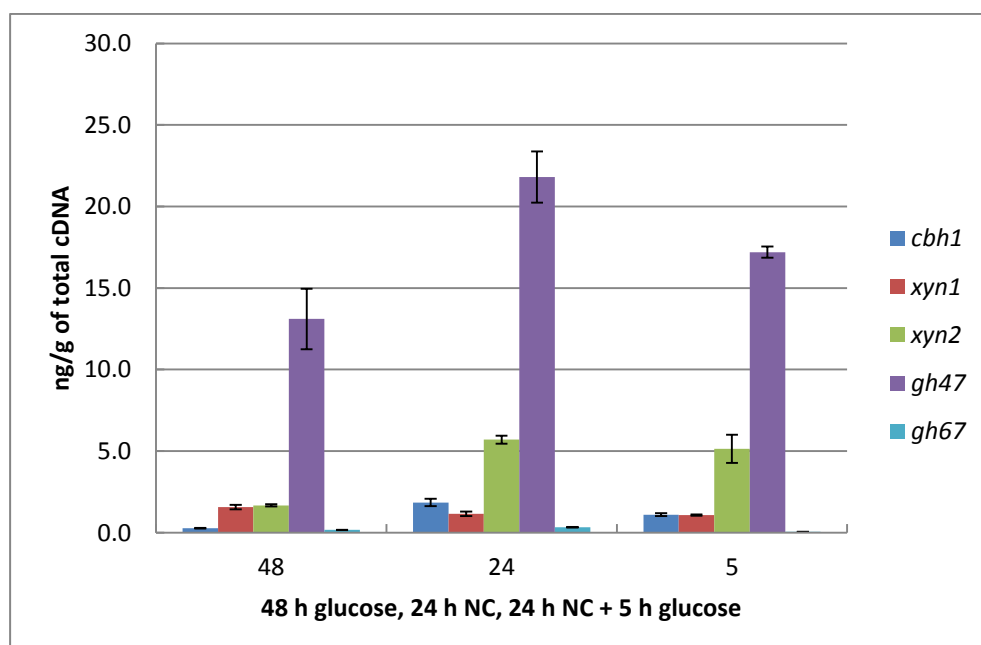
Figure 4.3.12: Expression of several glycoside hydrolase-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for 24 h into NC (no carbon

source) media and with the addition of glucose to NC media for 5 h. Error bars indicate the standard deviation for three replicates.

<u>4.3.4.2 Expression of CAZy and non-CAZy-encoding genes in mycelia grown for 48 h in glucose before being transferred for a 24 h time course in straw-rich or NC media</u>

The expression of CAZy and non-CAZy encoding genes was also studied in the Δ*cre1* strain when cultivated in straw or in no carbon source media for a 24 h time course (same conditions as described in 4.3.3.2; Figure 4.3.13). Transcript patterns of all of these genes were different to the ones seen in the wild-type strain. Expression bands of *cbh1* were visible after 3 h in straw and became stronger at 12 h and 24 h. Transcript bands of *xyn1* and *xyn2* were visible after 1 h and were already strongest after 6 h of incubation in straw. In the wild-type strain, *cbh1*, *xyn1* and *xyn2* transcript bands were strongest after 24 h cultivation in straw. Under NC conditions no expression of *cbh1* was recorded, whereas very faint bands of *xyn1* and *xyn2* transcripts were present at most time points. This was again in contrast to the wild-type strain where faint transcript bands of these genes were visible at the later time points. Transcript bands of *bgl1* were present at each of the time points in both conditions in the Δ*cre1* strain. This was in contrast to QM6a where *bgl1* expression was strongest after 24 h incubation in straw but transcript patterns were the same in both strains under carbon starvation conditions.

Expression patterns of the genes encoding GH61 proteins were also different between straw and NC media in the Δ*cre1* strain. Transcript bands of *gh61* were similar at each of the time points in the two conditions and also when compared to the wild-type expression pattern. Transcript bands of *gh61a* were present at each time point in straw and NC media and became stronger with increased incubation time in straw and weaker at the later time points under carbon starvation conditions. This was in contrast to the wild-type strain where *gh61a* transcripts were detected after 6 h of incubation in straw and at each time point under carbon starvation conditions. Weak expression of *gh61b* occurred after 24 h incubation in straw and at each time point in medium with no carbon source. In the wild-type strain, *gh61b* transcript bands were very strong at each time point when incubated in straw; whereas under carbon starvation conditions, *gh61b* expression bands were only visible after 6 h and increased in intensity at the subsequent time points.

The expression patterns of *hfb3*, *pth11* and *gprk* in the Δ*cre1* strain were the same in straw and in NC media: transcript bands were present in both conditions and at all the time points studied here. Expression of *swo1* became stronger with increased incubation time in straw and weaker at the later time

points under carbon starvation conditions. In the wild-type strain, expression pattern of *pth11* was the same as in the mutant strain whereas transcript bands of *hfb3* and *gprk* were only visible after 1 h to 3 h of incubation in either straw or NC media. Transcript bands of *swo1* were also present at all the time points in both conditions in QM6a and similar to the ones seen in the mutant strain.

Expression patterns of some of these genes (*gh61*, *gh61a*, *gh61b* and *pth11*) were also confirmed by qRT-PCR and results obtained from qRT-PCR were qualitatively in agreement with those obtained by RT-PCR (Figures 4.3.14 and 4.3.15). Expression levels and patterns of *pth11* were similar in both conditions with slightly higher transcript levels in straw; this was also similar to the levels observed in the wild-type strain. Expression of *gh61* was similar in straw and NC media whereas *gh61a* transcript levels increased in straw and decreased under carbon starvation conditions. Transcript levels of *gh61b* were highest at 24 h when incubated in straw whereas in NC medium they were high at all the time points studied here. QRT-PCR expression patterns of these genes differed from the ones obtained for the wild-type strain and these differences (described in the previous paragraphs) were in agreement with all the RT-PCR results obtained. See section 4.3.3.2 for comments on the appearance of multiple PCR bands for one gene.
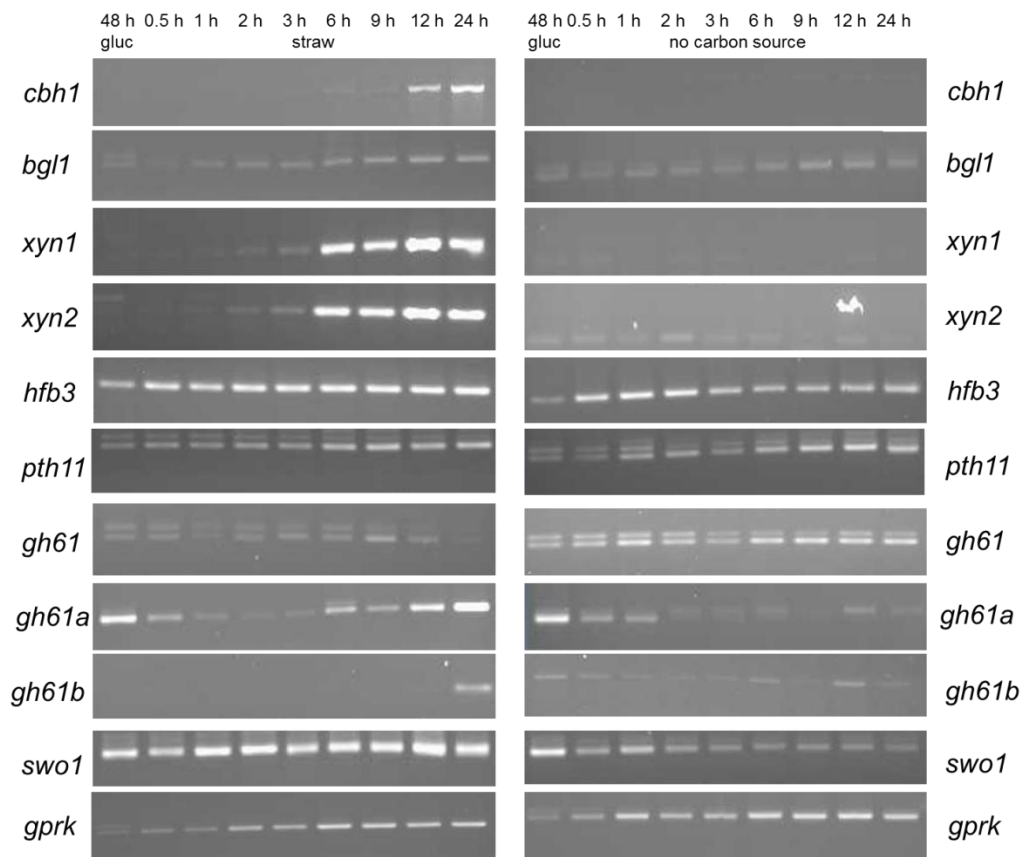
Figure 4.3.13: Products of RT-PCRs run on cDNA obtained from RNA extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into straw-rich or NC (no carbon source) media.
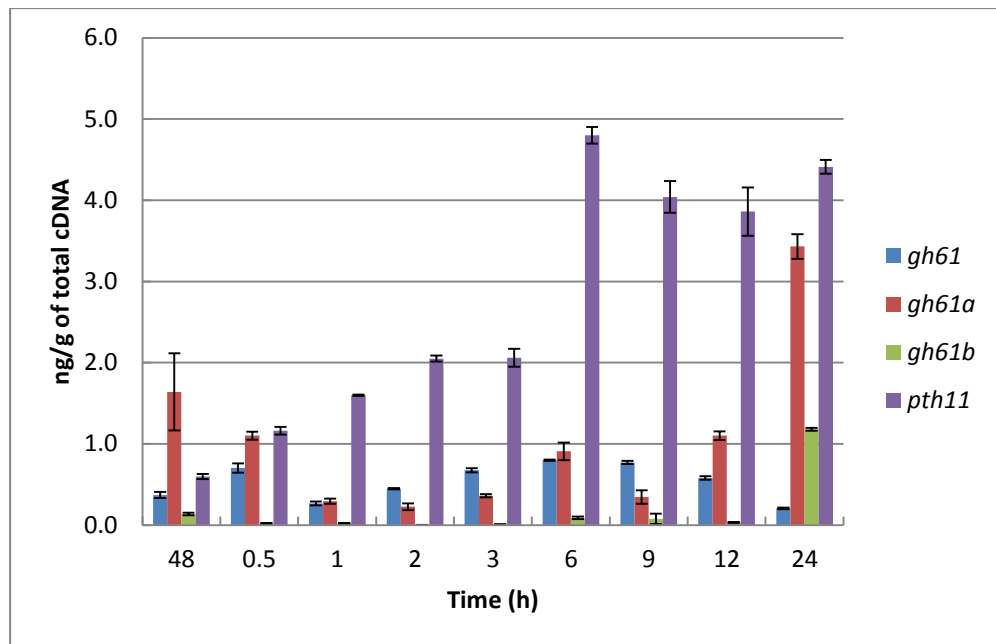
Figure 4.3.14: Expression of a few CAZy and non-CAZy-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into straw-rich media. Error bars indicate the standard deviation for three replicates.
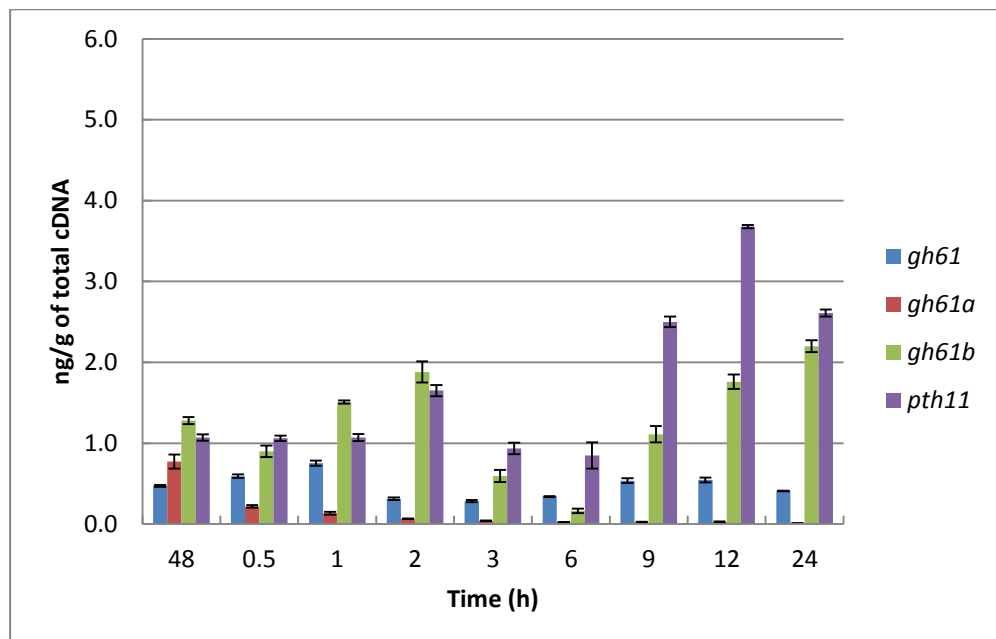
Figure 4.3.15: Expression of a few CAZy and non-CAZy-encoding genes monitored by qRT-PCR on cDNA obtained from RNA which was extracted from mycelia grown for 48 h in glucose then transferred for a 24 h time course into NC (no carbon source) media. Error bars indicate the standard deviation for three replicates.

4.3.5 Comparison of *A. niger* and *T. reesei* transcriptomes

<u>4.3.5.1 CAZy gene expression</u>

The genome of *T. reesei* encodes 201 GHs, 22 CEs and 5 PLs whereas the genome of *A. niger* encodes 246 GHs, 25 CEs and 8 PLs. There are differences between the two genomes in the families of enzymes they encode: CE families 8 and 12 are present in *A. niger* but not in *T. reesei*, whereas CE families 14 and 15 are encoded by *T. reesei* but not by *A. niger*; members from GH families 26, 29, 32, 51, 53, 88 and 114 are found in *A. niger* but not in *T. reesei* and GH families 24, 25, 39, 45, 64, 89 and 115 are found in *T. reesei* but not in *A. niger*. Both genomes encode completely different families of PLs: PLs 7, 8 and 20 are present in *T. reesei* whereas PLs 1 and 4 are present in *A. niger*. The genome of *A. niger* also does not encode expansin-like proteins (e.g. SWO1) or CIP1.

The percentage of RNA transcripts corresponding to CAZymes increased majorly in the presence of straw when compared to both glucose conditions and was much higher in *A. niger* (~19%) than in *T. reesei* (~8.3%) (Figure 4.3.16). The GH families with the highest expression levels after 24 h incubation in straw were very similar in both organisms and included glycoside hydrolase families 3, 6, 7, 11, 61 and 67 (Figure 4.3.17). There were differences within the transcription of CE-encoding genes for both organisms: whereas transcripts from CE family 1 (contains acetyl xylan esterases and feruloyl esterases) were most abundant in *A. niger* (~10% of total CAZy mRNA); transcript levels of acetyl xylan esterases of CE family 5 were highest in *T. reesei* (~5% of total CAZy mRNA). The genome of *T. reesei* encodes 3 acetyl xylan esterases and 1 cutinase, all belonging to CE family 5. The highest expression values are recorded for 2 acetyl xylan esterases (Transcript I.D.s 73632 and 54219) and one of these enzymes also contains a CBM1 module (73632). Acetyl xylan esterases remove acetyl groups at *O*-2 and *O*-3 positions of the xylose chain in arabinoglucuronoxylans, a process which has been shown to significantly enhance subsequent xylan and cellulose hydrolysis (Zhang *et al*., 2012). The genome of *A. niger* encodes 3 CE family 1 members: one acetyl xylan esterase, one feruloyl esterase and one unknown esterase (Dodd and Cann, 2009). Expression values of the acetyl xylan esterase and the feruloyl esterase were very high in the presence of straw (Delmas *et al*., 2012). Feruloyl esterases cleave ferulic acid groups which are esterified to the 5'-OH of arabinofuranosyl groups (arabinose residues linked to *O*-2 or *O*-3 of xylose) and which can be covalently linked to lignin or other ferulic acid groups in xylans (Dodd and Cann, 2009). The enzyme mix secreted by *A. niger* aids in loosening the lignin-hemicellulose structure in addition to de-

acetylating the xylan backbone in order to allow access of other CAZymes to the underlying hemicellulose and cellulose polysaccharides. Transcript abundance for PL-encoding genes was very low in both fungi (~0.01% of total CAZy mRNA in *T. reesei* and ~0.5% of total CAZy mRNA in *A. niger*) indicating that PLs are not as important as GHs and CEs for wheat straw degradation. Thus the bulk of GH enzymes used to degrade straw are from the same GH families in both organisms, whereas different CE and to a certain extent PL family members are used suggesting that both fungi specialised also in the cleavage of different bonds found within plant cell walls (www.cazy.org).

After 48 h growth in glucose, CAZy gene mRNA represented 0.58% of total RNA in *T. reesei* (in *A. niger* 3%), with members from GH families 16, 18 and 72 (glucanases, chitinases and glucanosyltransferases) representing approximately half (45%) of the total CAZy mRNA. In *A. niger* the glucoamylase *glaA* accounted for over 65% of total CAZy mRNA (Delmas *et al*., 2012). Thus in *T. reesei*, low levels of mRNA of genes encoding enzymes involved in complex carbohydrate, hemicellulose and chitin degradation are present when the fungus is cultivated in glucose-based media. In this media these enzymes are mainly involved in cell wall remodelling during hyphal extension as high growth rates are achieved in the presence of glucose in *T. reesei* (Fuji *et al*., 2010). In *A. niger* on the other hand, the majority of mRNA results from a gene encoding and enzyme involved in starch degradation in the presence of glucose. Addition of glucose to straw exerted strong carbon catabolite repression of the CAZy-encoding genes, and CAZy transcript abundance decreased to 0.23% of the total mRNA (in *A. niger* 1.5%) out of the total cellular mRNA with members from GH families 16, 18 and 72 (and GH15 in *A. niger*) becoming again the most expressed CAZy genes under this condition.
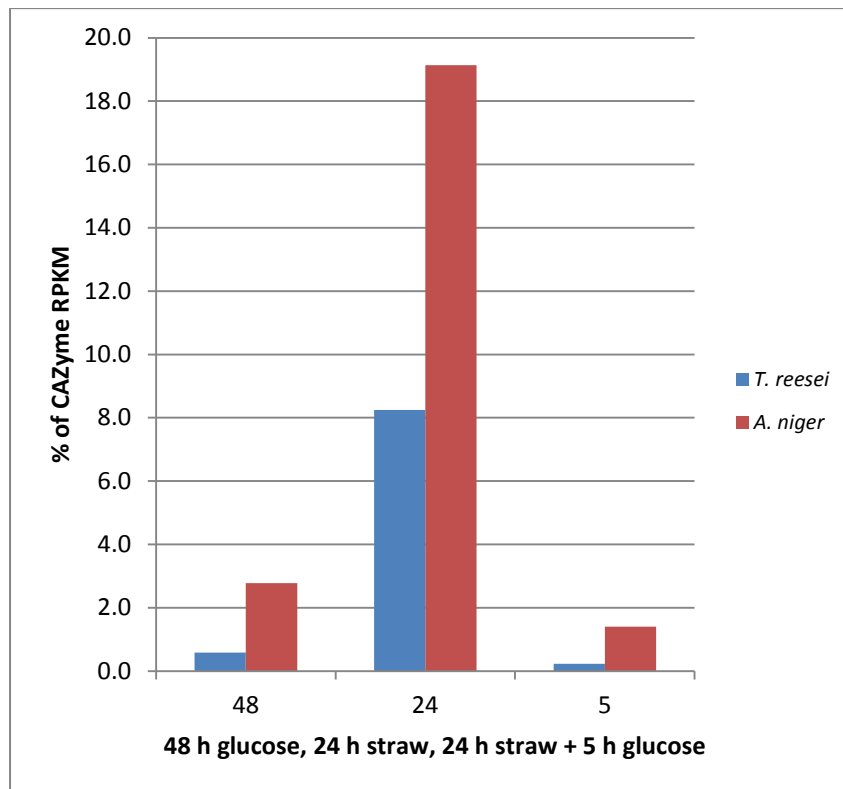
Figure 4.3.16: Comparison of the percentage of RNA transcripts corresponding to glycoside hydrolases, polysaccharide lyases and carbohydrate esterases in *T. reesei* and *A. niger* between 48 h glucose, 24 h straw and 24 h straw + 5 h glucose.
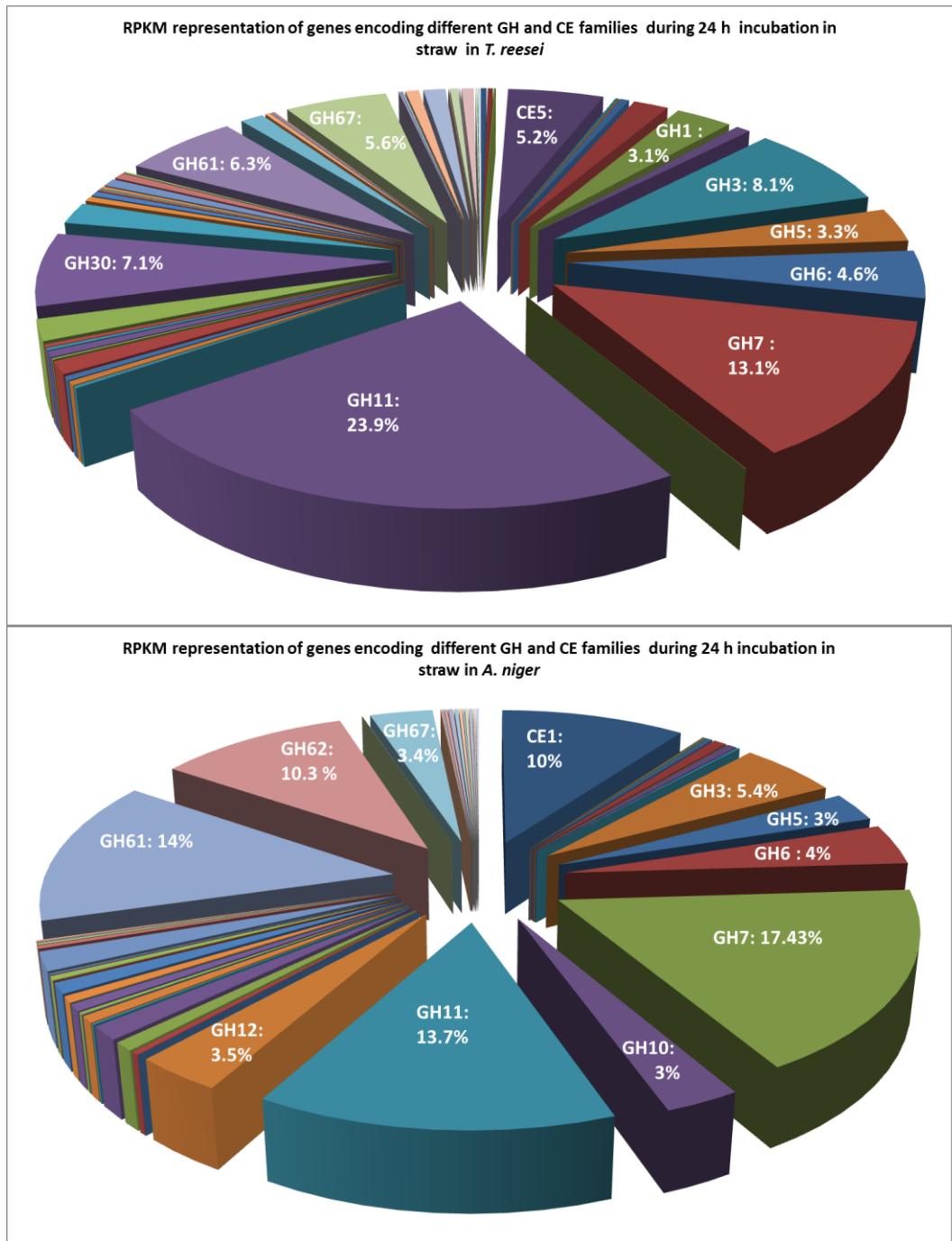
Figure 4.3.17: The relative amounts of the pre-dominant GH and CE families during 24 h incubation in straw.

Glycoside hydrolases belonging to family 61 were amongst the most highly expressed in straw in both organisms (~ 6% and ~ 14%). The genome of *T. reesei* encodes 6 GH61s whereas the one of *A. niger* encodes 7 GH61s (Appendix A5 Table A.5.T.1). Comparison of the primary protein sequences between individual GH61s from the same organism and from both fungi established that the 3 GH61s assayed in this study (GH61, GH61A and GH61B) from *T. reesei* were very different from each other whereas the ones from *A. niger* fell into two categories and clustered together (Figure 4.3.18). This was also confirmed in a recent study by Häkkinen *et al*. (2012): *gh61a* (73643) and *gh61b* (120961) belonged to different functional subgroups whereas *gh61* (22129) together with 2 other genes (31447 and 76065) belong to the same cluster and one gene (27554) did not cluster together with any other GH61 member (gene transcripts I.Ds are indicated in brackets). The expression of 3 *T. reesei* GH61-encoding genes was up-regulated in straw, 2 GH61-encoding genes were repressed and 1 did not change in expression at all (Appendix A5 Table A.5.T.1). In *A. niger* 5 of the 7 GH61-encoding genes were up-regulated in the presence of straw and the expression value of the combined glycoside hydrolase 61-encoding genes out of the total CAZy RNA was also much higher in *A. niger* (14%) than in *T. reesei* (6.3%) (Appendix A5 Table A.5.T.1).
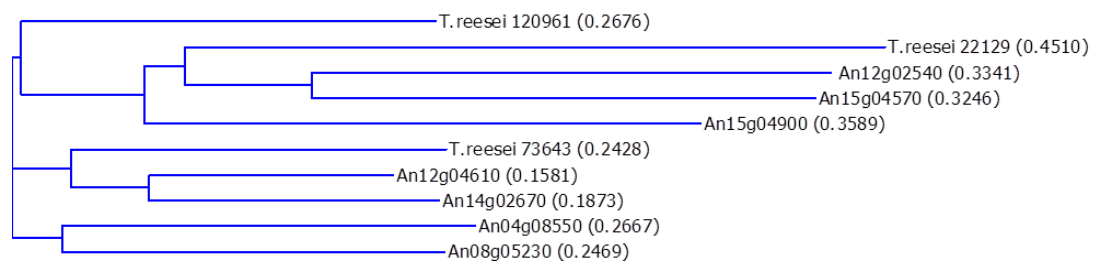
Figure 4.3.18: Sequence similarity tree based on the primary amino acid sequence of 3 *T. reesei* GH61s and 7 *A. niger* GH61s.

4.3.5.2 Non-CAZy gene expression

Twenty-seven genes which do not encode CAZy enzymes were up-regulated more than 20-fold and reached an RPKM greater than 50 after transfer for 24 hours from glucose to straw-based media (Appendix A4, Table A.4.T.1). In *A. niger*, highly induced genes encoding non-CAZy proteins were divided by Delmas *et al*. (2012) into 5 broad categories. Genes belonging to 4 of these functional categories (lipases, hydrophobic surface interacting proteins, transporters and carbon metabolism) were also highly induced in *T. reesei* suggesting a similar approach for both organisms when responding to the presence of a lignocellulosic substrate.

Lipases. One lipase-encoding gene was highly transcribed in the presence of straw whereas in both glucose conditions this gene was strongly repressed. BLAST analysis revealed that the protein encoded by this gene (Transcript I. D. 121418) is similar to a *A. niger* GDSL lipase/acylhdyrolase family protein (Gene I.D. in strain ATCC1510: 54865) which was also highly induced in the presence of straw and repressed in both glucose conditions (Delmas *et al*., 2012). The second gene listed in this category encodes a protein belonging to the neutral/alkaline non-lysosomal ceramidase family which hydrolyse the sphingolipid ceramide into sphingosine and free fatty acid, bioactive lipids serving as cellular messengers (IPR006823). The alkaline ceramidase described here is thought to be involved in cellulase signal transduction (JGI). BLAST analysis revealed the protein encoded by this gene (Transcript I.D. 64397) to be highly similar to a ceramidase (Gene I.D. ATCC1510: 120161) in *A. niger*, which was also highly induced in the presence of straw (Delmas *et al.*, 2012). Both genes contain a secretion signal sequence and have a similar transcriptional pattern to many GH and CE-encoding genes. It is likely that the *T. reesei* lipase and ceramidase are secreted with, and possibly regulated in a similar manner to the CAZy enzymes and participate in wheat straw deconstruction hence defining a new CAZy family.

Hydrophobic surface interacting proteins. Four genes encoding two hydrophobins and two cell wall proteins were induced more than 20-fold when switching from glucose to straw and all genes were repressed when glucose was added to the cultures. The genome of *T. reesei* encodes six hydrophobins of which three (HFB2, HFB3 and HFB5) were induced in the presence of straw; the genome of *A. niger* encodes seven hydrophobins of which four were induced in the presence of straw (HfbB, HfbC, HfbD and an unnamed protein) (Appendix A5 Table A.5.T.2 and Figure 4.3.19). These genes have a transcriptional profile similar to many genes of the CAZy group. In *Aspergillus oryzae*, the hydrophobin RolA recruits the esterase CutL to the

synthetic polyester polybutylene succinate-*co*-adipate (PBSA) and promotes its degradation (Ohtaki *et al*., 2006). A similar role for the two highly highly expressed hydrophobins of *T. reesei* could therefore be envisaged, as has been proposed for two *A. niger* hydrophobin-encoding genes which were also highly induced in the same conditions (Delmas *et al*., 2012).The other two highly induced cell wall protein-encoding genes are likely to also have a role in mediating interactions of the fungus with a solid substrate. The QI74 cell wall protein-encoding gene has been shown to be induced in *Trichoderma harzianium* when replacing glucose medium with chitin, simulating mycoparasitism conditions (Rey *et al*., 1998). The observation of the induction of hydrophobic surface interacting protein-encoding genes in both *T. reesei* and *A. niger* suggests that the recognition of solid surfaces is an important step in the fungal response to the plant cell wall.

Transporters. Seven genes encoding five transporters of the major facilitator superfamily (MFS), one xylose transporter and one oligo-peptide transporter were highly transcribed in straw and repressed in glucose-rich conditions. The MFS superfamily is a large family of transporters which can be divided into a further 17 families of which families 1, 5 and 7 mediate monosaccharide (hexoses, pentoses) and oligosaccharide transport into the cell by coupling it to proton symport or antiport (Pao *et al.*, 1998). The first gene listed here (Transcript I.D. 3405) is possibly involved in hexose and disaccharide transport as BLAST analysis of the encoded protein revealed 75% identity to a hexose transporter from *Glomerella graminicola* (Lingner *et al*., 2011) and 74% identity to a lactose permease from *Verticillium dahlia*. This transporter may belong to family 1 of MFS transporters which couple sugar uptake to proton symport and which are involved in the uptake of galactose, quinate, lactose, maltose and α-glycosides (Pao *et al.*, 1998). BLAST analysis of the protein encoded by the gene with transcript I.D. 50894 revealed 76% sequence identity to a high affinity glucose transporter from *Gaeumannomyces graminis* and may belong to family 7 of MFS transporters which couple hexose to proton import and are involved in the uptake of fucose, galactose and glucose (Pao *et al.*, 1998). The induction of transporter-encoding genes after 24 h in the presence of straw indicates that the cellulose and hemicellulose fractions of the wheat straw are being degraded, subsequently releasing simple sugars which are taken up by the fungus.

Carbon metabolism. As in *A. niger* (Delmas *et al*., 2012), genes of the xylose utilisation pathway such as xylose reductase (Appendix A4, Table A.4.T.1), xylitol dehydrogenase (*xdh1*, Transcript I.D. 81271) and xylulokinase (Transcript I.D. 123288) were up-regulated more than 5-fold when switching from glucose to straw. This indicates that after 24 h, *T. reesei* had begun the

degradation of the hemicellulose component xylan. On the other hand, *A. niger* expressed HsbA, a hydrophobic surface binding protein, in the presence of straw and this protein is absent from the genome of *T. reesei*. Figure 4.3.19 shows the expression patterns of HsbA ad several hydrophobin-encoding genes in *T. reesei* and *A. niger* as well as genes encoding SWO1 and one expansin-like protein.

PTH11-type GPCRs. The genome of *A. niger* encodes 70 potential PTH11-type GPCRs, similar to *Magnaporthe grisea* PTH11, which are thought to be involved in signalling the presence of various different substrates. This is double the number of genes encoding these types of proteins than are present in the genome of *T. reesei*. In *T. reesei*, 7 of these genes were expressed more than 10-fold when switching from glucose to straw, 9 were repressed in straw when compared to glucose, 3 genes did not manifest any changes in their expression values across all three conditions and the rest of the PTH11 encoding genes were up-regulated less than 10-fold (see Appendix A5 Table A.5.T.3 for details). In *A. niger*, 39 PTH11 encoding genes were up-regulated in the presence of straw when compared to glucose (of which 10 were up-regulated more than 10-fold) and 31 genes were repressed in the presence of straw when compared to glucose rich conditions (Appendix A5 Table A.5.T.3). Figure 4.3.20 shows the expression pattern of some PTH11 encoding genes in *T. reesei* and *A. niger*.
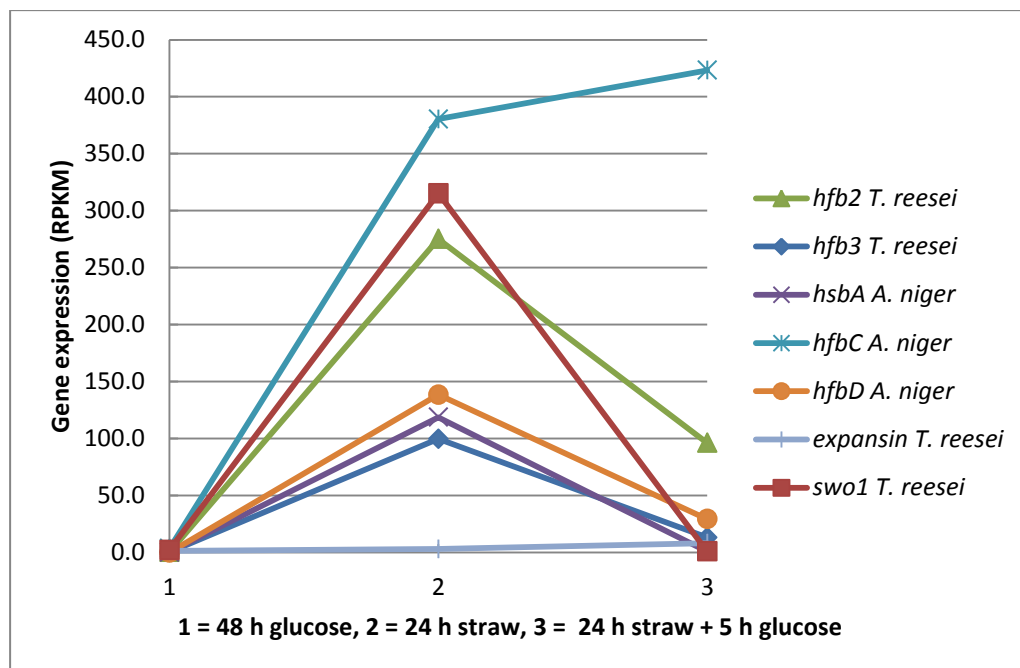
Figure 4.3.19: Expression patterns of genes encoding hydrophobins, HsbA, SWO1 and one expansin in *T. reesei* and *A. niger* when grown for 48 h in glucose then transferred for 24 h into straw-rich medium and with the addition of glucose to straw for 5 h. Details of gene expression values can be found in Appendix A5 Table A.5.T.2.
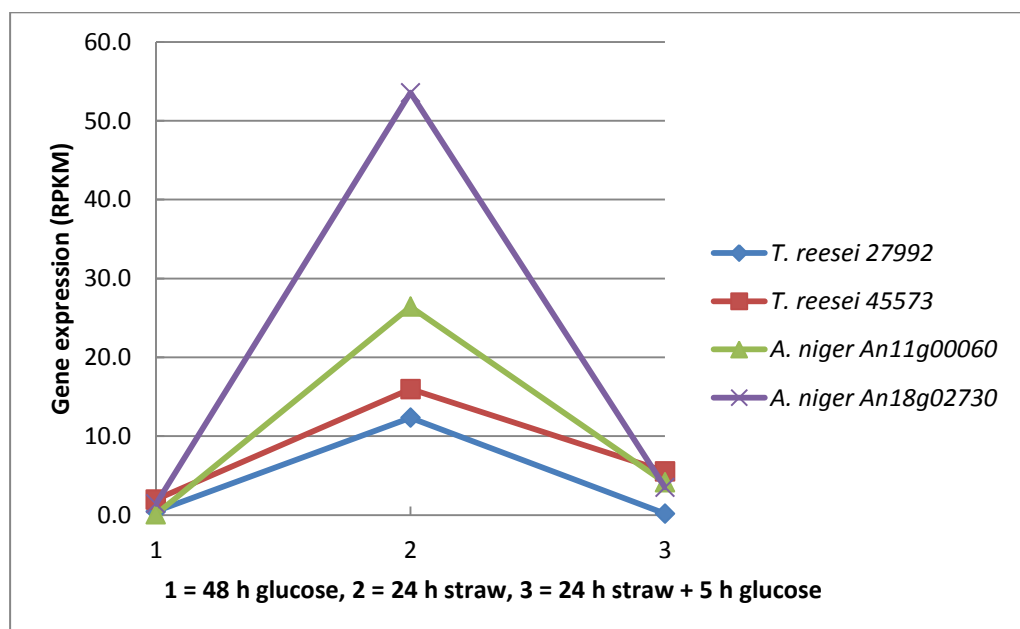
Figure 4.3.20: Expression patterns of genes encoding Pth11-type GPCRs in *T. reesei* and *A. niger* when grown for 48 h in glucose then transferred for 24 h into straw-rich medium and with the addition of glucose to straw for 5 h. Numbers indicate transcript I.Ds. Details of gene expression values can be found in Appendix A5 Table A.5.T.3.

4.3.5.3 Expression of CAZyme and non-CAZyme encoding genes in the presence of straw in the *A. niger* wild-type strain and a Δ*creA* mutant strain and with comparison to *T. reesei*

The expression of a few genes, encoding proteins that are putatively involved in carbohydrate deconstruction (mentioned under 4.3.5.2), were monitored by RT-PCR and compared between *A. niger* and *T. reesei* when cultured in glucose for 48 h, transferred into straw for 24 h and with the addition of 1% glucose for 5 h.

Genes encoding cellulolytic and hemicellulolytic enzymes were induced in the presence of straw in both organisms although glucose-mediated repression seemed to be more effective in *T. reesei* (Figure 4.3.21). In *A. niger*, genes encoding a GH61 protein, a Pth11-type GPCR, a hydrophobin HfbD and a non-hydrophobic hydrophobin HsbA were all up-regulated in the presence of straw but transcripts were also detectable in both glucose conditions. In *T. reesei*, this was also the case for *pth11*, *swo1* (*A. niger* lacks swollenin whereas *T. reesei* has no HsbA) and to a lesser extent for *gh61*. The gene encoding the hydrophobin HFB3 in *T. reesei* was repressed in 48 h glucose but induced in the presence of straw. The pattern of expression of these genes was fairly similar in both organisms (induction after 24 h incubation in straw and repression in glucose) although the onset of gene induction differed (Figure 4.3.23). In *A. niger*, *cbhA*, *gh61* and *hsbA* were fully induced after 9 h of incubation in straw whereas in *T. reesei cbh1* was fully induced at 24 h, *gh61* was fully expressed after 2 h of incubation in straw; and *swo1* was expressed at each time point studied here. Furthermore, in *T. reesei*, *pth11* transcripts were present at each time point studied here and the expression of *hfb3* occurred after 2 h of incubation in straw. In *A. niger pth11* induction occurred after 3 h in the presence of straw and *hfbD* transcripts were observed after 12 h of incubation in straw rich media.

The scenario is slightly different in carbon catabolite de-repressed strains (*A. niger* Δ*creA* and *T. reesei* Δ*cre1*). Transcript patterns of *cbhA*, *cbhB*, *gh61*, *pth11*, *hsbA* and *hfbD* in *A. niger* were almost identical between the wild-type and the Δ*creA* strains (Figure 4.3.22). In *T. reesei* this was also observed for *cbh1, cbh2, pth11* and one of the GH61-encoding genes. The expression pattern of *swo1* and *hfb3* was slightly altered in the Δ*cre1* strain where higher transcript abundance in straw and glucose was observed. Section 4.3.4 gives a detailed description of gene expression differences observed between *T. reesei* QM6a and *T. reesei* Δ*cre1* in the presence of straw.
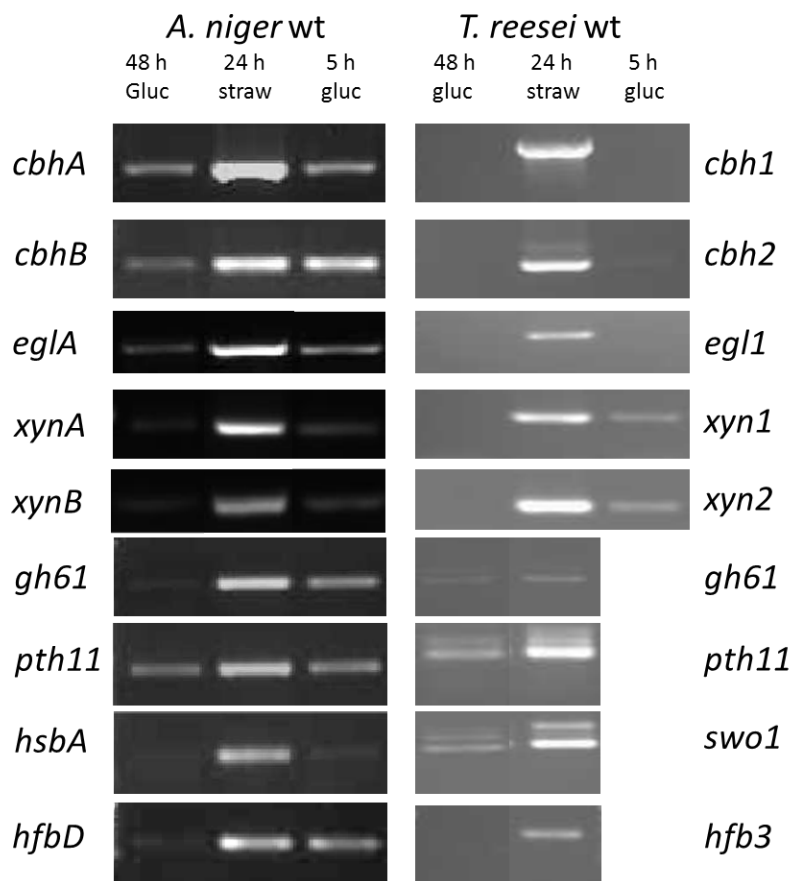
Figure 4.3.21: RT-PCR products of genes encoding glycoside hydrolases, hydrophobins and one PTH11-type GPCR in *A. niger* and *T. reesei* when grown in glucose for 48 h then transferred into straw-rich media for 24 h and with the addition of glucose to straw for 5 h (except for *T. reesei gh61*, *pth11*, *swo1* and *hfb3*). RT-PCRs were run on cDNA which was reverse-transcribed from RNA extracted from mycelia grown in the different conditions (see Materials and Methods).
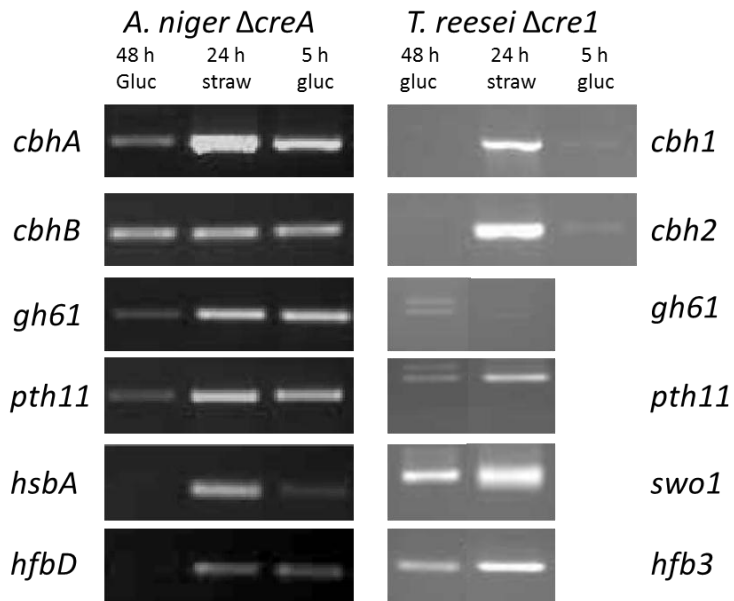
Figure 4.3.22: RT-PCR products of genes encoding glycoside hydrolases, hydrophobins and one PTH11-type GPCR in *A. niger* and *T. reesei* carbon catabolite de-repressed strains. Mycelia were grown in glucose for 48 h then transferred into straw-rich media for 24 h and with the addition of glucose to straw for 5 h (except for *T. reesei gh61*, *pth11*, *swo1* and *hfb3*). RT-PCRs were run on cDNA which was reverse-transcribed from RNA extracted from mycelia grown in the different conditions (see Materials and Methods).

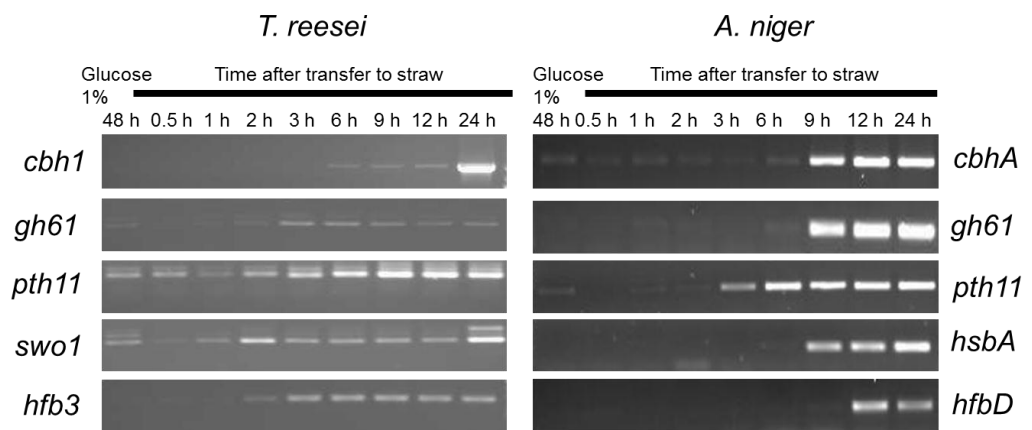**Figure 4.3.23**: RT-PCR products of genes encoding glycoside hydrolases, hydrophobins and one PTH11-type GPCR in *A. niger* and *T. reesei* when grown in straw for 24 h. Mycelia were grown in glucose for 48 h then transferred into straw-rich media for a time course of 24 h in total. RT-PCRs were run on cDNA which was reverse-transcribed from RNA extracted from mycelia grown in the different conditions (see Materials and Methods).

4.3.6 Natural antisense transcripts (NATs)

In order to find genes with NATs, raw reads from the sense and antisense strands from the transcriptome data were mapped to the genome of *T. reesei* using IGV. The programme differentiates between sense and antisense RNAs (different colours) and helps to visualise the direction and amounts of small RNAs for each gene. The genome was screened for genes which had NATs switching in direction between the glucose and straw conditions. Figures 4.3.24 and Figures A.7.F.2 – 4 in Appendix A7 show examples of 4 genes with NATs which switch in direction between the conditions studied here. These genes encode an amino acid transporter (transcript I.D. 81442), a β-glucuronidase (GH2, transcript I.D. 76852), a GH16 family member (transcript I.D. 70542) and a fungal specific transcription factor (transcript I.D. 121412). The number as well as the length of the NATs differed between the genes: the gene encoding a transcription factor had a very high number of long sense transcripts and a low number of short antisense RNAs in straw whereas the gene encoding the GH16 had a low number of RNA transcripts (sense and antisense) with only a part of the gene having short NATs. The gene encoding the amino acid transporter was repressed in straw and thus had a high number of mainly long antisense transcripts in this condition. The β-glucuronidase encoding gene had long NATs in 48 h glucose which switched mainly to long sense transcripts in straw. The number of sense RNAs for this gene in straw was much higher than the amount of small NATs. The addition of 1% glucose for 5 h caused the sense RNAs to disappear and the number of antisense transcripts became dominant again. The above mentioned examples show that differences in RNA transcript direction and abundance exist for different genes. Furthermore, introns of these genes were mainly spliced when this gene was expressed whereas, during repressive conditions, the introns were read through.

In total, 176 genes (1.93 % of the whole genome) were found to have NATs (Appendix A6 Table A.6.T.1). Most genes with NATs encode proteins with unknown function or general function prediction only (Figure 4.3.25). Many other genes encode components of the amino acid, carbohydrate and secondary metabolites transport and metabolism; proteins involved in post-translational modifications and protein turnover, and parts of the transcription machinery.
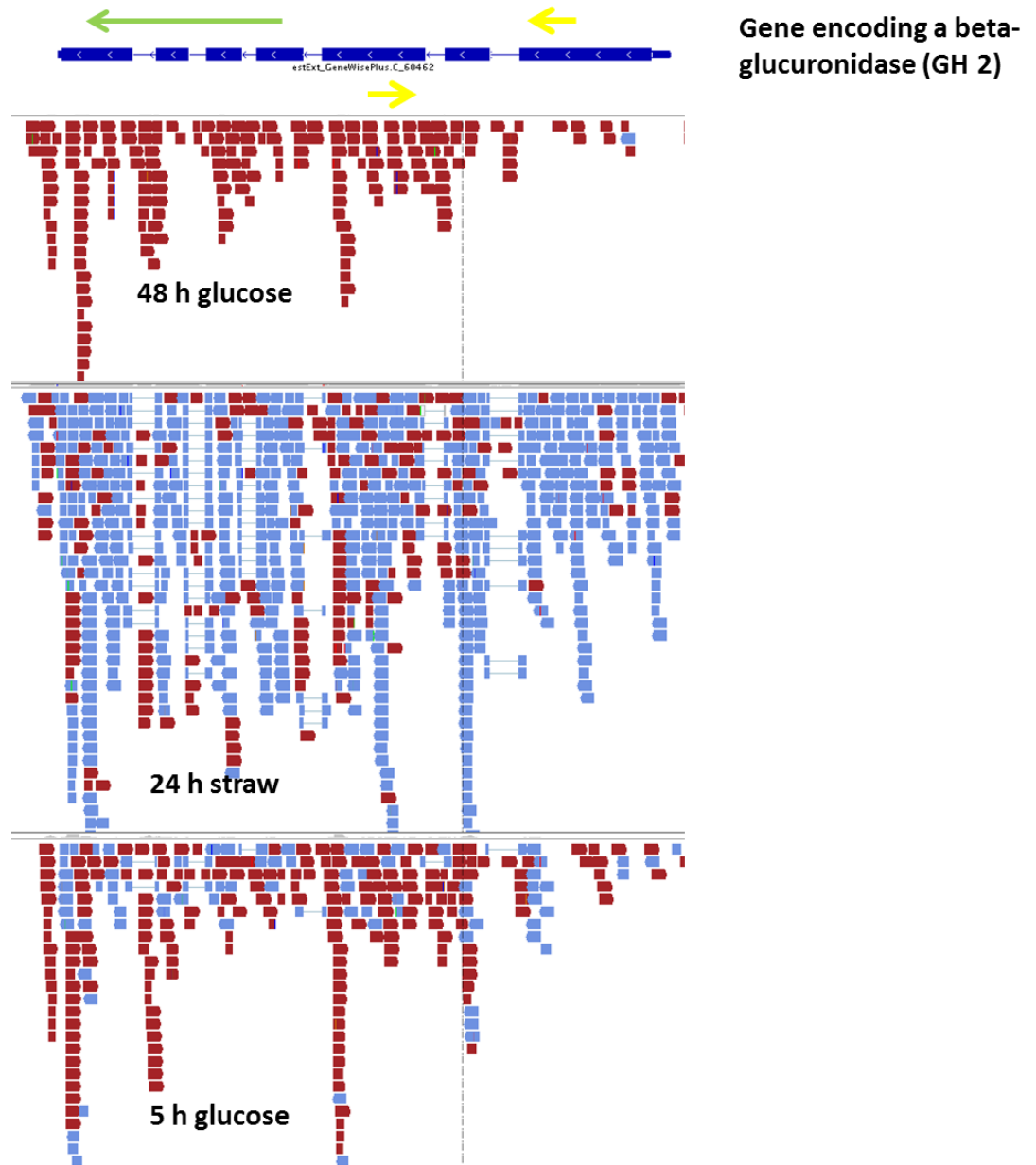
Figure 4.3.24: IGV output of all the sense (blue) and antisense (red) RNA transcripts mapped to a gene encoding a β-glucuronidase in QM6a mycelia which were grown in glucose for 48 h then transferred into straw-rich media for 24 h and with the addition of glucose to straw for 5 h. A diagram of the gene is shown in dark blue, the direction of the gene is indicated by a green arrow and yellow arrows indicate the location of the primers used for further studies. Intron splicing sites are portrayed by a light blue line between two blocks of RNA transcripts.
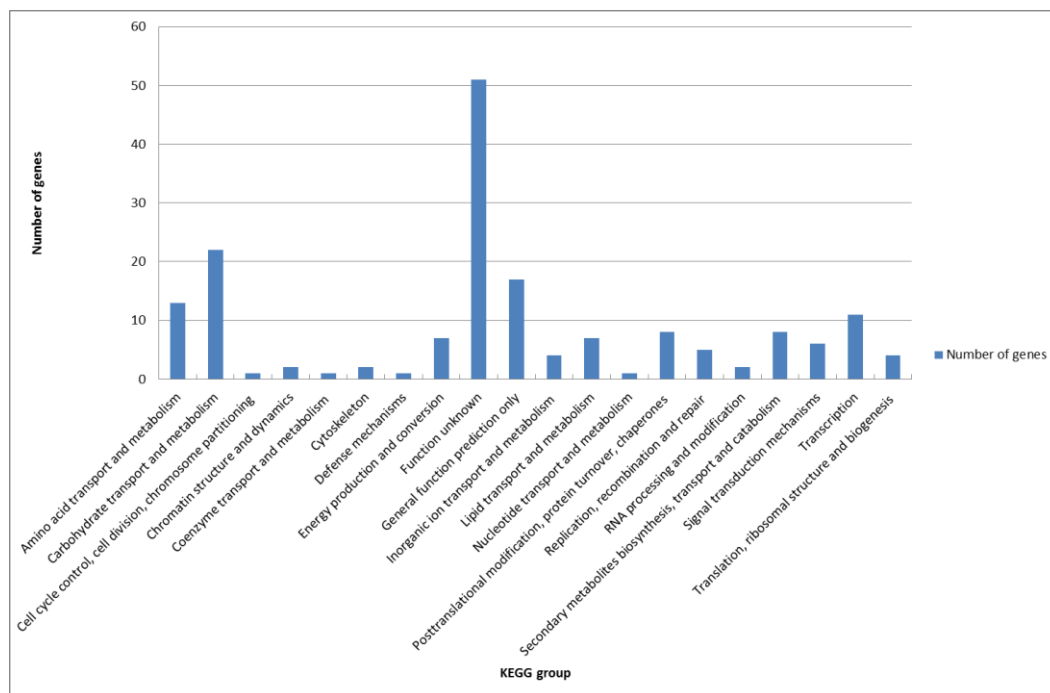
Figure 4.3.25: Number of genes with NATs classed according to their function into KEGG groups.

4.3.6.1 Confirmation of NATs

In order to confirm the presence of the NATs for these 4 selected genes, RT-PCRs and strand-specific PCRs (ssPCRs) were run on cDNA which was reverse transcribed from RNA extracted from mycelia grown in the above specified conditions (see Materials and Methods). The first PCR was run on gDNA and cDNA (from the three conditions) using the primers indicated in yellow in Figure 4.3.24 and Appendix A7 Figures A.7.F.2 – 4. PCR products were visualised on a 1% (w/v) agarose gel (Figure 4.3.26) and checked against their predicted sizes. The predicted sizes of the PCR products for the first gene, encoding an amino acid transporter, were the same for gDNA and 24 h straw cDNA. This is because this gene was repressed in straw and the introns were read through (326 bp); in glucose on the other hand, this gene was expressed and the introns were spliced (263 bp) giving a smaller DNA product (Appendix A7 Figure A.7.F.2). For the second gene, encoding a GH2 family member, the opposite was true: in both glucose conditions predicted DNA sizes were the same than with gDNA whereas in straw a band of a smaller size became more prominent. This gene was repressed in glucose and the introns were read through (670 bp); in straw on the other hand this gene was expressed and the introns were spliced (476 bp) although some antisense RNAs remained in

141

these locations (Figure 4.3.24). PCRs of the third gene which encodes a GH16 family member gave products of the same predicted size (344 bp) in all three conditions when compared to gDNA. This is because only 1 RNA transcript was spliced in all three conditions and this did not show on the gels as this method may not be sensitive enough to pick this up (Appendix A7 Figure A.7.F.3). This gene was expressed in glucose and repressed in straw by a low number of small antisense RNAs covering the region around the single intron. The final gene studied here encodes a transcription factor which was highly induced in straw when compared to both glucose conditions. In 48 h glucose, sense RNAs spanning the intron were spliced but read-through sense RNAs were also present, giving two differently sized predicted PCR products (299 bp and 240 bp). In straw on the other hand, spliced sense mRNA levels increased considerably as indicated by the most prominent band in panel 4 in Figure 4.3.26; a small fraction of antisense and sense RNAs which were non-spliced were also present, giving rise to further differently sized PCR products (Appendix A7 Figure A.7.F.4). In the condition 24 h straw + 5 h glucose, the number of sense RNAs decreased (as seen by a less intense DNA band in Figure 4.3.29) and as in 24 h straw, non-spliced sense and antisense RNAs remained. Thus all predicted PCR products obtained here were in perfect agreement with the IGV output data and the RNA sequencing data.
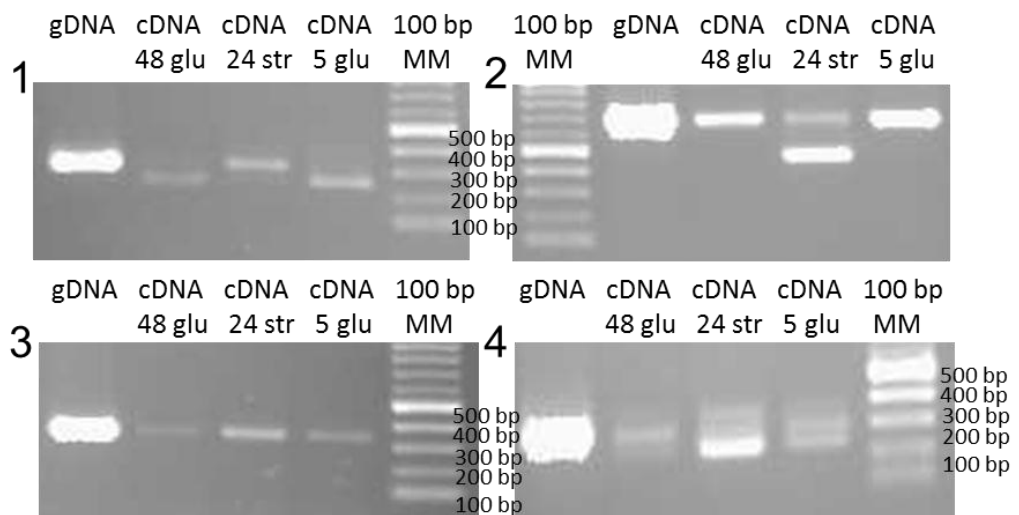
Figure 4.3.26: Products of PCR reactions run on gDNA and cDNA obtained from three conditions (48 h glucose, 24 h straw and 24 h straw + 5 h glucose). Genes encoding an amino acid transporter (1), a β-glucuronidase (2), a GH16 family protein (3) and a transcription factor (4) were assayed. A 100 bp ladder serves as the molecular marker (MM).

The second PCR was run on cDNA by using only one of the primers. This gave single-stranded cDNA PCR products. A further PCR was then run on these products using both primers this time and products were run on a 1% (w/v) agarose gel (Figure 4.3.27). As before, all results were in good agreement with the IGV mapping data. The gene encoding the amino acid transporter had stronger sense transcripts in both glucose conditions and a stronger antisense band in 24 h straw: this was because this gene was expressed under glucose rich conditions and repressed in the presence of straw. Sense RNAs were mapped to this gene in both glucose conditions whereas in straw-rich conditions most RNAs were antisense to this gene (Appendix Figure A.7.F.2). The second gene which was assayed here and encodes a GH2, was repressed in glucose and expressed in straw: at 48 h glucose this gene only had antisense RNAs mapped to it; upon transfer into straw media this gene was transcribed and thus the sense RNAs became more prominent. Addition of glucose to the straw media for 5 h caused an increase in antisense RNAs again. Different transcript sizes (PCR product bands) indicated that some sense RNAs were spliced for either one of the introns or both of the introns (see Figure 4.3.24). Antisense RNAs also had different lengths for this gene, thus giving rise to different transcript sizes. The gene encoding a GH16 had a very low number of sense and antisense transcripts (Appendix Figure A.7.F.3). An increase in antisense reads occurred for the region covering the intron when switching from glucose to straw and the amount of antisense transcripts decreased again when glucose was added to the straw media. This was also supported by the RT-PCR data (4.3.27). Sense transcripts also increased for a gene encoding a fungal specific transcription factor when incubated for 24 h in straw and decreased in both glucose conditions. This gene had a similar amount of antisense transcripts in both glucose conditions and this number increased only slightly in the presence of straw (Appendix Figure A.7.F.4). Upon transfer into straw rich media for 24 h, the amount of sense RNAs increased dramatically and diminished again when glucose was added (although not to the same low amount than under 48 h glucose). Sense and antisense RNAs covering the intron of this gene were either spliced or read through which gave rise to several different transcript sizes (Appendix Figure A.7.F.4). All of the above mentioned descriptions of sense and antisense RNAs for each of the 4 genes studied here were in good agreement with results obtained by RT-PCR, strand-specific PCR and also correspond to the mapping output by IGV. Furthermore amount of sense and antisense reads were also in agreement with RPKM data from the RNA sequencing.

Gene 1: amino acid transporter

Gene 2: β-glucuronidase

Gene 3: GH16

Gene 4: Transcription factor
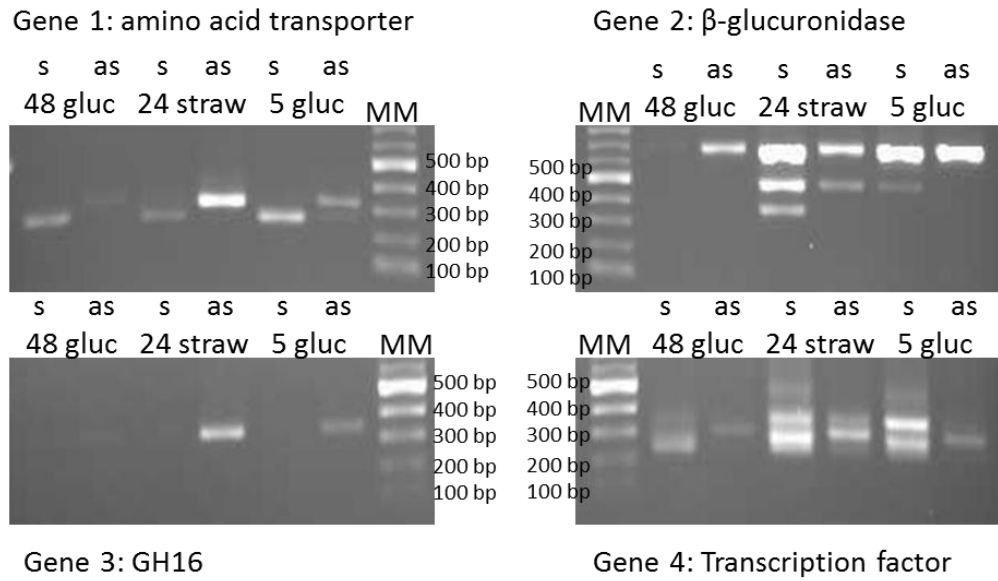
Figure 4.3.27: Products of PCR reactions run on sense (s) and antisense (as) cDNA strands (see materials and metods for strand-specific PCRs) from 3 different conditions: 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. Genes encoding an amino acid transporter, a β-glucuronidase, a GH16 family protein and a transcription factor were assayed (same genes as in Figure 4.3.26). A 100 bp ladder serves as the molecular marker (MM).

4.4 Discussion

The transcriptomes of *T. reesei* mycelia grown in three different conditions were sequenced and analysed. Mycelia were grown for 48 h in glucose then transferred into straw-based media for 24 h before glucose was added for another 5 hours. This gave three distinct conditions: repressive (48 h glucose), inductive (24 h straw) and inductive-repressive (24 h straw + 5 h glucose). Analysis of the transcriptomes between these three conditions revealed extensive differential gene regulation; the biggest shift in gene regulation was observed between 48 h glucose and 24 h straw, and between 24 h straw and 5 h glucose. There was not much difference between the transcriptomes of the two glucose conditions, indicating that the addition of glucose to straw caused a similar shift in gene expression patterns to the one for mycelia grown for 48 h in glucose. Furthermore, all significantly differently regulated genes were classed into KEGG groups according to the functions of the proteins they encode. Between the three conditions studied, most genes encoded proteins involved in carbohydrate and amino acid transport and metabolism, enzymes involved in energy production and conversion as well as genes encoding proteins required for transcription, post-translational modifications, protein turnover, signal transduction, lipid metabolism and secondary metabolite biosynthesis. Unfortunately many genes which were significantly differently regulated between these three conditions encoded proteins of unknown function. This is explained by the fact that about 1/3 (~3000 genes) of the *T. reesei* genome has an unknown associated function. The switch from glucose to straw and back to glucose again seems to require extensive remodelling of various metabolic pathways together with a change in transcription and protein synthesis. There were also differences between the three trancriptomes: the addition of glucose to straw caused the expression of genes encoding proteins involved in cytoskeleton and extracellular structure remodelling as well as proteins assisting in nucleolar transport and metabolism. Thus, the addition of glucose to straw seemed to induce a response in the nucleus (maybe by repressing the transcription of many genes) as well as having an effect on cell growth and morphology (see Chapter 3 for fungal growth in the presence of glucose). When comparing the two glucose transcriptomes, a lack of genes encoding proteins involved in cell cycle control, cell division, cytoskeleton and extracellular structures, nucleolar transport and metabolism was noticed. As discussed in Chapter 3, glucose is the preferred carbon source for *T. reesei* as it provides the means to quickly obtain high energy levels which are used to sustain growth and this could therefore explain the above observed results.

One of the main categories of genes significantly differently regulated between the three conditions studied here, was the one encoding proteins involved in carbohydrate transport and metabolism. These included extracellular enzymes (CAZy) which act on a certain substrate and degrade it into individual sugar monomers as well as transporters, enzymes which signal the presence of a specific sugar substrate and other proteins which assist in carbohydrate deconstruction. The amount of transcripts belonging to GHs, PLs and CEs increased from 0.45% in glucose to 8.25% in straw out of the total amount of cellular RNA (Figure 4.3.1). The presence of straw elicited a response resulting in the up-regulation of 79 genes encoding glycoside hydrolases and 7 CE-encoding genes required for straw degradation. These enzymes belonged to 42 different CAZy families with the largest number of genes encoding proteins from GH families 2, 3, 18, 27 and 92 (Figure 4.3.2). Thus the carbohydrate composition of the wheat straw (cellulose, hemicellulose and lignin) used in this study induced many genes encoding a variety of enzymes. Cellulase and hemicellulase as well as other glycoside hydrolase-encoding genes were all significantly up-regulated in the presence of straw; this probably occurred through the action of inducers such as sophorose and xylo-oligosaccharides resulting from the initial breakdown of the cellulose and xylan fractions of the straw (see Chapter 5 for details on the regulation of these genes). The onset of expression of cellulase (*cbh1*, *bgl1*) and hemicellulase (*xyn1*, *xyn2*)-encoding genes in straw-rich conditions occurred after 6 h for *cbh1* and *xyn2* and after 2 h for *xyn1* whereas *bgl1* transcripts were present at all the time points studied here (Figure 4.3.7). The expression of some cellulase and hemicellulase-encoding genes as well as other glycoside hydrolase-encoding genes was also confirmed by RT-PCR and qRT-PCR. QRT-PCR results for *cbh1*, *xyn1* and *xyn2* were different from the RNA-sequencing results though: RPKM data suggested that transcript levels are highest for *xyn2* followed by *cbh1* and then *xyn1*, whereas qRT-PCR data suggested that *xyn1* transcript levels are highest followed by *cbh1* and *xyn2* levels. The reason for this could be a difference in the methods used for RNA sequencing and qRT-PCR which could have led to differences in the total amount of RNA (see Materials and Methods). All these genes showed high transcript abundance at 24 h straw and levels of transcript abundance were either decreased or abolished in both glucose conditions (Figures 4.3.4, 4.3.5, 4.3.6). Exceptions to this expression pattern, were two genes encoding a GH47 α-1,2-mannosidase and a GH76 endo-1,6-β mannosidase which had similar transcript levels across all three conditions. This was in agreement with the RNA sequencing data and showed that not all glycoside hydrolase-encoding genes were expressed and repressed to the same levels.

Furthermore, genes encoding mannanases, acetyl xylan esterases, glucuronidases, arabinofuranosidases and galactosidases were also induced in straw (Appendix A3 Table A.3.T.1). These enzymes act on the hemicellulose fraction of the straw and are specialised in deconstructing the mannan backbone, removing acetyl groups from polymeric xylans and cleaving galactose and glucuronic acid side groups (Margolles-Clark *et al.*, 1997). *T. reesei* is capable of deconstructing arabinan in hemicellulose through the synergistic action of three arabinofuranosidases (ABF1-3) and BXL1. ABF1 and ABF3 both have cellulose binding domains indicating a role in cleaving interactions between cellulose microfibrils attached to matrix components through residues such as arabinan, xylan and galactan (Akel *et al.*, 2009). The genome of *T. reesei* does not encode any endo-arabinases but among the sordariomycetes it is the only fungus which contains two GH54 protein-encoding genes (ABF1 and ABF3). Multiple genes encoding enzymes with similar functions from the same families (e.g. glucosidases, xylosidases or mannanases) are also up-regulated in straw (Appendix A3 Table A.3.T.1); each of these enzymes may present a different effectiveness against the same substrate or act on a different location on the same substrate. In *T. reesei*, β-xylosidases can be classed into GH families 3 and 54 and especially GH3 family members exhibit a combination of different activities (Knob *et al.*, 2010). For example, *T. reesei* BXL1 not only possesses xylosidase activity but also arabinofuranosidase and glucosidase activities.

Glycoside hydrolases belonging to the GH61 family were also highly up-regulated and together with GH families 3, 5, 6, 7, 11, 30 and 67 made up the largest part of the enzyme cocktail secreted in the presence of straw (Figure 4.3.17). RT-PCR and qRT-PCR of the three GH61-encoding genes showed that each one was differently regulated: *gh61* and *gh61b* transcripts were present at each of the time points studied here with *gh61b* transcript abundance being much higher; whereas *gh61a* expression started at 6 h incubation in straw (Figure 4.3.7).

The expression patterns of genes encoding non-hydrolytic proteins putatively involved in carbohydrate deconstruction were also studied (Figure 4.3.7). These genes encode a PTH11-type GPCR, a GPRK, the hydrophobin HFB3 and the swollenin SWO1. RNA-sequencing data indicated that *pth11* and *gprk* were up-regulated more than 10-fold and *hfb3* and *swo1* were up-regulated over 20-fold and with an RPKM > 50 (Appendix A4 Table A.4.T.1) when switching from glucose to straw; hinting at a role of these proteins for initiating and enhancing polysaccharide degradation. A similar role for these proteins was proposed in *A. niger* (Delmas *et al.*, 2012). An additional gene thought to enhance cellulase activity is *grd1* encoding a glucose ribitol

dehydrogenase (Schuster *et al*., 2011). This gene was down-regulated in the presence of straw and up-regulated in glucose-rich conditions supporting a role of this enzyme in the oxidation of glucose to gluconolactone (Schuster *et al*., 2011). GRD1 was reported to enhance cellulase activity on pure crystalline cellulose and sophorose by keeping the activity of intracellular BGL in check and preventing inhibitory glucose build-up (Schuster *et al*., 2011). The down-regulation of this gene in a more complex carbon source may suggest that very small quantities of glucose are released from straw after 24 h or that this gene may not actually have a significant role in enhancing cellulase activity. RT-PCR and qRT-PCR (Figure 4.3.8, *pth11* only) showed that *pth11* and *swo1* transcripts were present at all the time points studied here. Transcripts of *gprk* (encoding the G-protein coupled receptor kinase) and *hfb3* were present from 3 h and 2 h in straw respectively. Swollenin is a non-hydrolytic protein but also contains a CBM1 module and it is thought to enhance cellulase activity by disrupting cellulose microfibrils. Regulation of *swo1* has been reported to be similar to the regulation of cellulase-encoding genes although expression levels of this gene are generally lower than the cellulase ones (Saloheimo *et al.*, 2002). Transcripts of *swo1* were observed at all the time points studied here and this suggests that the onset of early expression of this gene (when compared to *cbh1* for example) is advantageous for when the cellulases are finally secreted. This fits with the above described role of SWO1. Hydrophobins are small hydrophobic proteins which have a role in mediating the interaction between the fungus and its environment (e.g. carbon source). The genome of *T. reesei* encodes 6 hydrophobins of which HFB2, HFB3 and HFB5 are expressed in the presence of straw. Detectable expression of *hfb3* occurred after just 2 h of incubation in straw hinting at a role for this protein in helping to attach the fungus to the straw particles and thus initiate straw breakdown. Transcript abundance of *pth11* was observed at each time point studied here but increased at the later time points. PTH11-type GPCRs have been associated with pathogenicity in *M. grisea* as they sense plant surface cues (such as cutin monomers) and relay the signal for appressorium differentiation (Wilson and Talbot, 2009). The increase in transcript abundance of this gene in *T. reesei* during 24 h incubation in straw indicates that it is likely that this GPCR may sense the presence of straw (possibly by binding to a small plant cell wall oligosaccharide as is the case in *M. grisea*) and relays this signal to the intracellular compartments. In *M. grisea*, the hydrophobin MPG1 acts as an additional cue for PTH11-mediated signalling. HFB3 or any of the other two up-regulated hydrophobins may have a similar role in *T. reesei*. A further six PTH11-type GPCRs encoding genes were expressed more than 10-fold in the presence of straw so it is entirely possible that they may also have a role in signal transduction. The later onset of *gprk*

(when compared to *pth11*) expression may suggest a role of this G-protein in sustaining the straw-mediated signal (if it is involved in carbohydrate-mediated signalling) or by signalling the presence of different sugar polymers resulting from the degradation of various straw carbohydrate chains and side groups. Thus, different GPCRs may work together to signal the presence of a potential carbon source and to sustain the signal in order to allow degradation of the substrate.

In order to test whether the results observed under straw-rich conditions were not the outcome of the fungus being unable to recognise and degrade this substrate, the same RT-PCR and qRT-PCR reactions were carried out when the fungus was incubated for 48 h in glucose then transferred into no carbon source (NC) medium for 24 h and with the addition of glucose to straw for 5 h (Figures 4.3.4 and 4.3.6). A 24 h time course was also established under carbon starvation conditions (Figure 4.3.7). Transcription patterns of all the above mentioned genes were different under carbon starvation conditions when compared to straw-rich conditions. There was very little transcript abundance of all the glycoside hydrolase-encoding genes when grown in media with no carbon source (including the NC 24 h time course); the only exceptions were *gh2* and *gh47*. Transcript patterns of *gh61* were similar under NC and straw conditions whereas *gh61a* transcripts were present at each time point and *gh61b* expression started at 6 h under carbon starvation conditions. This was also different from the results obtained for these genes under straw-rich conditions. Transcripts of *pth11* and *swo1* were still present at all the time points studied here, although transcript abundance was much less than in straw as was confirmed by *pth11* qRT-PCR (Figure 4.3.9). The opposite was the case for expression of *hfb3* and *gprk*: transcripts of these two genes were observed at earlier time points under carbon starvation conditions (1 h and 2 h) when compared to straw-rich conditions and transcript abundance was also higher although this requires further confirmation. HFB3 could therefore act as a "scavenger" by looking for and binding to potential substrates and subsequently act as a cue for PTH11 signalling and glycoside hydrolase expression. These results suggest that *T. reesei* is able to recognise straw as a potential carbon source and initiate its breakdown. Delmas *et al.* (2012) analysed the sequenced transcriptome of *A. niger* when grown in the same conditions as described in this study, and they found that genes encoding HsbA and a Pth11-type GPCR as well as an additional subset of genes encoding CAZy (CbhB, GH5, AbfB) and non-CAZy (lipase, esterase) were up-regulated at early time points (6 h) after transfer from glucose to straw. They proposed a model, in which carbon starvation during the first few hours after transfer into straw-rich media, alleviates carbon catabolite-mediated repression by CreA which subsequently allows the expression of genes

encoding CAZy and non-CAZy proteins involved in initiating and signalling carbohydrate degradation. These early expressed proteins then cleave off inducer molecules, which are taken up into the fungal cells and fully induce the expression of the major hydrolytic system. A similar mechanism could also operate in *T. reesei*: temporally differential up-regulation of *hfb3*, *gprk* and *pth11* (as well as other CAZy and non-CAZy-encoding genes not analysed in this study) mediated by CRE1 alleviation, suggests signalling events as well as a "scavenger" role for these proteins in the absence of a carbon source. Carbon starvation could possibly trigger the induction of genes encoding "scouting" proteins (e.g. HFB3 and others, not discovered here) which mediate the attachment of the fungus to the substrate. This could serve as a cue for PTH11 (as is the case in *M. grisea*) and/or GPRK (and maybe others) activation which could then signal the presence of wheat straw to the cells and this would then trigger the induction of CAZy-encoding genes necessary for polysaccharide degradation. Low basal expression (Kubicek *et al.*, 2009) and the early presence of transcripts (2 h, 6 h after incubation in straw) of genes encoding major hydrolytic enzymes such as CBH1, XYN1, XYN2 and BGL1 may further aid in the formation of inducers which serve as an additional signal for CAZy-gene induction. Expression of MFS transporters (over 20-fold induction, see Appendix A4 Table A.4.T.1) in the cell membrane would allow for efficient inducer and sugar uptake. Two genes, wich were up-regulated over 20-fold between 48 h glucose and 24 h straw, also encoded fungal specific transcription factors which may play a role in regulating genes involved in carbohydrate utilisation. Recently, two new transcription factor-encoding genes termed *clr-1* and *clr-2* were described in *N. crassa* and which are thought to mediate cellulose degradation by up-regulating genes encoding 16 cellulases, 6 hemicellulases, 15 enzymes with activity on polysaccharides as well as proteins involved in the secretion pathway and the UPR (Coradetti *et al.*, 2012). Orthologues of *clr-1* and *clr-2* in *T. reesei* (transcript I.Ds 27600 and 26163) were up-regulated over 5-fold when switching from glucose to straw and these transcription factors may also play a role in regulating cellulase-encoding genes in addition to other known *T. reesei* regulators (XYR1, ACE1, ACE2, CRE1 and HAP2/3/5 complex; see Chapter 5). Induction of cellulase and hemicellulase-encoding genes in *N. crassa* in the presence of substrates such as cellulose (Znameroski *et al.*, 2012) was proposed to occur through the following mechanism: low constitutive levels of extracellular or cell wall-associated enzymes in *N. crassa* release metabolites from plant cell walls which are taken up by the fungus and act as inducers of cellulase and hemicellulase-encoding genes through activating various transcription factors which independently and synergistically regulate the expression of the major hydrolytic genes (Sun *et al.*, 2012).

The same experiments as mentioned above were also performed in a strain carrying a *cre1* gene deletion. RT-PCR and qRT-PCR indicated that transcript levels of all glycoside hydrolase-encoding genes increased after 24 h in straw medium when compared to both glucose conditions but were much higher in the wild-type strain. Transcript levels were similar in carbon source-depleted conditions for both strains (compare Figures 4.3.4 and 4.3.10). As discussed in Chapter 5, the deletion of the main carbon catabolite repressor caused a small increase in glycoside hydrolase transcript levels in the presence of glucose because full gene induction requires the presence of inducers (such as cello-oligosaccharides and xylo-oligosaccharides probably released from straw breakdown). Furthermore, deletion of *cre1* seemed to impair cellulase and hemicellulase gene induction in QM6a: this indicates a role for CRE1 being important for the correct regulation of these genes in the presence of a complex carbon source such as straw (in contrast to simple carbon sources such as sophorose and cellobiose, see Chapter 5). Expression patterns which were also different between the two strains are those for *gh76* and *gh92* in straw-rich conditions. Transcripts of *gh92* were completely absent whereas transcript abundance of *gh76* was very low in Δ*cre1*. This further supports a role for CRE1 in the expression of some GHs in the presence of an utilisable carbon source (these differences are not observed under carbon starvation conditions). CRE1 is described as a cell-wide regulator and it was shown that deletion of this gene causes the down-regulation of many other genes (Portnoy *et al.*, 2011). CRE1 is also likely to interact with genes for other pathways which may be activated by the different substrates present in straw (e.g. pentose utilisation pathways) and that knock-out of *cre1* causes a general defect also affecting cellulase and hemicellulase gene regulation. It is also possible that QM6a and QM6a Δ*cre1* carry unmapped mutations in their genomes which may affect glycoside hydrolase gene regulation. CRE1 was also shown to be important for correct nucleosome positioning within the *cbh1* coding region during *cbh1* repressing conditions by probably interacting with chromatin-remodelling factors (see Chapter 5). The differential temporal regulation of these genes was also different between the two strains. The *cre1* deletion caused a much earlier onset of *cbh1*, *xyn1* and *xyn2* expression in straw rich conditions; this effect was less severe under carbon starvation conditions where these genes were barely transcribed. In the wild-type strain, the presence of straw caused full induction of *bgl1* at 24 h whereas transcript abundance in the Δ*cre1* strain was much lower and similar at each time point. Thus, CRE1 represses the main cellulolytic and hemicellulolytic genes quite tightly probably until enough inducer is present within the cell which can alleviate these genes from carbon catabolite-mediated repression. In the

mutant strain no such repression is possible and gene induction most likely occurs after the uptake of the first inducers into the mycelia.

The expression patterns of the three GH61-encoding genes also differed between straw-rich and NC conditions in the Δ*cre1* strain and these patterns were different from the ones seen in the wild-type strain. Expression of *gh61* in the wild-type strain and in the mutant strain under carbon starvation and straw-rich conditions was very similar although transcript abundance was higher in QM6a under carbon starvation conditions (compare qRT-PCR data). In the Δ*cre1* strain, *gh61a* transcripts were detected at each assayed time point in straw-rich conditions whereas under carbon starvation conditions *gh61a* transcript abundance decreased over the 24 h time course. This was in contrast to the situation in the wild-type strain where *gh61a* transcripts were only detected after 6 h of incubation in straw and at each time point in NC media. Transcripts of *gh61b* were only detected at 24 h in straw and at each of the time point under carbon starvation conditions in the Δ*cre1* strain. In the wild-type strain, *gh61b* was expressed at each time point in straw-rich conditions and from 6 h onwards in carbon-depleted conditions. The gene encoding the GH61 enzyme is probably not under the control of CRE1 as no big differences in the expression pattern between strains were found. GH61A may be regulated by carbon catabolite repression as the early onset of expression in straw-rich conditions in the Δ*cre1* strain when compared to the wild-type strain was similar to the expression pattern observed for other glycoside hydrolase-encoding genes. This scenario seems to be reversed for GH61B: it appears that this gene requires CRE1 for full expression under straw-rich conditions and that a *cre1* knock-out severely delays and decreases *gh61b* transcript formation. The GH61 family of glycoside hydrolases lack the structure typical for GHs; instead they were described to be copper-dependent oxidases (Quinlan *et al*., 2011; see introduction). GH61s are capable of significantly enhancing cellulase activity (Harris *et al.*, 2010). The genome of *T. reesei* encodes 6 GH61s with low sequence similarity to each other and which make up a significant proportion of the enzyme cocktail secreted here (Figure 4.3.17); only one of the proteins has a CBM1 module (GH61A). It can therefore be expected that all three genes are likely to be subject to different regulatory mechanisms and that GH61A, which also possesses a CBM module, is regulated in a manner similar to the cellulases and hemicellulases.

The impact of the Δ*cre1* deletion was less severe on the expression patterns of genes encoding non-glycoside hydrolase proteins. There were some differences though between the two strains, suggesting that CRE1 may interact indirectly with pathways in which these proteins are involved. The

expression patterns of *pth11* and *swo1* were very similar in all conditions between the two strains (compare *pth11* qRT-PCR data). In the Δ*cre1* strain, *swo1* transcript abundance was much higher than in the wild-type under straw-rich conditions, and under carbon starvation conditions transcript abundance of this gene decreased; this is consistent with findings in this study and with previous reports that swollenin is regulated in a similar manner to cellulases: *cre1* deletion induced higher transcript levels of cellulases in inducing conditions whereas under carbon starvation conditions low transcript abundance was observed. Transcript patterns of *gprk* and *hfb3* were similar under carbon starvation and straw-rich conditions in the Δ*cre1* strain and this was also the case in the wild-type strain. In QM6a, the onset of *gprk* and *hfb3* gene expression happened at 2 h and 3 h respectively in straw-rich conditions and at 1 h and 2 h respectively under carbon starvation conditions. In the mutant strain, transcripts for these two genes were observed in both conditions at each one of the time points. Thus *swo1*, *gprk* and *hfb3* are likely to be directly or indirectly under the influence of the carbon catabolite repressor CRE1 whereas *pth11* is not.

The expression patterns of genes required for carbohydrate deconstruction in *T. reesei* were also compared to their orthologues (if present) in the *A. niger* strain N402 (short spore chain mutant of N400). Both organisms employ a certain set of enzymes from the same GH families to deconstruct straw but secrete members of different CEs families as well as additional GHs families which help the breakdown of this carbon source (see section 4.3.5.1 for details of the differences between the two organisms).

One of the most highly transcribed glycoside hydrolases were those belonging to GH family 61. In *T. reesei* GH61 transcripts made up ~6% of the total secreted enzyme mix whereas in *A. niger* they made up ~14% of the total amount of secreted proteins. The number of genes encoding GH61s was similar in *T. reesei* (6) and in *A. niger* (7): 4 genes were induced in the presence of straw in *T. reesei* whereas 5 GH61-encoding genes were up-regulated in the presence of straw in *A. niger*. Furthermore, sequence comparison revealed that the 7 *A. niger* GH61 proteins cluster into two groups based on their primary amino acid sequence similarity, whereas the protein sequences of the 3 *T. reesei* GH61s studied here, have low sequence similarity and are not grouped together (Figure 4.3.18). Häkkinen *et al*. (2012) reached a similar conclusion when doing a cluster analysis of the 6 *T. reesei* GH61s. This suggests that although *T. reesei* GH61 transcript levels are less abundant than those in *A. niger*, they may fulfil different roles and are probably subject to different regulatory mechanisms (see above) whereas an expansion of closely related GH61 proteins in *A. niger* suggest activity on similar substrates. The *A.*

niger GH61-encoding gene An12g0460 was not induced until after 9 h of exposure to straw (Delmas *et al*., 2012), suggesting that GH61s in *A. niger* are not part of the early response to carbon starvation but are secreted together with the major cellulases and hemicellulases.

Genes encoding non-CAZy enzymes which are likely to be involved in carbohydrate deconstruction were similar between *A. niger* and *T. reesei* suggesting a similar approach of both fungi for carbohydrate deconstruction (Delmas *et al*., 2012). These non-CAZy enzymes included lipases, hydrophobins, transporters and enzymes involved in xylan metabolism (see section 4.3.5.2 for details). Differences lay within non-hydrolytic CAZy-encoding genes (e.g. SWO1, CIP1 which are absent from the genome of *A. niger*) and the secreted non-hydrophobic hydrophobin HsbA (hydrophobic surface binding protein A) which is absent from the genome of *T. reesei*. Thus, the approach used to attach mycelia to straw particles and subsequent initiation of breakdown is similar but different in the detail in both organisms. Furthermore, the genome of *A. niger* encodes 70 potential PTH11-type GPCRs which is double the number of genes encoding these types of proteins than in the genome of *T. reesei* (Appendix A5 Table A.5.T.3). *T. reesei* may make up for the lack of these genes by having a higher number of genes encoding cAMP-receptor like proteins, *H. sapiens* mPR-like proteins (steroid receptors) and *Aspergillus* GPRK-like receptors (Brunner *et al.*, 2008). In *T. reesei*, 23 of PTH11-type GPCR-encoding genes were up-regulated in the presence of straw with 7 genes being expressed more than 10-fold when switching from glucose to straw. In *A. niger*, 37 PTH11-encoding genes were up-regulated in the presence of straw when compared to glucose-rich conditions and 8 of these genes were up-regulated more than 10-fold

To summarise, similarities in the transcriptional response to straw between *A. niger* and *T. reesei* include the composition of the secreted enzyme cocktail whereas differences between both organisms are found in the number and versatility of CAZy and hydrophobin-encoding genes, the presence/absence of swollenins, expansins and HsbA from either genome and the number of PTH11-encoding genes in both organisms. This suggests that both organisms use a similar array of enzymes to degrade straw but that the amounts of each enzyme secreted as well as the steps mediating carbon source-related recognition and signal transduction are essentially different. This study provides a basis for further analysis and characterisation of genes shown to be highly induced in the presence of a lignocellulosic substrate in order to attempt to elucidate the mechanism of solid substrate recognition and subsequent degradation in *T. reesei* and provide information on the conserved and diverged approaches used by saprobic fungi for lignocellulose

degradation. Transcription of some of the above mentioned genes (having high RPKM values in straw and low RPKM values in glucose) were also studied by RT-PCR when cultured in glucose for 48 h then transferred into straw-rich conditions for 24 h and with the addition of glucose to straw for 5 h (Figure 4.3.21). Repression of cellulase-encoding genes was more efficient in *T. reesei* as transcript bands were present in both glucose conditions for *A. niger cbhA* and *cbhB* (Delmas *et al.*, 2012). Regulation of cellulolytic genes by CRE1/CreA and XYR1/XlnR (Chapter 5) has been reported to differ between the two organisms (Stricker *et al.*, 2008). On the other hand transcript patterns for *A. niger hsbA*, *gh61*, *pth11* and for *T. reesei swo1*, *gh61* and *pth11* were similar and present at 48 h glucose and 24 h straw. The *T. reesei* gene encoding the hydrophobin HFB3 was repressed in glucose whereas *A. niger hfbD* transcripts were observed at 48 h glucose. Differences in the patterns of these gene transcripts were also observed when doing a 24 h time course in straw (see Figure 4.3.23 and section 4.3.5.3).

In the *A. niger* carbon catabolite de-repressed strain, Δ*creA*, transcript patterns of *cbhA*, *cbhB*, *gh61*, *pth11*, *hsbA* and *hfbD* were very similar between the wild-type and the *creA* knock-out strain (Figure 4.3.22; Delmas *et al.*, 2012). This was not the case in *T. reesei* (described in detail in section 4.3.4 and in the first part of the discussion) indicating again that CRE1/CreA-mediated carbon catabolite repression is different between the two organisms studied here.

Finally, the transcriptome of *T. reesei* was also scanned for genes with NATs between the three conditions studied here. The *T. reesei* genome was specifically screened for genes with NATs which were considered to be open reading frames (ORFs) with RNA transcripts which change in direction (sense/antisense to antisense/sense) between different conditions (Faghihi and Wahlestedt, 2009), as gene regulation by these types of NATs was easily identifiable and most interesting. About 2% of all the genes had NATs which varied in abundance and length (Appendix A6 A.6.T.1). Most of these genes encode proteins of unknown function whereas others are involved in carbohydrate and amino acid transport and metabolism and transcription (Figure 4.3.25). Four genes with NATs were chosen to be confirmed by PCR and results were in good agreement with the RPKM data and the IGV mapping data (Figures 4.3.26 and 4.3.27). It is known that NATs have several different regulatory roles (Faghihi and Wahlestedt, 2009) but it is impossible to know, without further experiments, what the roles are for each gene assayed here.

The genes encoding an amino acid transporter and a GH2 enzyme both had long NATs which also extended into the promoter and the 3' regions of these

genes. This suggests a tight regulatory mechanism where, under repressive conditions, all of the mRNA is bound by its antisense counterpart and gene expression is completely shut down. It cannot be predicted what the exact roles of these long double-stranded RNA molecules are (initiation of RNAi is one possibility) without further experimental investigation. The GH16-encoding gene on the other hand had a very low amount of sense transcripts (as supported by the RNA sequencing data). Mainly, the region covering the intron had an increase in antisense transcripts when switching from glucose to straw. These NATS were quite short and it is possible that genes with low expression require only a portion of their sense RNA to be bound by their antisense counterparts for efficient shut down. Alternatively, repression of this gene may not be as important as efficient shut down of the GH2 or amino acid transporter-encoding genes. The gene encoding a transcription factor saw the biggest increase in sense transcripts (mRNA, gene expression) when switching from glucose to straw. It is interesting to note that the number of antisense transcripts did not change much across the three conditions studied here. In 48 h glucose, the levels of antisense transcripts equalled the levels of sense RNAs, whereas in 24 h straw, the amount of sense transcripts increased considerably. In 5 h glucose the number of sense transcripts had decreased but still outweighed the amount of antisense reads. This could hint at a threshold role of NATs for this gene: under repressive conditions antisense RNAs ensure that the gene is not expressed whereas under inducing conditions the number of sense transcripts exceeds the antisense ones and the protein is expressed. Another interesting feature of the genes encoding the transcription factor and the β-glucuronidase proteins is that several splice variants exist for these genes (IGV output data). The reason for this is unknown and it is also questionable whether all splice variants are actually translated to different protein isoforms.

4.4 Conclusion

The sequenced transcriptomes of *T. reesei* exposed to three conditions (repressive, inductive and repressive-inductive) studied here revealed extensive differential gene expression between each one of them and especially when switching from glucose to straw and when adding glucose to straw-rich media. The transcriptomes between these two conditions were also different, indicating metabolic differences between the inductive and the repressive-inductive conditions. A major drawback for a more extensive analysis of these transcriptomes is the lack of annotation of proteins encoded by 1/3 of the *T. reesei* genome. One of the main categories of genes which were differentially regulated between the three conditions studied here were those encoding proteins involved in carbohydrate deconstruction, transport and metabolism due to the transfer of the fungus from a simple carbon source (glucose) to a complex one (straw). *T. reesei* secreted versatile cocktail of enzymes involved in wheat straw breakdown (86 enzymes belonging to 42 different families) but the number and amount of these secreted enzymes was still predicted to be less than in *A. niger*. Both organisms secreted the majority of enzymes from the same GH families but accessory enzymes and proteins from other families varied between them. There were also slight differences between *A. niger* and *T. reesei* in the approach they use to mediate attachment of the mycelia to the substrate, how they initiate straw degradation and probably in the signal transduction mechanisms. The GH61 family proteins were amongst the highest expressed in both organisms and the three (out of six in total) GH61-encoding genes in *T. reesei* studied here are likely to be subject to different regulatory mechanisms. *T. reesei* was able to recognise and use straw as a potential carbon source by the initial secretion of "scouting" enzymes as well as attachment of the fungus to straw particles through the expression of hydrophobins (Delmas *et al*., 2012). The attachment may then act as a cue for PTH11 GPCR signalling which up-regulates the expression of genes encoding several hydrolytic and non-hydrolytic (SWO1) proteins. The "scouting" proteins, as well as cell wall associated enzymes, could then lead to production of inducer molecules from the straw cellulose and hemicellulose components and fully induce other GH and CE-encoding genes (Delmas *et al*., 2012). CRE1 also appears to play a role in ensuring correct GH gene expression and it also mediates the onset of expression of these genes. Thus the roles for CRE1 are numerous: not only does CRE1 repress a large set of genes; it also ensures correct expression of these genes and is likely to be involved either directly or indirectly in other pathways. This is in contrast to the general belief that CRE1 acts mainly as a repressor. The identification of genes which are most likely involved in straw-

mediated signalling, enzyme expression, secretion and degradation provides a basic platform for further investigation. More experiments are needed to characterise the roles of each of the hydrophobin and PTH11 GPCR-encoding genes. Furthermore, this study indicates that different regulatory mechanisms act on the GH61-encoding genes and further detailed investigation into this may clarify the specific roles of this enzyme family. This study also identified the presence of NATs for around 2% of *T. reesei* genes. These NATs present a further mechanism of regulation (mainly post-transcriptionally). Detailed experiments are required to analyse the specific role of each one of these NATs but length and abundance of these RNA transcripts can already hint at possible gene regulatory mechanisms.

Chapter 5: Nucleosome positioning within the *cbh1* promoter in *T. reesei* in inducing and repressing conditions

5.1 Introduction

5.1.1 Overview

The genome of *T. reesei* encodes 10 main cellulases (e.g. CBH1, CBH2, EGL1-7 and BGL1) and 16 major hemicellulases (e.g. XYN1-5, BXL1, BXL2, ABF1-3) which have been shown to be important for industrial applications. These genes are regulated at the transcriptional level by an interplay of 3 activators (XYR1, ACE2 and the HAP 2/3/5 complex) and 2 repressors (CRE1 and ACE1) (Portnoy *et al.*, 2011). XYR1 (xylanase regulator 1) is the main inducer for cellulase and hemicellulase-encoding genes and it is thought that ACE1 (activator of cellulases 1) and ACE2 fine tune the activity of XYR1. XYR1 is up-regulated by a variety of substrates (Mach-Aigner *et al.*, 2008) and induces different sets of enzymes depending on the carbon source present (Figure 5.1). The HAP2/3/5 complex may have a role as a general transcriptional enhancer by facilitating access of other regulators to the promoter regions. CRE1 is the main repressor of cellulase and hemicellulase genes in the presence of glucose.

XYR1, CRE1, ACE1 and ACE2 are all zinc binuclear cluster DNA binding proteins (Zn(II)2Cys6) and this class of proteins is unique to fungi. The N-terminal part of the protein contains six cysteines which coordinate two zinc ions responsible for contacting three base pairs in the major groove of the DNA double helix (Todd *et al.*, 1997). Thereby, zinc binuclear cluster proteins are involved in regulating the transcription of genes. Furthermore the C-terminal domain of these proteins can have a leucine zipper-like heptad motif which is involved in protein-protein interaction and dimerization (Todd *et al.*, 1997). A linker region between the C-terminal and N-terminal domains is thought to confer at least part of the DNA binding specificity through recognising the spacing between the 3 nucleotides.

Genomic DNA of eukaryotic cells is compacted by a factor of c. 10,000 by conserved histone proteins to form a highly organised structure termed chromatin. This ensures that chromosomes are packed properly into the nucleus and thus presents an additional level of gene regulation in eukaryotes. Chromatin consists of several layers of organisation (Figure 5.2)

and it is a highly dynamic structure. The basic unit of chromatin is the nucleosome consisting of a core histone octamer and associated DNA. Approximately 147 bp of DNA are wrapped around these histones in 1.75 left-handed superhelical turns and the nucleosome repeat length is 159 bp +/- 7 bp in filamentous fungi (Scazzocchio and Ramón, 2008). An H3/H4 tetramer is the core of the nucleosome and serves as a scaffold for two H2A/H2B dimers. Changes in chromatin structure are accompanied by short- or long-term alterations in transcriptional activity and can be triggered by intrinsic cellular programs and/or environmental factors (Brosch *et al.*, 2008). Furthermore, chromatin is essential for genome stability and promoters, genes and termination regions have characteristic signatures of histone modifications, histone replacements and nucleosome positions (Cockerill, 2011). Two types of chromatin can be distinguished: euchromatin describes a more lose interaction between DNA and histone proteins and is usually associated with transcribed regions of chromosomes; whereas heterochromatin is a highly compacted structure and is associated with gene repression as well as telomere and centromere structures. Changes in chromatin structure can be triggered by various different factors: 1) chromatin can undergo ATP-dependent remodelling and can include the exchange of histone variants, 2) histones are subject to posttranslational modifications with structural and functional consequences, 3) DNA itself can be modified by methylation, 4) enzyme complexes that use DNA for transcription, replication and recombination (e.g. transcription factors, RNA/DNA polymerases) can contribute to chromatin modifications and 5) non-coding RNAs also affect chromatin structure. These chromatin modifications are catalysed by large enzyme complexes and these different modifications are interlinked or act together in order to ensure proper restructuring of chromatin.
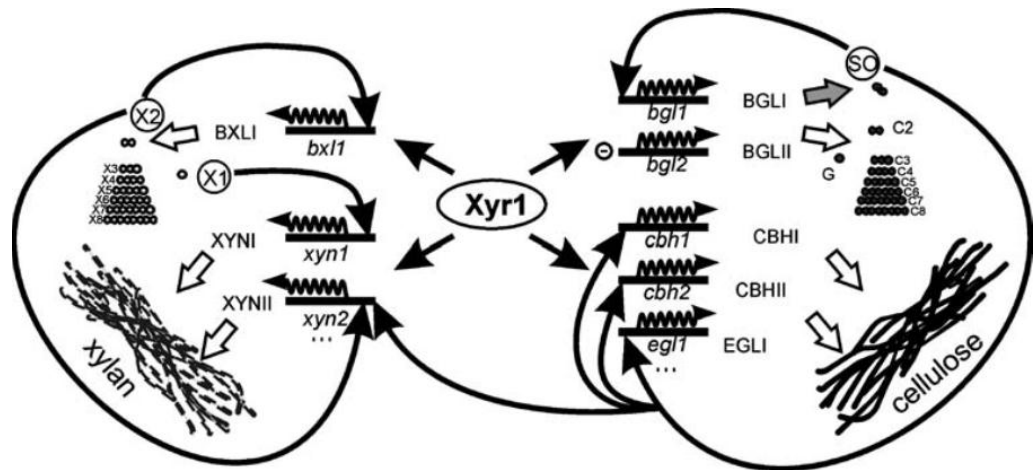
Figure 5.1: Diagrammatic representation of the transcriptional up-regulation of several cellulase and hemicellulase genes by XYR1 and their respective carbon source (X1 = D-xylose, X2 = xylobiose, SO = sophorose, C2 = cellobiose, G = glucose). Taken from Stricker *et al*. (2008).
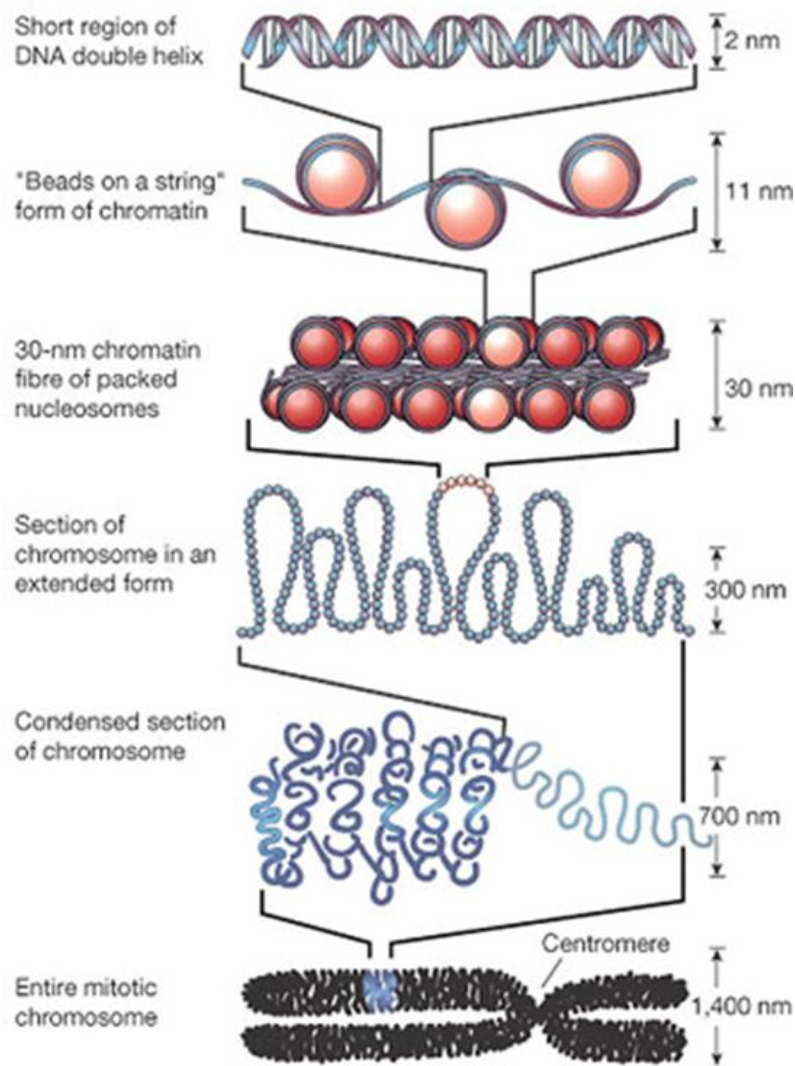
Figure 5.2: Diagram depicting the different steps and layers leading to chromatin condensation. The basic unit of chromatin is the 11 nm fibre or "beads on the string" (10-fold compaction when compared to naked DNA) which can be further compacted to the 30-nm (6 nucleosomes wound together in a solenoidal structure) fibre through histone H1 binding. Additional levels of compaction also exist and they can lead to highly condensed chromosomes as those seen during the M-phase of mitosis. Taken from Felsenfeld and Groudine (2003).

5.1.2 XYR1 – the main cellulase and hemicellulase gene inducer

The transcriptional regulation of the major cellulase and hemicellulase-encoding genes (*cbh1*, *cbh2*, *bgl1*, *xyn1-5*, *bxl1* and *egl1-7*) as well as *abf2* (L-α-arabinofuranosidase) is dependent on XYR1 in the presence of different carbon sources (Portnoy *et al.*, 2011). The exceptions for this are *bgl2* (encoding a cell wall-bound enzyme), *egl8*, *abf1* and *abf3* (Furukawa *et al.*, 2009; Akel *et al.*, 2009). It was shown that in a strain which constantly expresses *xyr1*, the transcription of all these genes is up-regulated when compared to the parental strain (Pucher *et al.*, 2011). Furthermore deletion of *xyr1* abolishes cellulase induction in the presence of inducers and also impairs hemicellulase gene transcription (Portnoy *et al.*, 2011, Furukawa *et al.*, 2009; Stricker *et al.*, 2006).

The expression of *xyr1* is mediated by carbon catabolite de-repression in the presence of different inducing substrates as no higher expression levels of *xyr1* were detected in *T. reesei* RutC30, a carbon catabolite de-repressed strain. Furthermore *xyr1* transcription is subject to CRE1 and ACE1 repression in the presence of glucose and D-xylose (Mach-Aigner *et al.*, 2008). Therefore, CRE1 can indirectly repress genes which do not possess CRE1 motifs within their promoter regions (e.g. *cbh2*). Furthermore, *xyr1* is expressed at low basal levels and induction of cellulolytic and hemicellulolytic genes, does not require *de novo* XYR1 synthesis which is also part of the carbon source recognition mechanism (Mach-Aigner *et al.*, 2010). The promoter region of *xyr1* contains 2 ACE1 and 10 CRE1 binding sites although binding of these regulators to this gene promoter has not been confirmed experimentally yet. Furthermore, no XYR1 binding sites within its own promoter were found, eliminating the possibility of auto-regulation. *In silico* analysis of the *xyr1* promoter showed an additional 4 putative ACE2 binding sites. It remains to be determined how *xyr1* expression is mediated by ACE1, ACE2 and CRE1.

XYR1 binds to a 5'-GGCTAA-motif arranged as an inverted repeat spaced by 10 to 12 bases in the promoter region of its target genes. The 2nd G within this motif has been shown to be essential for DNA binding. Furthermore, Furukawa *et al.* (2009) found that XYR1 also binds to single 5'-GGC(A/T)$_3$-3' motifs (mainly 5'-GGCTAA-3', 5'-GGCAAA-3' and 5'-GCCTAT-3'). *In silico* analysis showed that these motifs are all overrepresented within all major cellulase and hemicellulase-encoding genes. They also provided direct evidence of binding of XYR1 to these motifs within the *xyn3* promoter region *in vitro*. It remains to be determined whether this is the case *in vivo* as well. It is thought that post-translational modifications of XYR1 and dimerization are required for full expression of its target genes (Stricker *et al.*, 2008).

XYR1 also regulates D-xylose, L-arabinose and D-galactose metabolism because it activates D-xylose reductase (XYL1) by binding to two motifs in its promoter region (Stricker *et al.*, 2007). XYL1 catalyses the reduction of D-xylose to xylitol, which is converted into D-xylulose and this substrate then enters the pentose phosphate pathway (Stricker *et al.*, 2006). XYL1 also converts L-arabinose to L-arabinitol which is also reduced to D-xylulose and has the same metabolic fate as D-xylose (Akel *et al.*, 2009). Furthermore, the genome of *T. reesei* encodes 3 arabinofuranosidases (*abf1*, *abf2* and *abf3*) and *bxl1* (has an arabinofuranosidase domain) which are all involved in arabinan and L-arabinose metabolism; *abf2* and *bxl1* expression is also mediated by XYR1. XYR1 also indirectly influences *bga1* transcription by making galactitol, a substrate formed during the alternative lactose pathway (Pucher *et al.*, 2011). *T. reesei* therefore uses the same enzyme to catalyse the first step in the degradation of different hemicelluloses and has them under the control of a common regulator, which is beneficial in its habitat where different hemicellulose sugars are found in close proximity.

5.1.3 CRE1 – the main carbon catabolite repressor

Carbon catabolite repression (CCR) is a way by which cells prioritise the uptake of readily metabolisable sugars such as glucose which are of high nutritional value (Cziferszky *et al.*, 2002). In the presence of glucose, transport of inducers such as sophorose (formed by the transglycosylation activity of extracellular BGL) into the mycelia is inhibited (Zeilinger *et al.*, 2003). CRE1 is the main carbon catabolite repressor protein which represses either directly or indirectly genes involved in polysaccharide degradation as well as genes involved in the utilisation of ethanol and amino acids as potential carbon sources (Portnoy *et al.*, 2011). CRE1 is involved in the regulation of the main inducer XYR1 by repressing *xyr1* transcription in the presence of high concentrations of D-xylose hence it also indirectly influences hemicellulase gene transcription (Mach-Aigner *et al.*, 2010). Furthermore, CRE1 also plays a role in regulating fungal growth on solid substrates and the formation of aerial hyphae, conidia and fungal colonies (Nakari-Setälä *et al.*, 2009). This was shown by a deletion of *cre1* which reduced growth on several substrates such as D-xylose, D-galactose, L-sorbose and maltose. CCR is therefore dependent on the fungal growth rate. In a genome-wide study, Portnoy *et al.* (2011) showed that 250 genes are significantly differentially regulated by CRE1 of which 47.3% of those genes are repressed by CRE1 (considering different growth rates), 29% of these genes are induced by CRE1 and 17.2% of genes are regulated independently of CRE1, supporting the role of CRE1 as a cell-

wide master regulator of carbon metabolism. The genes which are repressed by CRE1 encode xylanolytic and cellulolytic enzymes, proteins involved in carbohydrate transport such as membrane permeases and those involved in transcriptional regulation such as chromatin re-modellers during high growth rates.

CRE1 has an N-terminal Zinc finger DNA-binding domain which is followed by a glutamine and histidine rich region. CRE1 is a phosphoprotein which can be phosphorylated at Ser^{241} within the acidic domain by a casein kinase II; this post-translational modification is essential for DNA binding and to ensure full repression by CRE1 (Cziferszky *et al.*, 2002). It is also thought that CRE1 binds as a dimer. The homologue of CRE1 in *A. nidulans*, CreA, is subject to a switch in compartmentation during CCR conditions, such that this protein is transported into the nucleus in the presence of glucose (Cziferszky *et al.*, 2002). The *cre1* gene is also transcribed in the presence of other (more inducing) carbon sources indicating a role in the fine tuning of gene expression.

CRE1 binds the DNA double helix as a dimer at two closely spaced 5'-SYGGRG binding motifs and has been shown to bind directly to the promoters of *cbh1* and *xyn1* whereas it indirectly suppresses the transcription of *cbh2*, *egl1*, *egl2* and *egl5* and other genes by controlling *xyr1* transcription. Functional CRE1 binding sites are arranged in tandem (*cbh1*) or inverted (*xyn1*) repeats (Mach-Aigner *et al.*, 2008). Furthermore, CRE1 represses *bga1* transcription at the basal level and is also required during induction in the presence of D-galactose and lactose (Fekete *et al.*, 2007). The genome of *T. reesei* also encodes another three CRE proteins (*cre2*, *cre3* and *cre4*) which encode a de-ubiquitinating enzyme, a WD-40 repeat protein and an arrestin and PY-containing motif protein. De-ubiquitinating enzymes are cysteine proteases and are thought to target the activation domains of certain transcription factors. Ubiquitination serves as a marker on proteins for targeting them to the proteasome. They also interact with ubiquitin ligases and together this could be an additional level of regulation by controlling the amounts of transcription factors present during carbon catabolite repression (Kubicek *et al.*, 2009).

5.1.4 Additional regulators ACE1, ACE2 and HAP2/3/5

ACE1 is a repressor and ACE2 is an activator of cellulolytic and xylanolytic genes. ACE1 and ACE2 are described as narrow domain regulators, which modulate the activity of the general activator XYR1 (Mach-Aigner *et al.*, 2008). The expression of *ace1* and *ace2* is regulated by CRE1 and by the presence of different carbon sources (Portnoy *et al.*, 2011). CRE1 is needed for *ace2* induction whereas *ace1* is subject to CCR in the presence of lactose. ACE2 is only found in *Trichoderma* spp. and no homologue has been found in any *Aspergillus* spp. or other filamentous fungi.

ACE1 contains 3 Cys_2-His_2-type zinc fingers and binds to the 5'-AGGCA motif within gene promoters. Deletion of *ace1* resulted in better growth in cellulose-rich media and increased cellulase (*cbh1*, *cbh2*, *egl1* and *egl2*) and hemicellulase (*xyn1* and *xyn2*) gene expression in the presence of cellulose and sophorose (Aro *et al.*, 2003). Furthermore, *ace1* deletion led to impaired growth on sorbitol suggesting that ACE1 is involved in the regulation of other genes. ACE1 is most similar to a protein encoded by *stzA* in *A. nidulans* which alleviates sensitivity to salt and DNA-damaging agents in this organism. The *ace1* gene is transcribed in the presence of different carbon sources and was shown to be independent of ACE2 and CCR although 11 putative CRE1 binding sites are found within its promoter region (Aro *et al.*, 2003).

ACE2 binds to the 5'-GGCTAATAA-3' motif of gene promoters and was shown to bind *in vitro* to the *cbh1* promoter. It also positively regulates *cbh2*, *egl1*, *egl2*, *bgl2* and *xyn2* in cellulose-induced conditions (Aro *et al.*, 2003; Verbeke *et al.*, 2009). DNA binding motifs for ACE2 and XYR1 are very similar which suggests interplay of these two factors to mediate gene induction. ACE2 also has an N-terminal zinc binuclear cluster domain followed by a histidine rich region and this protein is 341 amino acids long (38 kDa) (Aro *et al.*, 2001). Deletion of *ace2* results in a lower growth rate and decreases the levels of cellulase and hemicellulase expression but does not fully abolish them in the presence of cellulose (Aro *et al.*, 2001). The reduced levels of cellulase expression in an *ace2* mutant strain are greater during the early stages of induction by cellulose indicating that ACE2 increases the induction rate. Furthermore an *ace2* deletion has no effect on the induction of cellulolytic and xylanolytic genes in the presence of sophorose suggesting that slightly different mechanisms operate during cellulose and sophorose-mediated gene induction. ACE2 is also likely to be phosphorylated by a Ca^{2+}/calmodulin-dependent protein kinase and needs to dimerise in order to bind to promoter regions of genes such as *xyn2* (Portnoy *et al.*, 2011).

The HAP2/3/5 protein binds to the 5'-CCAAT-3' motif in cell-free lysates. This motif is found within many gene promoters including *cbh2*, *xyn1* and *xyn2* and the CCAAT sequence is one of the most ubiquitous motifs in eukaryotic promoters underlining its importance in creating accessibility to promoters. There is a HAP2/3/5 complex binding site within the *cbh1* promoter at around -700 bp (from the start codon) but it is not known in which conditions the complex binds (if at all) to this promoter; and deletion of CRE1 binding sites and introduction of multiple HAP2/3/5 and ACE2 binding motifs was shown to increase *cbh1* transcription even in repressing conditions (Liu *et al.*, 2008). The HAP2/3/5 complex was shown to bind to the *cbh2* promoter where it keeps an open chromatin structure even in the presence of glucose. The three HAP proteins are homologous to the mammalian NF-Y (nuclear transcription factor Y) complex which is thought to have acetyltransferase activity. Therefore the CCAAT sequences in the promoters of cellulase genes are thought to play a role in generating a more open chromatin structure thus allowing gene transcription to take place.

5.1.5 Regulation and promoter architecture of the main cellulolytic (*cbh1*, *cbh2*) and xylanolytic genes (*xyn1*, *xyn2*)

The homologue of XYR1 in *A. niger* is XlnR, induced by xylose, which co-ordinates the expression of all major cellulase and hemicellulase genes in this organism. This is not the case in *T. reesei* where XYR1 is up-regulated in the presence of different substrates such as sophorose, lactose, xylose or xylobiose and does not induce the full set of cellulase and hemicellulase-encoding genes (Stricker *et al.*, 2008). There are differences in the time-course of induction of these genes but also in the extent and level of response to various different inducing molecules (Verbeke *et al.*, 2009). In contrast to XlnR of *A. niger*, XYR1 needs to bind as a dimer to ensure full gene induction. The transcriptional regulation of cellulase and hemicellulase-encoding genes in *T. reesei* thus requires finer tuning and tighter control when compared to *A. niger*. In the next two sections, the regulation at the promoter level of two main cellulase (*cbh1* and *cbh2*) and two main xylanase (*xyn1* and *xyn2*)-encoding genes will be described as most research on transcriptional regulation mechanisms has been carried out in these genes due to their importance for industrial applications.

5.1.5.1 Regulation and promoter architecture of *cbh1* and *cbh2*

The major cellulase genes are co-ordinately expressed by various different inducing substrates, suggesting the presence of a common regulatory

mechanism (Furukawa *et al.*, 2009). The regulation of the genes encoding the main cellulases is subject to interplay of several different factors as has been found by various different studies.

CBH1 is the most abundant secreted protein in *T. reesei* when incubated in the presence of an inducing substrate such as sophorose. Regulation of *cbh1* has been studied in great detail at the transcription factor level (see Figure 5.3) but this has not been linked to any accompanying changes in chromatin structure. In the case of *cbh1*, it was shown that CRE1, XYR1, ACE1 and ACE2 but not the HAP2/3/5 complex bind to specific sites within the *cbh1* promoter region. The promoter region of *cbh1* contains 4 CRE1 binding sites (Takashima *et al.*, 1996) at -691 bp, -699 bp, -725 bp and -1154 bp (relative to ATG); 5 putative ACE2 binding sites at -209 bp, -513 bp, -779 bp, -803 bp (*in vitro*) and -829 bp (Aro *et al.*, 2001); 7 sites which have been shown to bind ACE1 in *vitro* at positions -152 bp, -254 bp, -298 bp, -392 bp, -402 bp, -907 bp and -1034 bp (Saloheimo *et al.*, 2000; Ling *et al.*, 2009) and two XYR1 binding sites at -320 bp and -733 bp. All binding sites are indicated as positions relative to the ATG translational start site. The *cbh1* promoter also has two activities: basal and induced (El-Gogary *et al.*, 1989). In the absence of an inducer low basal transcript levels of *cbh1* are produced which can act on insoluble substrates such as cellulose. Two *cis*-acting elements were identified in the promoter region which are thought to mediate induction by cellulose and basal expression of this gene: the sequence which mediates basal expression is located next to the TATA box at -135 bp to -204 bp; adjacent to this, is the region mediating induction by cellulose at -205 bp to -374 bp (Henrique-Silva *et al.*, 1996) and which contains 1 XYR binding site, 1 ACE2 binding site and 2 ACE1 binding sites (red box in Figure 5.3). Furthermore, the region which mediates basal expression also partially overlaps with the region associated to sophorose induction located between -1 bp to -162 bp (green box in Figure 5.3) and contains the TATA box and on ACE1 binding site. This suggests that sophorose and cellulose induce *cbh1* via different molecular mechanisms (Mach and Zeilinger, 2003).

Only CRE1 dimer binding has been shown to confer *cbh1* repression and there are 3 closely spaced CRE1 binding motifs at -700 bp in this promoter. Deletion of the CRE1 binding sites within the *cbh1* promoter increases its expression levels although not to the same levels than when cellulose or sophorose are present (Nakari-Setälä *et al.*, 2007). This suggests that a further inducer is required for full *cbh1* induction. It has also been proposed that the presence of low basal levels of *cbh1* expression is different from the mechanism which governs full gene induction (Torigoi *et al.*, 1996). Elevated levels of XYR1 expression also increased transcription of *cbh1* indicating the requirement of

this protein for gene induction (Pucher *et al.*, 2011). The exact role of the 8 putative ACE1 binding sites in *cbh*1 regulation remains largely unknown although all sites are able to bind ACE1 *in vitro*. In the presence of glucose, ACE1 was shown to bind to AGGCA motifs at -152 bp, -254 bp and -281 bp in the *cbh1* promoter region thus helping to mediate repression by glucose (Ling *et al.*, 2009). Deletion of *ace2* was shown to diminish *cbh1* expression levels in the presence of cellulose but not sophorose (Stricker *et al.*, 2008).
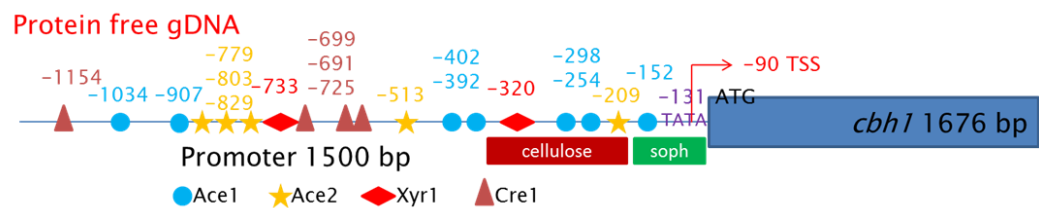
Figure 5.3: Diagram of the *cbh1* promoter region showing all putative regulatory factor binding sites, the TATA box and TSS (transcription start site) as well as the regions required for induction by cellulose and sophorose (red and green box).

The *cbh2* promoter contains a CAE (*cbh2* activating element) at position -295 bp to -193 bp relative to the translation start codon ATG. This CAE has two binding motifs: one for the HAP2/3/5 complex and one for an unidentified protein (Mach *et al.*, 2003). These two protein binding motifs are adjacent to each other: a GTAATA box (resembles ACE2/XYR1 binding motifs) on the coding strand and a CCAAT box (HAP2/3/5 binding) on the non-coding strand and mutations of these motifs showed a reduction of *cbh2* expression in the presence of sophorose. The mutation within the ACE2/XYR1 binding motif showed a more severe reduction of *cbh2* transcript levels (55%) suggesting that ACE2 and/or XYR1 also act as inducers within this promoter in the presence of sophorose and cellulose (Zeilinger *et al.*, 2003). The HAP2/3/5 complex remains bound to this motif during inducing and repressing conditions, keeping the CAE nucleosome free and possibly interacting with other proteins required for nucleosome dislocation or transcription initiation. There is also a CRE1 binding site at -222 bp but this has not been investigated *in vivo* yet and basal expression of *cbh2* seems to be unaffected by CCR (Zeilinger *et al.*, 2003). The analysis of chromatin structure with respect to gene induction and repression in *T. reesei* has so far only been elucidated for

the *cbh2* promoter region. The CAE within *cbh2* promoter region is nucleosome-free, flanked by five well positioned nucleosomes at positions -1, -2, +1, +2 and +3 (downstream and upstream from the CAE) during repressing conditions. Upon *cbh2* induction, nucleosomes -1 and -2 are displaced, freeing the TATA box (Zeilinger *et al.*, 2003). The CCAAT box is contacted by the HAP2/3/5 complex whereas the GTAATA motif is bound by an unknown protein but which is likely to be ACE2 or XYR1. Zeilinger *et al.* (2003) showed that both CAE-contacting proteins and CRE1 bind to the *cbh2* promoter under repressing and inducing conditions and that all three of them are required for correct nucleosome positioning (as loss of these proteins results in a random nucleosome positioning pattern) and that HAP2/3/5 complex is involved in keeping the CAE region nucleosome free. They further proposed that the two CAE binding proteins are possibly involved in binding to another protein or protein complex which in turn interacts with nucleosome remodelling complexes and the transcriptional machinery.

5.1.5.2 Regulation and promoter architecture of *xyn1* and *xyn2*

The transcription of hemicellulase-encoding genes is also mediated by different carbon sources but, in contrast to cellulase-encoding genes, the expression of xylanolytic genes does not appear to be co-regulated. Cultivating *T. reesei* on xylan, xylose, lactose or cellulose induces high levels of XYN1, XYN2 and EGL1; growth in the presence of sophorose induces XYN2 and EGL1 and in the presence of xylobiose XYN1, XYN2 and BXL1 are expressed (Stricker *et al.*, 2008; Mach-Aigner *et al.*, 2008). Only *xyn1* has been shown to be induced by xylose when compared to the other cellulase- and hemicellulase-encoding genes.

It was shown that a 214 bp fragment between -321 bp to -534 bp (relative to the translational start codon) within *xyn1* contains all the information necessary for *xyn1* regulation. The consensus sequences required for ACE1/ACE2 and XYR1 binding are very similar in the *xyn1* and *xyn2* promoters and it has been shown that ACE1 and XYR1 compete for binding sites within the *xyn1* promoter (Stricker *et al.*, 2008). This establishes a role for ACE1 in the fine tuning of the functionality of XYR1 within the *xyn1* promoter region. XYR1 is bound within the *xyn1* promoter during all conditions studied, but as dimerization is required for full gene induction, the binding of a XYR1 monomer and binding of the ACE1 protein (does not need to dimerize to be fully functional) may mediate gene repression. Support for this theory would be the requirement of post-translational modifications (e.g. phosphorylation) on XYR1 which would activate this regulator (by dimerization) and thus induce

xyn1 (Mach-Aigner *et al.*, 2008). Furthermore, CRE1 also binds to *xyn1* in the presence of glucose, repressing its transcription.

In the *xyn2* promoter region, the presence of a 55 bp fragment contains all the information necessary for its regulation. The induction as well as the basal expression of *xyn2* is mediated via the presence of an undecameric motif 5'-GGGTAAATTGG-3' named XAE (xylanase activating element) within this 55 bp fragment. Adjacent to XAE is an inverted sequence repeat separated by approximately 12 bases and which supports the binding of ACE2, the HAP2/3/5 complex and XYR1. Both motifs also show a striking resemblance to an inverted version of CAE found within the *cbh2* promoter. ACE2 can bind to the XYR1 motifs within the *xyn2* promoter and was shown to be essential for basal *xyn2* transcription. ACE2 also acts in a dual role: ACE2 is an antagonist during early induction and then acts as an enhancer of a continuous extension of *xyn2* gene expression but not in the presence of sophorose (Stricker *et al.*, 2008). XYR1 is indispensable for *xyn2* induction and low basal transcription (Stricker *et al.*, 2006). Furthermore, ACE2 and the HAP2/3/5 complex bind during inducing and repressing conditions. This induction complex is counteracted by a yet unidentified DNA-binding protein (complex) binding to the 5'-AGAA-3' upstream of XAE in the presence of glucose and which is thought to be involved in *xyn2* repression. Mach-Aigner *et al.* (2010) isolated this protein of 55 kDa which they named Xpp1 (Xylanase promoter binding protein 1). Xpp1 is likely to be an E-box binding HLH (helix-loop-helix) protein as it contains a C-terminal helix-loop-helix region required for dimerization and an N-terminal DNA binding domain (two alpha helices). Furthermore, the gene encoding Xpp1 is up-regulated in repressive carbon sources and down-regulated in the presence of inducers. ACE1 does not bind to the *xyn2* promoter, whereas ACE2 has no binding consensus within the *xyn1* promoter. It is thought that interplay of XYR1 and ACE2, including phosphorylation steps and hetero-dimerisation is required for *xyn2* expression in the presence of inducers such as xylobiose (Stricker *et al.*, 2008).

There is also a difference in the transcriptional response to glucose of *xyn1* and *xyn2* (Mach *et al.*, 2003). It was demonstrated that *xyn2* is transcribed at a low basal level in the presence of glucose and up-regulated with the addition of xylan, xylobiose or sophorose (Stricker *et al.*, 2007). On the other hand, glucose abolishes the expression of *xyn1* through the binding of CRE1 at two consensus sequences. It is likely that more proteins exist which mediate this interplay but which have not been identified yet.

In eukaryotes such as *T. reesei*, transcriptional regulation of a gene can be mediated by the binding of different factors to multiple binding sites within

the promoter region. Thus the synergistic action of repressors and activators can either modulate the levels of gene expression or completely switch them on or off. In *T. reesei*, cellulase and hemicellulase gene expression is governed by the balance of positive and negative regulators such as ACE1, ACE2, the HAP2/3/5 complex and CRE1; and that they also interact with the essential, positive regulator XYR1. The effect of the interplay of all these factors is graded so that they are individually or together only partially responsible for complete regulation of cellulolytic and xylanolytic genes (Aro *et al.*, 2003). Furthermore, the ratios at which cellulase- and hemicellulase-encoding genes are expressed also vary due to different promoter strengths. The gene with the strongest promoter in *T. reesei* is *cbh1* and CBH1 can make up 65% of the secreted enzyme cocktail under production conditions (Seiboth *et al.*, 2007). Disrupting the coding region also diminished expression suggesting that an intact gene is required for transcriptional regulation (Rahman *et al.*, 2009).

5.1.6 Nucleosomes – the basic units of chromatin

Chromatin is itself a highly repressive structure and the process leading to transcriptional activation involves the ordered recruitment of factors that cooperatively act together to ensure unravelling of the chromatin structure. Changes in chromatin structure are initiated by DNA-binding proteins which scan the chromatin fibre in search of their preferred recognition sequences. Nucleosomes, the basic units of chromatin, can provide platforms for regulatory factors and other chromatin remodelling enzymes. Nucleosomes are composed of a core histone octamer (H3/H4 tetramer and H2A/H2B tetramer) and associated DNA. The DNA double helix contacts the histone surface at 14 sites with clusters of hydrogen bonds and salt links (Becker, 2002).

In filamentous ascomycetes, core histone (H2A, H2B, H3 and H4)-encoding genes contain, in contrast to other eukaryotes, introns. *T. reesei* also possess additional core histone variants: H2A.Z, H3.3v, H4v.1 and H4v.2. Eukaryotes use chromatin assembly factors (CAF) and RCAF (replication-coupled assembly factor) to guide histones prior to their assembly into nucleosomes at sites which are nucleosome free; as is the case after replication (Borkovich *et al.*, 2004). In yeast, positioning of nucleosomes along the genome occurs in the wake of DNA replication and is catalysed by CAF-1 in conjunction with the H3/H4 histone chaperone ASF1 (Tolkunov *et al.*, 2011). After initial deposition, nucleosomes can then undergo replication- and transcription-independent histone exchange in coding and promoter regions. Histone chaperones such as the *S. cerevisiae* ASF1, SPT6 and SPT16 proteins interact with histones and

contribute to the disassembly and reassembly of nucleosomes (Hainer and Martens, 2011). See Appendix A8 Table A.8.T.1 for a list of histone protein and histone assembly factor-encoding genes in *T. reesei*.

The amino acids on the N-terminal tails of histones are subject to several post-translational modifications: acetylation of lysines, phosphorylation of serines and threonines, methylation of lysines and arginines, ubiquitination of lysines, ADP-ribosylation of arginines and lysines, SUMOylation of lysines as well as glycosylation, carbonylation and biotinylation of different residues (Brosch *et al.*, 2008). Modified histone residues can provide docking stations for proteins which promote an open or a closed chromatin state and acetylation and methylation are the best studied modifications. The N-terminal tails of the core histone proteins can have a multitude of post-translational modifications on them, giving rise to the concept of a "histone code". It is thought that a combination of modifications creates a pattern that serves as a receptor-like docking station for regulatory proteins which recognise and interpret these signals. The "histone code" is not like the genetic code and it may actually be the wrong term to use due to the many combinatorial possibilities of these modifications. A few specific histone modification marks have been identified and linked to destabilisation of higher order chromatin structures or to an increase in repressive chromatin. Acetylation of lysine 16 on histone H4 was shown to reduce higher order chromatin structure by disrupting H4 tail secondary structure and salt bridging within adjacent stacks of nucleosomes in the 30-nm fibre and thus prevent heterochromatin spreading. Some histone modifications also have a dual function: in *S. cerevisiae* H3-K4 tri-methylation marks gene activation in euchromatic regions whereas the same mark signals an increase in heterochromatin at mating-type loci and at subtelomeric regions (Strauss and Reyes-Domingues, 2011).

Chromatin marks have also been shown to "cross-talk" between each other meaning that certain marks on the DNA and the histones influence or inhibit subsequent chromatin modifications catalysed by ATP-dependent remodelling complexes and/or proteins which establish further post-translational histone marks. These enzymes can also work together to interpret a multitude of different marks; for example DOT1 (see section 5.1.8) methylation activity is stimulated by ubiquitylated histone H2BK34 for intra-nucleosomal H3K4me3 and H3K79me2 (Fischle, 2008; Werner and Ruthenburg, 2011). In *N. crassa*, H3-S10 phosphorylation negatively affects the tri-methylation of H3-K9 (repressive) and the subsequent recruitment of DNA methyltransferases which methylate DNA cytosine residues.

An additional histone, termed H1, binds the nucleosome at the point where the DNA enters and exits the core and also binds the DNA linker region between two nucleosomes and catalyses the formation of the 30 nm fibre from the 11 nm fibre (Figure 5.2). Phosphorylation of histone H1 weakens it's interactions with DNA and relieves the nucleosome of H1-induced constraints. The length of the linker region is variable and differs between organisms and tissues within the same organism. There is only one histone H1 found to date in filamentous fungi (see Appendix A8 Table A.8.T.1) whereas higher eukaryotes have multiple H1-encoding genes. In contrast to higher eukaryotes where deletion of H1 encoding genes is lethal; knock-out of *hhoA* (histone H1 encoding gene) in *A. nidulans* did not result in any specific phenotype (Brosch *et al.*, 2008). In other filamentous fungi, deletion of histone H1 resulted in other phenotypes (none of which is lethal) indicating that this protein has different functions in different fungal species but the detail of these functions is unknown (this has not been investigated in *T. reesei* yet).

5.1.7 Nucleosome positioning

Positioning of nucleosomes is a non-random process which can modulate transcription factor binding and RNAPII (RNA Polymerase II) assembly, preventing unregulated transcription which could otherwise be detrimental to the cell. Nucleosome positioning is usually associated with gene repression whereas nucleosome displacement or eviction is linked to gene transcriptional activation. The positioning of nucleosomes is closely entwined with histone modifications and chromatin remodelling and several factors exist which dictate where and how nucleosomes are (re-) positioned. Changes in nucleosome positioning are catalysed by proteins creating post-translational modifications on histones and/or by large ATP-dependent remodelling complexes (section 5.1.8).

The nucleosome positioning landscape varies between organisms and tissues of the same organism. The nucleosome repeat length (NRL) describes the average length of DNA associated with one nucleosome and variability depends on the linker DNA between two nucleosomes (Arya *et al.*, 2010). Thus, the same tracts of DNA can have different nucleosome positioning patterns, depending on which tissues they are part of. The yeast genome has relatively short linker lengths (~20 bp) when compared to higher eukaryotes (human ~ 40 bp) and this may reflect the increased transcriptional activity observed for this genome. Gene coding regions have higher nucleosome occupancy than intergenic or non-coding regions. The pattern of nucleosome positioning at gene promoter regions allows classification into 1) open

promoter and 2) closed or occupied promoters (Arya *et al.*, 2010). Open promoters generally lack a TATA box and have a large (~150 bp) nucleosome depleted region (NDR) upstream of the transcription start site (TSS). In yeast the NDR is flanked upstream (-1 nucleosome) and downstream (+1 nucleosome) by two well-positioned nucleosomes and nucleosome occupancy increases downstream of the TSS. Occupied promoters do not contain an NDR but have positioned nucleosomes across the promoter region. These promoters also contain a TATA box which is often buried inside the nucleosome edge and is inaccessible to the RNAPII complex. In contrast to open promoters, the purpose of occupied promoters is to exclude transcription factors. Closed promoters are usually associated to genes which require strict regulation such as those required in stress response. A nucleosome also exists at the 3' termination site of many genes which is thought to facilitate the dis-assembly of RNAPII.

Choosing the location for nucleosome positioning can depend on the DNA sequence, on physical forces, on the action by different enzymes or on a combination of these factors. Around 123 bp of DNA are bent by 590° around the histone octamer core (Travers *et al.*, 2010). The nucleosome's preference for assembling on particular DNA stretches does not depend on the bases of the double helix but rather on the sequence-dependent mechanics and physico-chemical properties of the DNA (Sekinger *et al.*, 2005). The histone octamer lacks the base-specific sequence selectivity which is found in transcription factors. The close contacts between histones and DNA are mainly non-specific, involving the positively charged amino acid side chain of proteins and the negatively charged sugar-phosphate backbone on DNA. Enhanced binding of histones to specific DNA sequences can be further achieved through the intrinsic structure and/or deformability of the double helix (Xu, and Olson, 2010). DNA sequences containing the 10 bp consensus sequence $(GGAAATTTCC)_n$ and variations of it favour nucleosome positioning (Trifonov, 2010). It is thought that AA/TT/AT di-nucleotides allow the expansion of the major groove of the DNA whilst GC base-pairs facilitate the bending of the DNA (anisotropic flexibility) around the histone proteins and thus nucleosomes favour GC rich sequences (Chung *et al.*, 2010). In yeast, the NDR is composed of long poly(dA:dT) tracts which exclude nucleosomes. The CG di-nucleotides are centred at the minor grooves oriented towards the histone octamer whereas AT base pairs are orientated at minor grooves which point outwards. Furthermore, methylation and de-methylation on the CG nucleotides would allow modulating nucleosome stability (Trifonov, 2010). The temperature, salt concentration and the pH of the cellular environment also influence the properties of the double helix and therefore also influence preferential DNA-histone interactions. Sequence-dependent nucleosome

positioning is a subject of constant debate and it is likely that some nucleosomes, but not all, are positioned according to the DNA sequence *in vivo* whereas positioning of other nucleosomes depends on several other factors. Interestingly the concept of a "nucleosome-positioning code" based on the relative affinities of specific DNA sequences for histone octamers has been introduced although the existence of this code lacks conclusive experimental data. It is proposed that the nucleosome positioning code facilitates nucleosome spacing with regards to transcription factors, remodelling proteins and RNA polymerases which will influence the distribution of nucleosomes at specific gene loci (Clark, 2010). The information encoded by a certain DNA sequence important for nucleosome positioning is likely to be degenerate and specifies multiple overlapping nucleosome assembly positions. A combination of different structural and energetic parameters such as the rotational positioning and sequence periodicity signals of a specific stretch of DNA may resolve this issue.

Physical forces which could influence nucleosome positioning include statistical positioning, RNAPII binding, nucleosome packing and interaction constraints and interaction with transcription factors. The positioning of one nucleosome on a certain stretch of DNA probably restricts the location of positioning of the next nucleosome by acting as a physical barrier. Binding of RNAPII has also been shown to induce strong phasing of nucleosomes down-stream of the TSS and in the direction of transcription (Arya *et al.*, 2010). The exact mechanism underlying this is unclear but the bound pre-initiation RNAPII complex could also act as a physical barrier or recruit various chromatin remodelling factors. As most of the chromatin exists in the form of the 30-nm fibre, nucleosome positioning may also influence folding of the 11-nm fibre into higher order chromatin structures and thus nucleosome occupancy must be carefully determined in order to allow inter-nucleosomal interactions mediated by the histone tails and prevent repulsive interactions between the wound DNA in the 30-nm fibre. There could also be competition between transcription factors (TFs) and nucleosomes to bind to certain DNA sequences. Closely spaced TF binding sites could co-operate to displace nucleosomes through recognising histone modifications or DNA methylation patterns. Finally, ATP-dependent chromatin remodelling factors also play a role in evicting, displacing or re-positioning nucleosomes (see below).

An example of changing nucleosome positioning patterns between different conditions involves the *A. nidulans prnD-prnB* intergenic region. Regulation of these genes depends on a variety of signals: in the presence of proline, PrnA induces *prnD* and *prnB*; repression is mediated via CreA in the simultaneous presence of a repressing carbon (glucose) and nitrogen (ammonia) source

whereas addition of one of these substrates does not repress both genes (Scazzocchio and Ramón, 2008). In the absence of proline, 8 well positioned nucleosomes are found in the *prnD-prnB* intergenic region which are lost upon induction and a new nucleosome is partially positioned (García *et al*., 2004). Addition of repressive substrates to mycelia grown in the presence of proline results in a different nucleosome positioning pattern than the one observed under non-inducing conditions. PrnA is responsible for *prnD-prnB* transcriptional activation and nucleosome eviction upon induction by probably interacting with nucleosome remodelling factors such as HATs; and CreA is essential for positioning under repressing conditions (simultaneous presence of ammonia and glucose) by probably negating the interaction of PrnA with chromatin remodelling factors and promoting histone de-acetylation (García *et al*., 2004).

5.1.8 Enzymes involved in the modifications of the chromatin structure

5.1.8.1 Post-translational modifications

In fungi, histone acetyltransferases (HATs) and histone deacetylases (HDACs) catalyse the transfer (from acetyl-coA)/removal of an acetyl group to defined lysine residues of histones H2A, H2B, H3 and H4 and of non-histone proteins such as HMG (high-mobility group) proteins and transcription factors TFIIE and TFIIF (Graessle *et al*., 2001). HATs and HDACs are themselves subject to post-translational modification such as methylation (see below). HATs and HDACs function in large multi-subunit complexes (see section 5.1.9 for an example) and acetylation is usually associated with the loosening of the chromatin structure through the neutralisation of the positive charge of the ε-amino group of lysine whereas de-acetylation is associated with gene repression. Thus, HATs are usually associated with gene activation whereas HDACs mainly mediate gene repression. The protonated state of lysine facilitates the electrostatic binding of lysines' positively charged side-chain to the negatively charged DNA (Chichewicz, 2010). HATS can be divided into several families: the GNAT (GCN5-related N-acetyltransferase), the MYST (MOZ, YBF, SAS2, TIP60), the p300/CBP (chromatin binding protein) and the TAF (transcription-associated factor) II250 family. CBP and p300 have multiple specific interaction surfaces that can interact with several different transcription factors and proteins within the same complex (Cockerill, 2011). All HAT families contain one or two (TAFII250 proteins) bromodomains required for stable occupancy of the complex on acetylated promoter

nucleosomes. Filamentous fungi possess multiple members of each HAT family but there is a lack of data on these enzymes (Appendix A8 Table A.8.T.1 for genes encoding HAT and HDAC in *T. reesei*).

HDACs can be categorised into three families: the sirtuins, the classical HDACs and the H2-like enzymes (absent in filamentous fungi). The sirtuins are NAD$^+$-dependent HDACs and several genes encoding these enzymes exist in filamentous fungi. The classical HDACs can be further divided into 3 classes: the RPD3 (reduced potassium dependency)-type proteins (class I), the HDA1 (histone deacetylase)-type proteins (class II) and the class 4 type enzymes which are absent in filamentous fungi (Tribus *et al.*, 2010). The genome of *A. nidulans* encodes two class I (HosA, RpdA) and two class II (HdaA, HosB) HDACs. The RPD3-type enzymes possess an extension of ~200 amino acids of the C-terminal region in filamentous fungi which is essential for binding of fungal-specific complex partners and a target for post-translational modifications which are determining factors stability, localisation and activity. A negatively charged insertion of ~20 amino acids in the C-terminus is conserved amongst filamentous fungal RPD3-type HDACs and it is essential for catalytic activity of the enzyme. Cells lacking RpdA are non-viable indicating that this HDAC is essential for viability, growth and sporulation by de-acetylating histones H3 and H4 (Tribus *et al.*, 2010).

Protein methylation and de-methylation involves the transfer [from S-adenosyl-L-methionine (SAM)]/removal of a methyl-group to either lysine or arginine residues of histone and non-histone proteins. Histone lysine methyl-transferases (HKMTs) can mono-, di- or tri-methylate lysine residues and are divided into the SET [Su(var)3-9, Enhancer-of-zeste, Trithorax] or the DOT1 (disruptor of telomeric silencing) families of HKMTs. The SET family of HKMTs can be further divided into four families: the SUV39, SET1, SET2 and RIZ (absent in *Aspergillus* spp.) families and are classed according to other domains surrounding the SET domains. The SUV39 family proteins mono- and tri-methylate H3-K9 and contain two cysteine-rich regions flanking the SET domain. Knock-out of the SUV39-encoding gene (*clrD*) in *A. fumigatus* resulted in the reduction of radial growth, conidia production and delayed conidiation due to transcriptional impairment as a result of the lack of methylation within the *A. fumigatus* genome (Palmer *et al.*, 2008). The SET1 family of HKMTs (of which 7 are present in *A. nidulans* and 4 SET domain protein-encoding genes were found in *T. reesei*, see Appendix A8 Table A.8.T.1) have three distinct regions of clear homology: an RNA recognition motif, the SET domain, post-SET region (cysteine rich region) as well as a coiled-coil domain. The DOT1 family of HKMTs lacks the SET domain and

methylates H3-K79, a modification required for gene silencing at telomeres, rRNA gene loci and mating-type loci in *S. cerevisiae*.

Two classes of enzymes de-methylate lysine residues: LSD1 (lysine-specific de-methylase 1) and JmjC (jumonji C) domain containing proteins. LSD1 de-methylates H3-K4 by amine oxidation to generate unmodified lysine and formaldehyde in an FAD (flavin adenine dinucleotide)-dependent manner. LSD1 only de-methylates mono- or di-methylated residues. JmjC proteins are hydroxylases which de-methylate mono-, di- and tri-methylated residues. They can be further divided into several subclasses: JHDM1 (JmjC domain-containing histone demethylase 1A) de-methylates H3-K36, JHDM2 de-methylates H3-K9 residues and JHDM3 target H3-K9 and H3-K36 and also remove tri-methylation marks. In contrast to mammalian JHDM1 proteins, fungal JHDM1 enzymes lack an F-box domain. Two homologues, termed JHDM2 and JHDM3 are present in filamentous fungi but differ from their mammalian counterparts by additional PHD (plant homeodomain) and HTH (helix turn helix) motifs and by a lack of Tudor domains (structural motif consisting of 5 β-strands involved in recognising symmetrical dimethylated arginine residues).

Protein arginine methyltransferases and peptidyl arginine deiminases (PRMTs/PADs) catalyse the transfer (from SAM)/removal of a methyl group to arginine residues within the N-terminal tails of histones H3 and H4. They all possess a highly conserved SAM binding motif and also methylate an extensive number of non-histone proteins such as those involved in RNA metabolism (e.g. poly(A)-binding protein II). PRMTs can either mono- or dimethylate arginine residues in an asymmetrical (type 1 enzymes) or in a symmetrical (type 2 enzymes) fashion. Mammalian genomes encode 9 of these enzymes (6 class I and 3 class II) of which 2 class I (RmtA and RmtC which are homologous to human PRMT1 and PRMT3) and 1 class II (RmtB homologous to human PRMT5) enzymes are present within the genome of *A. nidulans* (Trojer *et al.*, 2004) and *T. reesei* (Appendix A8 Table A.8.T.1).

Histone kinases catalyse the phosphorylation of serines and threonines in core and linker histone N-terminal tails and these modifications have been linked to be important for chromosome condensation, signalling of active versus repressive chromatin states, transcription, apoptosis and DNA repair. Phosphorylation is a very transient process and requires the permanent presence of the modifying kinases. The genome of *N. crassa* encodes two histone kinases (and so does the *T. reesei* one, see Appendix A8 Table A.8.T.1), termed NimA and Aurora B. NimA is mainly located at the spindle pole body in filamentous fungi whereas Aurora B is the enzymatic component of the

chromosomal passenger complex found at most chromosomal locations (Fischle, 2008). The additional cell cycle kinase (NimA) also phosphorylates H3-S10, a pre-requisite for H3-K14 acetylation, a modification linked to gene activation; and H2-S10 phosphorylation, which is further required for cell cycle progression and correct chromosome segregation.

In eukaryotes, ubiquitination of proteins mainly occurs on lysine residues in histones H2A and H2B. In *S. cerevisiae* H2A-K119 and H2B-K123 can be ubiquitinated and this modification is usually associated with gene repression although ubiquitylation can also serve as gene activation marks. In yeast, the PAF-transcription elongation complex bridges the E2 and E3 ubiquitin machinery with elongating RNAPII (Werner and Ruthenburg, 2011). Histone H2B ubiquitination is thought to facilitate removal and replacement of nucleosomes during RNAPII passage by recruiting various remodelling factors.

Poly (ADP-ribose) polymerases (PARPs) catalyse poly-ADP ribosylation of glutamic acid in histones and these modifications can result in either gene expression or repression. This is by far the least well understood post-translational mechanism. At repressed loci, PARP can bind the two strands of DNA at the point at which they exit from the nucleosome thereby opposing the actions of transcriptional activators. In the presence of NAD^+ (nicotinamide adenine dinucleotide), PARP is activated and ribosylates histones and itself causing a reduction in chromatin condensation. The genome of *N. crassa* encodes a single *parp* gene. The *S. cerevisiae* HDAC SIR2 (sirtuin-type HDAC) was also reported to possess ADP-ribosylation together with de-acetylase activity (Graessle *et al.*, 2001). PARP-encoding and Sirtuin-type HDACs are also present within the genome of *T. reesei* (Appendix A8 Table A.8.T.1).

SUMOylation describes the catalysis of SUMO, an ubiquitin-like protein, to lysine residues of histone and non-histone proteins. Deletion of this protein encoded by *sumO* (see Appendix A8 Table A.8.T.1 for orthologue in *T. reesei*) in *A. nidulans* resulted in the change of secondary metabolite production. Not much else is known about this modification in filamentous fungi.

5.1.8.2 ATP-dependent chromatin re-modellers

ATP-dependent chromatin remodelling proteins have numerous functions and different families are specialised in recognising and catalysing different reactions. ATP-dependent chromatin remodelling involves the sliding of nucleosomes over distances of up to 100 bp; it includes the exchange of histone variants or the complete ejection of the histone core by breaking DNA-histone contacts. ATP-dependent chromatin-remodelling enzymes use

the energy derived from ATP hydrolysis to catalyse changes in chromatin structure, and are found in large multi-subunit complexes. Interactions with various protein partners, as well as post-translational modifications (e.g. phosphorylation) of some subunits dictate the biochemical properties of the enzyme complex. The ATPase subunit of the chromatin remodelling complexes belongs to the SNF2 (sucrose non-fermenting) superfamily of proteins and is a translocase (involving several cycles of ATP hydrolysis) which draws DNA in from one side of the nucleosome and expels it from the other. Chromatin re-modellers of the SNF2 family of ATPases can be classed into the SWI (switch)/SNF family, the ISWI (imitation switch) family, the INO80 (inositol biosynthesis) family and the family of CHD (chromodomain helicase DNA-binding) family (Marfella, and Imbalzano, 2007). Low-resolution microscopy of the *S. cerevisiae* SWI/SNF remodelling structures revealed a bowl-like shape with a central depression of appropriate size to hold a nucleosome (Flaus, and Owen-Hughes, 2011). These large remodelling complexes are proposed to envelop the entire nucleosome and control its dynamic properties; or they could cantilever across the nucleosome as was shown for the *S. cerevisiae* ISW1 protein.

The SWI/SNF family contains a bromodomain which contacts acetylated residues on histone tails and are involved in gene activation. The ISWI family members have a C-terminal SANT (Switching-defective protein 3, Adaptor 2, Nuclear receptor co-repressor, Transcription factor IIIB) domain which binds to non-modified histone tails, coupling histone binding to enzyme catalysis, and a chromodomain which recognises and binds to tri-methylated H3K4 and are associated to gene repression. The INO80 SNF2 family members are the only chromatin remodelling factors for which DNA helicase activity has been observed (Marfella and Imbalzano, 2011). The CHD family is characterised by tandem chromodomains in the N-terminal region. CHD proteins can be classed into three subfamilies of which only one (class I) is present in lower eukaryotes such as yeast and which contains an additional DNA binding domain in the C-terminal region of the protein.

Transcription factors often recruit ATP-dependent remodelling proteins required for gene activation (and repression) by binding to their recognition sequences which are exposed on the nucleosome surface, the nucleosome edge or somewhere along the higher-structured chromatin fibre. An example of this is the yeast *rec104* promoter where the repressor Ume6p recruits Isw2 (ISWI family) leading to shifts in nucleosome positioning towards the repressed state (Becker, 2002). Octamer transfer to alternative DNA sequences and partially "peeled off" DNA have been observed for SWI/SNF family proteins. ISWI as well as CHD family members are associated with

repressive nucleosome sliding in the region of active genes. In yeast, the loss of ISW2 (ISWI) resulted in an increase of non-coding transcripts due to the aberrant nucleosome positioning at the 3'-end of Isw2 target genes (Piatti *et al.*, 2011). Likewise, yeast CHD1 deletion resulted in transcription termination defects and aberrant nucleosome arrangements at the 3'-ends of the *cyc1* and *asc1* genes. The incorporation of individual histone variants also requires distinct types of ATP-dependent factors (most remain unknown) together with specialised histone chaperones (Piatti *et al.*, 2011). Incorporation of histone variants may result in different post-translational modifications, alter interactions with protein complexes or may affect the histone-DNA interactions. For example, incorporation of histone variant H2A.Z is catalysed by members of the INO80 family of chromatin remodelers such as Swr1 which recognises H2A.Z and acetylated H3/H4. In yeast, Nap1p chaperones the cytoplasmic pool of H2A.Z and together with an additional chaperone Chz1p enables H2A.Z/H2B dimer exchange.

5.1.9 Steps leading to transcriptional activation – a specific example

Transcription factor activation (through environmental signals), ATP-dependent chromatin remodelling and histone tail modifications are entwined processes which ensure, through a complex cascade of steps, changes in chromatin structure which allow gene expression and repression. Usually, a cycle of transcription commences with the recruitment of transcription factors and co-factors bound at the promoter and which modify the local chromatin structure in order to allow binding of the pre-initiation complex (Cockerill, P. N., 2011). Very often, transcription factors also co-operate to destabilise nucleosomes.

In *C. albicans*, the NUA4 complex consists of 13 subunits with EPL1, ESA1 and YNG2 making up the HAT subunit of this complex. ESA1 is the main acetylase and YNG2 aids in acetylating histone H4 on lysines 5, 8 and 12. Without YNG2, ESA1 would only be able to acetylate free histones. Two other subunits, EAF7 and EAF5, which interact with each other, are involved in another set of cellular functions possibly through the recruitment of NUA4 to distinct loci. In *C. albicans* NUA4 is recruited by the transcription factor EFG1 (it is not known though with which subunits EFG1 interacts) to target genes, such as those involved in hyphal development (Lu *et al.*, 2008). An increase in cAMP levels and subsequent protein kinase activity (see Chapter 3) are thought to mediate activation of EFG1 which binds to its target gene promoters. NUA4 mediates

acetylation of histone H4 at lysine residues 5 and 16 and an increase in acetylation triggers the dimorphic switch of *C. albicans* yeast cells to hyphae. This increase in gene transcription is mediated via the SWI/SNF complex which is recruited to the acetylated gene region by NuA4 and which can bind to acetylated residues through its' bromodomain and consequently move nucleosomes (how many are moved is not known). Furthermore, NUA4 acetylation leads to the recruitment of transcription factor II H (TFIIH) which phosphorylates serines at position 5 within the heptapeptide repeats of the C-terminal domain (CTD) of RNAPII. The phosphorylated CTD domain can then recruit directly or indirectly complexes such as SAGA, the Elongator complex (contains Elp3, a GNAT HAT) and NuA3 which all acetylate histone H3 and aid their eviction during transcription. Phosphorylation of CTD by TFIIH further promotes the recruitment of HMTs such as SET1 and MLL1 which are part of the COMPASS complex. The Elongator and COMPASS complexes can travel together with the elongating polymerase and promote histone acetylation during active transcription or mediate gene repression after transcription. The HMTs can introduce H3-K4 di- or tri-methylation marks at the 5' ends of active or recently transcribed genes. This triggers an increase in H3-K36 tri-methylation and indicates the return to a de-acetylated state of the gene region after the cycle of transcription is completed.

5.1.10 Aims

The aims of this chapter are to investigate the regulation of hemicellulase and cellulase-encoding genes by:

- Describing the expression patterns of several cellulase and hemicellulase-encoding genes in the presence of different carbon sources which may either induce or repress these genes
- Characterising protein profiles of mycelia grown in the presence of the same carbon sources as mentioned in the previous point and determine whether they agree with the observed transcriptional response
- Determining the chromatin structure within the *cbh1* promoter region during repressing conditions and investigate whether *cbh1* induction results in a different nucleosome positioning pattern
- Investigating whether CRE1 could be one of the factors involved in determining nucleosome positioning in the *cbh1* promoter region

5.2 Materials and Methods

5.2.1 Growth conditions

TCM (Chapter 2 section 2.4) supplemented with glucose or cellulose (SigmaCell type 20, Sigma), was inoculated with 10^5 *T. reesei* QM6a, RutC30, Δ*cre1* or *cre1-1* spores/mL for 24 h (QM6a) or 48 h (mutants) in glucose and 72 h in cellulose. TMM (Chapter 2 section 2.4) complemented with various carbon sources was inoculated with 10^5 QM6a spores/mL (from PDA, PCA, PGA and PXA slopes) for 24 h (glucose and cellobiose) or 36 h (xylose, glycerol and sophorose). Mycelia were cultured in 100 mL duplicates at 150 rpm at 28°C. The mycelia were harvested (Chapter 2 section 2.5) and ground to a fine powder under liquid nitrogen. RNA was extracted (Chapter 2 section 2.7) and reverse transcribed to cDNA (Chapter 2 section 2.11).

5.2.2 Real-time PCR on cellulase and hemicellulase-encoding genes

QRT-PCR was run on cDNA (see section 5.2.1) on the following genes: *cbh1*, *cbh2*, *egl1*, *bgl1* and *xyn1*, *xyn2* and *bxl1* (primer details in Table 5.1). QRT-PCR was run and set up in exactly the same way as described in Chapter 3 section 3.2.5.

Table 5.1: Primer sequences of the 7 cellulase and hemicellulase-encoding genes. Annealing temperatures and calculated product lengths of the PCR products from gDNA and cDNA are also shown

Name	Forward sequence (5' to 3')	Annealing temperature (°C)	Reverse sequence (5' to 3')	Annealing temperature (°C)	Expected product size (bp)	Expected product size (bp) no intron
cbh1	cagacaagggc ggcctgactc	60.6	cgctgtagccaa taccgccg	60.1	459	396
cbh2	ccttggccgaca tccgca	60.0	gggcaggtaat gttccacccg	59.9	589	499
egl1	tgccggttgcgg cttcaa	59.3	ccatgccgctgc tcaggg	59.7	348	278
bgl1	cgccaatatcac gccgcg	60.5	ccggctgctcgc tccca	60.4	479	415
xyn1	tccgtcaactgg tccaactcgg	60.7	tgcggcggacg gaccag	61.1	430	323
xyn2	cgggcagttctc cgtcaactg	59.2	ggacgggttgta ggtgccaaag	59.3	202	141
bxl1	ggcaaacccctt gtcgtcctg	60.2	gcctcgaaggtg aagacggga	60.0	440	440

5.2.3 Total protein concentration and CBH1 assays

After the separation of mycelia from the growth media, 50 mL of supernatant were concentrated to 2 mL using the Vivaspin6 column (Sartorius) with a 5,000 MWCO (molecular weight cut-off) filter.

To quantify the total amount of protein in each concentrated sample, a micro and standard Bradford assay were performed according to the manufacturer's instructions (Sigma). A standard curve was generated based on known bovine serum albumin (BSA, NEB) protein concentrations.

To assay for CBH1 by activity, 50 μL of a 1 M glucose solution (dissolved in 0.05M acetate buffer, pH 5) was added to 250 μL of a 1 mM substrate solution (25 mg 4-methylumbelliferyl-β-D-lactoside dissolved in 50 mL 0.05M acetate buffer, pH 5) and incubated at 50°C for 10 min. Then 200 μL of the concentrated growth medium was added and reactions were incubated at 50°C for 10 min. To stop the reaction, 1 mL of a 1 M Na_2CO_3 (dissolved in ddH_2O) solution was added and the absorbance was read at 370 nm in the Ultrospec 2000 spectrophotometer (Pharmacia Biotech, Little Chalfont,

Buckinghamshire, U.K.). A standard CBH1 curve was generated using known concentrations of *Aspergillus oryzae* Cbh1 (Sigma).

5.2.4 Separation of proteins through polyacrylamide gels

To separate proteins in concentrated samples, 1.5 mg/mL (quantified by Nanodrop spectrometry, see Chapter 2 section 2.6) of sample together with 10 µL loading buffer (Invitrogen) and 2 µL of 0.5 M DTT (Invitrogen) was loaded onto a 4-20% Tris-glycine Novex SDS-PAGE protein gel (Invitrogen) and run for 3 h at 100 V in 650 mL 1 x running buffer (10 x stock solution: 30.2 g Tris base, 188 g glycine, 50 mL of a 20% SDS solution). Samples were heated at 100°C for 5 min prior to loading. The protein gel was then silver stained.

5.2.5 Silver staining of protein gels

Protein gels were first fixed two times for 15 min in a 50 mL solution containing 5% (v/v) acetic acid and 20% (v/v) methanol. They were then sensitised for 30 min in a 50 mL solution containing 15% (v/v) methanol, 0.1 g sodium thiosulfate and 3.4 g sodium acetate. Gels were washed 3 x 5 min in ddH$_2$O and then silver stained for 20 min in a 0.1% (w/v) silver nitrate solution. Gels were washed 2 x 1 min in ddH$_2$O and then developed in a 100 mL solution containing 2.5 g Na$_2$CO$_3$ and 0.04 % (v/v) formaldehyde until all bands were clearly visible. Development was stopped by placing gels in a 50 mL 1.5% (w/v) EDTA solution for 10 min. Gels were washed another 3 x 5 min with ddH$_2$O. Bands of interest were cut out with a scalpel, stored in 500 µL ddH$_2$O at 4°C before being examined by mass spectrometry.

5.2.6 Protein sample sequencing

All sequencing was carried out by Susan Liddell (Proteomics lab, Division of Animal Sciences, Sutton Bonington Campus, Loughborough, Leicestershire, U.K.).

5.2.6.1 Sample processing and digestion by trypsin

Excised gel fragments were diced into cubes with the ProteomeWorks Mass PREP robotic liquid handling station (Waters, Elstree, U.K.) and then incubated 3 times for 10 min in 100 µL de-stain solution (50 mM ammonium bicarbonate, 50% acetonitrile). The samples were then dehydrated in 50 µL

acetonitrile for 5 min, then acetonitrile was removed by pipetting and evaporation was allowed for 10 min. Samples were then further processed for 30 min in 50 μL reducing solution (10 mM DTT, 100 mM ammonium bicarbonate) and then for 20 min in 50 μL alkylation solution (55 mM iodoacetamide, 100 mM ammonium bicarbonate). Washes were carried out for 10 min in 50 μL 100 mM ammonium bicarbonate, then for 5 min in 50 μL acetonitrile and samples were dehydrated with two consecutive washes and evaporation for 5 min each in acetonitrile. Samples on the microtitre plate were cooled down to 6°C for 10 min before the addition of 25 μL per well of trypsin gold (Promega) which was diluted to 10 ng/μL in trypsin digestion buffer (50 mM ammonium bicarbonate) prior to addition to the samples. The microtitre plates were then incubated at 6°C for 15 min then at 40°C for 4 hours. Unless otherwise specified all steps were carried out at room temperature.

5.2.6.2 Mass spectrometry (MS)

Sample sequencing was carried out in a LC-ESI-tandem MS on a Q-TOFII mass spectrometer with a nanoflow ESI source (Waters Ltd). All peptides from the trypsin digest were separated on a C18 reverse phase, 75 μm i.d., 15-cm column (made in-house, Dr. David Tooth) and fed to the MS via a CapLC HPLC system (Waters Ltd). A capillary voltage of 3000 V was applied in the positive ion mode with argon as a collision gas. Sequencing data were generated by using an automated data-dependent switching between MS and MS/MS scanning method which was based on ion intensity, mass and charge state (DDA[TM] data directed analysis). Doubly, triply and quadruply charged peptides were recognised for fragmentation by MassLynx 4.0 software and up to 4 precursor masses were chosen at a time for tandem MS. Collision energy was between 15 eV and 55 eV depending on charge and mass of each peptide.

5.2.6.3 Database searching

Non-interpreted MS data were put into pkl (peak list) files and the software ProteinLynxGlobalServer version 2.0 (Waters, Ltd) was used to search pkls against NCBInr and Swissprot databases using the web version of the MASCOT MS/MS ions search tool (http://www.matrixscience.com). Variable modifications of carbamidomethylation of cysteine and oxidation of methionine were included as well as semi-tryptic digestion and additional modifications depending on the search. All remaining search values were the preset defaults except for file type (micromass pkl) and instrument type (ESI-QUAD-TOF). Only search results with a probability above a threshold of $P < 0.05$ were accepted.

Protein sequencing can sometimes produce mass spectra which are not very clear due to large hydrophobic proteins for example, and thus need *de novo* searching and interpretation. The order of two amino acids can be switched due to two residues being "joined together" on the spectrum given by the BioLynx Peptide Sequencing report. Furthermore, leucine and isoleucine are isobaric and cannot be distinguished in this type of analysis and usually alignments with other known sequences will confirm the type of amino acid. The same is true for glutamine and lysine as well as phenylalanine and oxidised methionine. Consecutive glycine residues as well as glycine-alanine also do not produce clear cleavages between them for sequencing; glycine-glycine is isomeric to asparagine, glycine-alanine is isomeric to glutamine and isobaric to lysine. Considering all these facts about sequencing by ESI tandem MS, this leaves room for interpretation of a given sequence. Furthermore, re-sequencing of the more obscure peaks produced by weaker peptides as well as analysing the blank runs will also help to determine a certain protein sequence. On the other hand iodoacetamide treatment during processing produces well defined cysteine derivatives and cleavage due to enzymatic digestion adjacent to proline residues produces abundant y-type ions which are also easily identifiable for sequencing.

5.2.7 *cbh1* probe

The *cbh1* probe for Southern hybridisations was generated by PCR amplification from *T. reesei* QM6a gDNA and is primarily derived from the 3' end of the *cbh1* gene (Figure 5.2.1). Primers were the same as in Table 5.1. PCR conditions were the same as described in section 2.8 of Chapter 2 and 3 x 50 μL reactions were run for 35 cycles. The PCR fragment had a size 459 bp. Product size was confirmed on a 1% (w/v) agarose gel (Chapter 2 section 2.9) before the probe was purified with the Geneflow (Staffordshire, U.K.) "Gel extraction and PCR purification kit" following the manufacturer's instructions. Probe concentration was quantified by Nanodrop spectrometry.
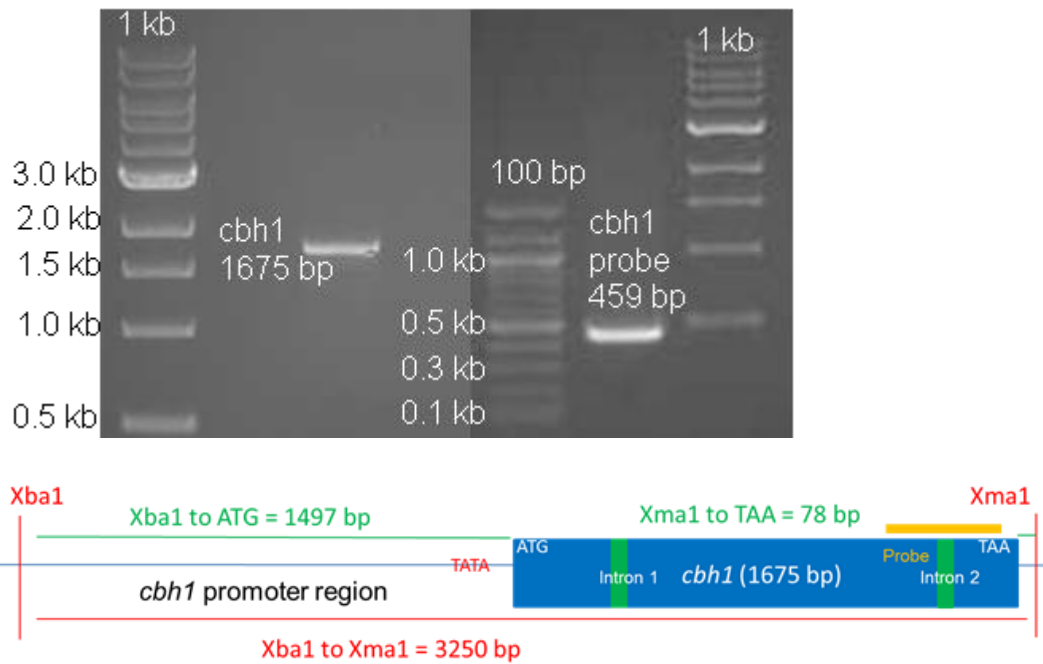
Figure 5.2.1: Gel picture showing the *cbh1* gene and corresponding probe, both amplified by PCR on gDNA (1 kb and 100 bp both refer to molecular marker ladders). Below, a diagram of the *cbh1* gene architecture with restriction sites, fragment lengths and probe binding site (yellow) is also shown.

5.2.8 Nucleosomal assay using microccocal nuclease (MNase)

Mycelia were cultivated for 24 h (wild-type) or 48 h ($\Delta cre1$, *cre1-1*) in 100 mL TCM + 2% glucose, 36 h (wild-type) or 48 h ($\Delta cre1$, *cre1-1*) in 100 mL TCM + 2% (w/v) xylose or 2% (w/v) cellobiose or 2% (w/v) glycerol and 0.025% (w/v) sophorose; or for 72 h in 100 mL TCM + 0.25% (w/v) cellulose (Sigmacell type 20, Sigma). Media inoculation was as described in section 2.4 (Chapter 2) and cultivation for each strain and condition was carried out in duplicate. Mycelia were harvested (Chapter 2 section 2.5), ground to a fine powder under liquid N_2 and weighed out into samples of 200 mg. Samples were kept on ice at all times. Stop buffer (40mM EDTA, 2% SDS) was prepared and DTT was added to MNase digestion buffer (250 mM sucrose, 60 mM KCl, 15 mM NaCl, 0.05 mM $CaCl_2$, 3 mM $MgCl_2$, 15 mM Tris-HCl pH 7.5) to a final concentration of 0.5 mM. The 200 mg samples were then re-suspended in 1 mL MNase digestion buffer and vortexed until no clumps were visible. MNase (Fermentas, Sankt Leon-Rot, Germany) was diluted to 1 U/μL in MNase digestion buffer and 0 U, 25 U, 30 U and 35 U MNase were added to one re-suspended 200 mg mycelial

sample, vortexed and incubated at 30°C for 5 min. Reactions were stopped by the addition of 200 µL stop buffer, vortexed and placed immediately on ice.

DNA was then extracted as described in section 2.6 (chapter 2), starting with the first phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) step. DNA pellets were re-suspended in 200 µL ddH$_2$O on ice for 1 h and RNA was degraded with 50 mg/mL RNAse A [New England Biolabs, (NEB), Ipswich, MA, U.S.A.] at room temperature for 1 h.

Samples were then subjected to a restriction digest at 37°C overnight with the following components: 3µL *Xba*1, 6 µL *Xma*1, 10 µL Buffer IV and 4 µL 10 mg/mL BSA (all components were obtained from NEB). After the digest, DNA was precipitated with 0.1V sodium acetate (1M, pH 5.2) and 0.7 V of 100% ethanol by centrifugation for 10 min at RT. Pellets were washed with 70% (v/v) ethanol, air-dried at RT for 10 min then re-suspended in 50 µL TE buffer at 4°C for 30 min.

5.2.9 MNase digestion of gDNA

Extracted gDNA (Chapter 2 section 2.6) was treated with MNase as described in section 5.2.8 with the exception that the phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) steps were skipped.

5.2.10 Southern blotting by Vacuum

The MNase-treated samples were run on a 1% (w/v) agarose gel at 200 V for 3 h. The DNA was then acid nicked in 1 L 0.25 M HCl (Fisher Scientific) for 20 min and washed for 10 min in 1 L ddH$_2$O. The gel was placed into 1 L of a 1.5 M NaCl, 0.5 M NaOH solution for 45 min before it was placed into the vacuum device. All steps were performed at 100 rpm at room temperature.

In the meantime the Zeta-Probe® GT Genomics membrane (BioRad) was cut to 20 cm x 26 cm membrane and soaked in ddH$_2$O for 5 min then soaked in 1.5 M NaCl, 0.5 M NaOH for 5 min. To transfer the DNA by Vacuum from the gel to the membrane the Pharmacia Biotech VacuGene XL Vacuum Blotting Unit (Pharmacia) with the Vacuum Regulator Pump (BioRad) were used. The porous screen was rinsed in ddH$_2$O and placed inside the vacuum base unit. The membrane was placed on the screen and covered with the polyethylene mask and then with the frame. Gel was placed upright onto the membrane and the vacuum pump turned on. The device was clamped into place and the DNA transferred for 1 h at 50 mbar.

After the transfer, the excess liquid was drained off and the frame and mask removed. The membrane was removed and excess liquid blotted off on chromatography paper (Fisher Scientific). The membrane was then neutralised in 500 mL 2 x SSPE (20 x stock: 3.6 M NaCl, 0.2 M NaH_2PO_4, and 0.02 M EDTA pH 7.7). DNA was then UV cross-linked (120 mJ/cm^2) to the membrane. The membrane was dried on Whatman paper, wrapped in Saran foil and stored at room temperature until hybridisation.

5.2.11 Pre-hybridisation and hybridisation of Southern blots

First, 800 μL of 10 mg/mL DNA-MB grade salmon sperm DNA (Roche, Welwyn Garden City, Hertfordshire, U.K.) was boiled for 5 min, then cooled down on ice for 30 s before being added (200 μg/mL) to 50 mL of pre-hybridisation buffer [6 x SSPE, 1 % (w/v) SDS and 5 x Denhardt's (100 x stock in 50 mL ddH$_2$O: 1 g Ficoll 400, 1 g PVP 360, 1 g BSA Fraction V)]. Membrane was transferred to Hybaid tube (Bibby Sterilin) and pre-hybridisation buffer was added for 4 h to 5 h at 65°C with rotation.

Then 45 ng of *cbh1* probe and 5 ng of 1 kb ladder (NEB) were diluted in a total volume of 14 μL ddH$_2$O and boiled for 10 min. Denatured DNA was cooled on ice for 30 s and liquid spun down at 15,000 x g for 15 s before 4 μL of labelling enzyme ("High Prime end-labelling kit", Roche) and 10 μCi of α^{32}P dCTP (Elmin-Parker) were added. Reactions were incubated at 37°C for 10 min before probe was purified through Micro Bio-Spin 30 columns (Bio-Rad). Purified DNA was added to 450 μL salmon sperm gDNA and boiled for 4 min. Labelled DNA and fish DNA (150 μg/mL) were then added to 30 mL hybridisation buffer [5 % (w/v) dextran sulphate sodium salt (Fisher Scientific), 6 x SSPE and 1 % SDS). After discarding the pre-hybridisation buffer, the hybridisation buffer with the labelled DNA and fish DNA was added to the blot for overnight hybridisation at 65°C with rotation. Steps described in sections 5.2.8, 5.2.10 and 5.2.11 are summarised in Figure 5.2.2.
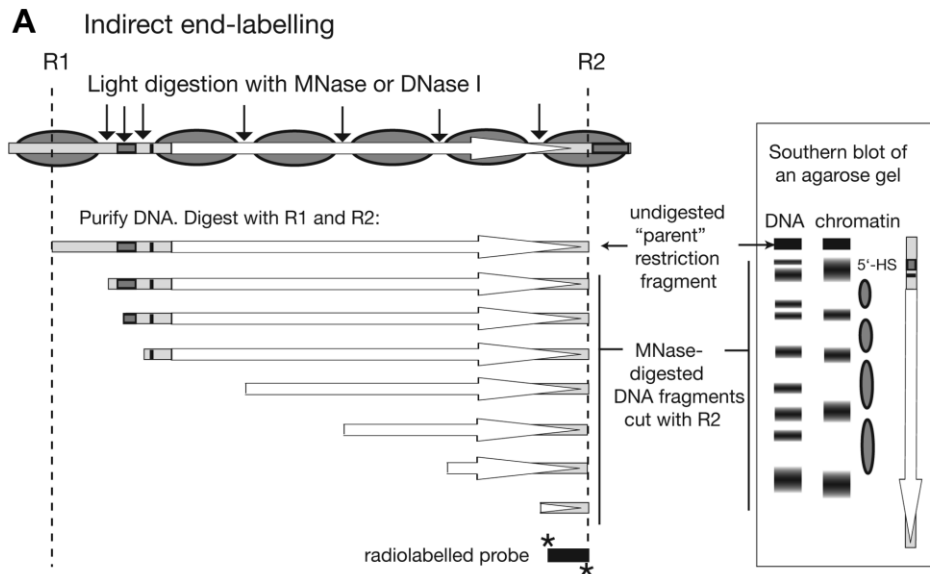
Figure 5.2.2: Schematic representation of the steps describing the indirect end-labelling method used for chromatin analysis. Taken from Clark (2010).

5.2.12 Blot washes, development and analysis

After overnight hybridisation, the hybridisation buffer was discarded and the blot membrane was washed 2 x with 50 mL of 2 x SSPE, 0.5% SDS (10 min then 30 min) and 2 x with 50 mL of 0.2 x SSPE, 0.5% SDS (2 x 30 min) at 65°C with rotation. Buffers were discarded after each wash and excess liquid from membrane was blotted off on Whatman paper. The blot was then wrapped in Saran foil and exposed to film inside a Fujifilm BAS cassette 2040 (Fujifilm, Duesseldorf, Germany) for a few weeks. The membrane was scanned with the Storm 840 scanner and Storm Scanner Control programme from Molecular Dynamics (Amersham Pharmacia Biotech, Sunnyvale, California, U.S.A.) and the picture exported as a .tif file. Pictures were enhanced with the Adobe Photoshop v.5.

The 1 kb ladder was used to create a standard log curve by plotting the known molecular marker sizes against their run distance on the gel. Sample band sizes were then mapped to it. The *cbh1* DNA sequence (JGI website) was used to map MNase restriction sites and predict nucleosome positioning.

5.3 Results

5.3.1 Expression of cellulase and hemicellulase-encoding genes in *T. reesei* QM6a in different carbon sources

Expression of 4 cellulase (*cbh1*, *cbh2*, *egl1* and *bgl1*) and 3 hemicellulase (*xyn1*, *xyn2* and *bxl1*)-encoding genes was studied in QM6a mycelia grown in the presence of glucose, cellobiose, xylose, glycerol and sophorose. All gene expression was recorded at time points where fungal biomass reached similar levels.

All cellulase-encoding genes were repressed in glucose, cellobiose, xylose and glycerol; and they were induced in the presence of sophorose and cellulose (Figures 5.3.1 and 5.3.4). Out of all the genes studied, *cbh1* and *egl1* were expressed the highest in sophorose whereas *egl1* and *bxl1* were expressed the highest in cellulose. Sophorose was a better inducer of cellulase-encoding genes in QM6a than cellulose; expression of cellulolytic genes was roughly half of that in cellulose when compared to sophorose (compare figures 5.3.4 and 5.3.1). In repressing conditions, *cbh1* transcript levels were the lowest. There were also differences between the repressive effects of the different carbon sources: glucose was most repressive for *cbh1* and *cbh2* whereas glycerol and xylose were most repressive for *egl1* and *bgl1*.

The expression levels of hemicellulase-encoding genes were different from the cellulolytic genes when grown in the same conditions (compare Figures 5.3.1 and 5.3.2). Sophorose and cellulose did not induce *xyn2* but induced expression of *xyn1* and *bxl1* although not to the same levels as for *cbh1* or *egl1* (Figure 5.3.2). The expression of *xyn2* was diminished in the presence of sophorose and cellulose. Hemicellulase gene expression levels were higher in glucose, cellobiose, xylose and glycerol than were those for the cellulase-encoding genes. Especially, *xyn2* seemed to be constitutively expressed at higher levels in these carbon sources when compared to *xyn1*.The levels of *bxl1* transcripts increased in the presence of xylose and glycerol when compared to glucose but were not as high as when sophorose was added.
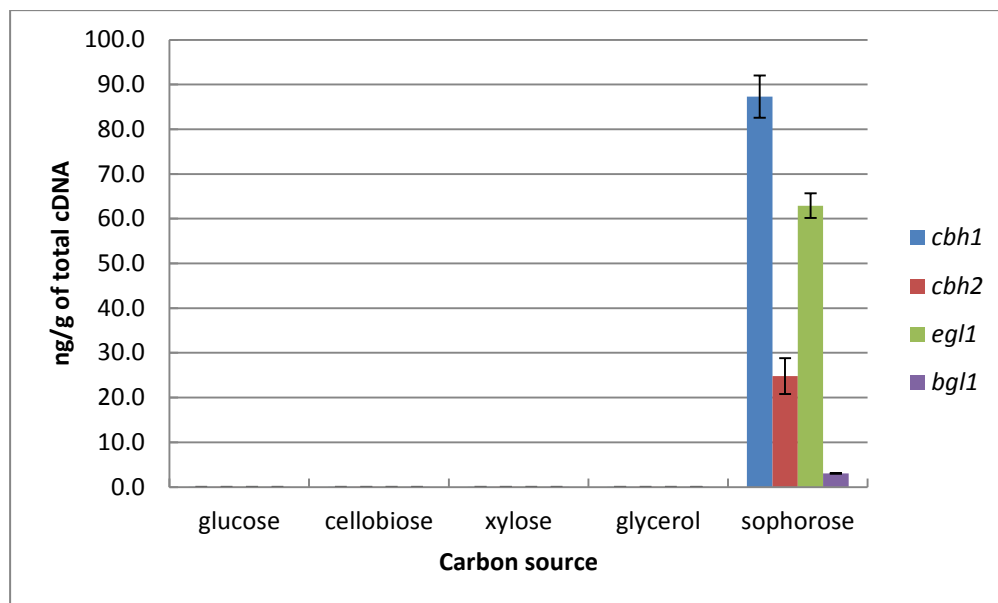
Figure 5.3.1: Transcript levels of cellulase-encoding genes in *T. reesei* QM6a when grown in TMM supplemented with different carbon sources. Standard deviations of three replicates are shown as error bars.
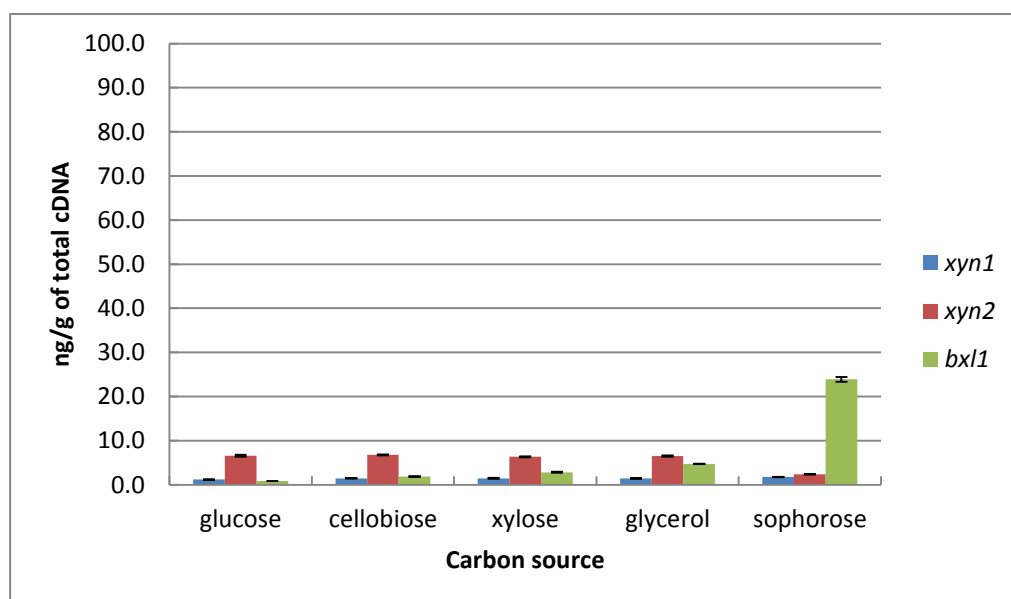
Figure 5.3.2: Transcript levels of hemicellulase-encoding genes in *T. reesei* QM6a when grown in TMM supplemented with different carbon sources. Standard deviations of three replicates are shown as error bars.

5.3.2 Expression of cellulase and hemicellulase-encoding genes in CCR de-repressed strains in the presence of glucose and cellulose

The expression of the same 4 cellulase (*cbh1*, *cbh2*, *egl1* and *bgl1*) and 3 hemicellulase (*xyn1*, *xyn2* and *bxl1*)-encoding genes was also studied in 3 strains which are relieved of carbon catabolite repression (CCR) when cultured in glucose and cellulose rich media. These strains were a *cre1* deletion strain (Δ*cre1*), a *cre1* truncated strain (*cre1-1*) and strain RUTC30. Expression levels were also recorded in the presence of cellulose in *T. reesei* QM6a.

Expression levels of all cellulase and hemicellulase-encoding genes in the CCR-relieved strains were much higher in cellulose than they were in glucose, despite these mutants having a non-functional *cre1* gene (compare Figures 5.3.3 and 5.3.4). Especially the *cbh1* and *cbh2* genes were induced the most in the mutant strains in the presence of cellulose whereas expression levels of the other genes was similar in all strains (Figure 5.3.3).

In QM6a, expression levels of all genes increased in the presence of cellulose when compared to growth in glucose (Figures 5.3.2, 5.3.2 and 5.3.4). Despite the induced state of the cellulase and hemicellulase-encoding genes, transcript levels in QM6a were similar to those in the carbon catabolite de-repressed strains when grown in glucose.

In the presence of glucose, expression of the different genes varied between the carbon catabolite de-repressed strains. Whereas *bxl1* transcript levels were high in all strains, *egl1* was induced the strongest in Δ*cre1* and RutC30; *bgl1* was expressed the highest in *cre1-1* and RutC30. Apart from these two differences, gene expression in the *T. reesei* Δ*cre1* and *cre1-1* strains was similar. In contrast, general transcript levels of *cbh1*, *cbh2* and *egl1* as well as *xyn1* and *xyn2* were higher in RutC30 than in the other two mutant strains (Figure 5.3.4).
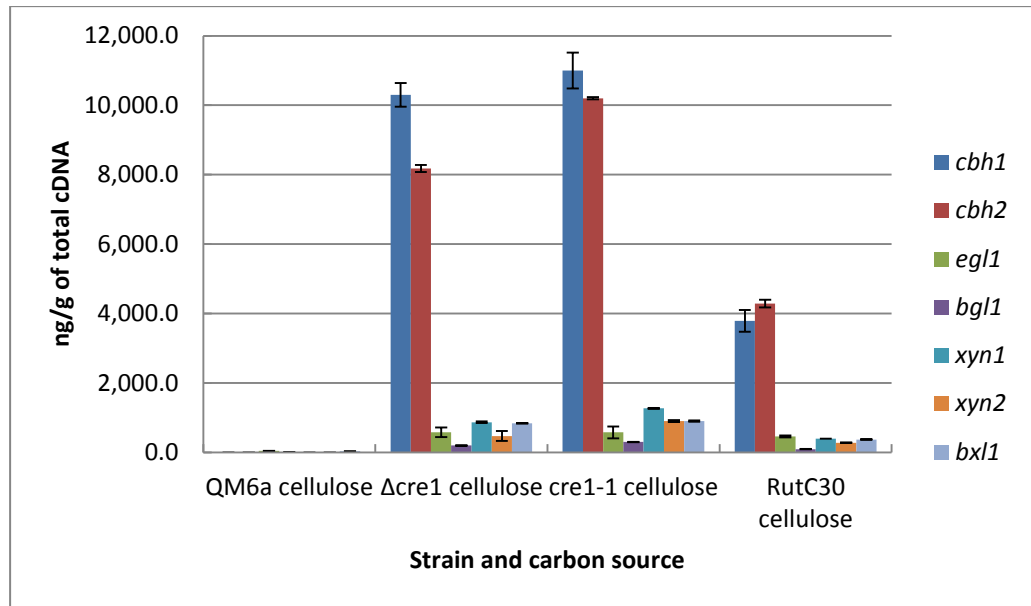
Figure 5.3.3: Transcript levels of cellulase and hemicellulase-encoding genes in *T. reesei* QM6a and CCR de-repressed strains during growth in TCM supplemented with cellulose. Standard deviations of three replicates are shown as error bars.
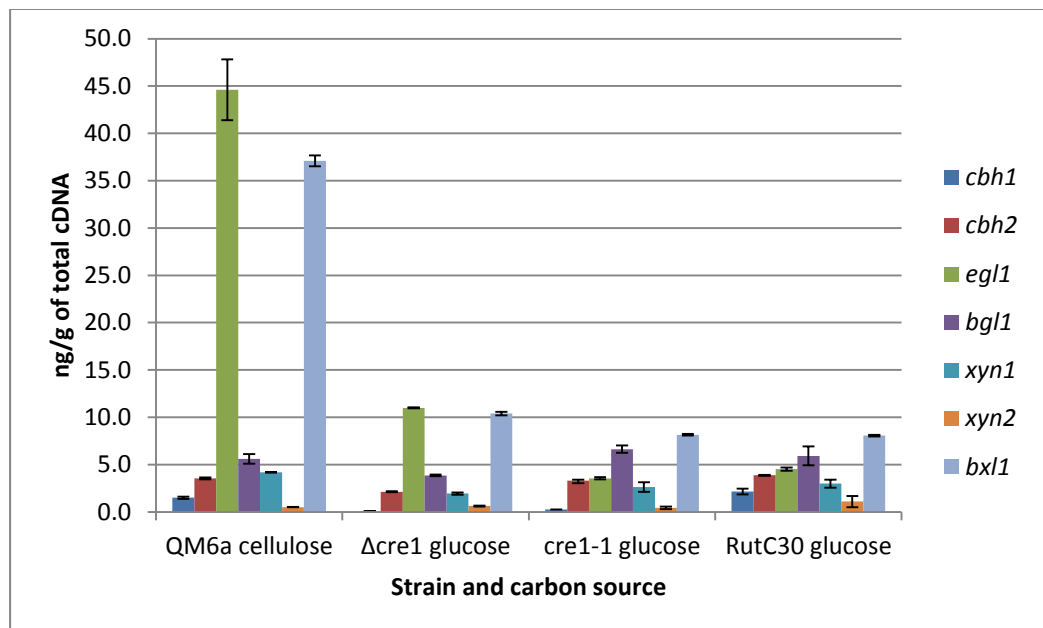
Figure 5.3.4: Transcript levels of cellulase and hemicellulase-encoding genes in *T. reesei* CCR de-repressed strains during growth in TCM supplemented with glucose. Standard deviations of three replicates are shown as error bars. Transcript levels of the same genes in the wild-type QM6a strain when grown in the presence of cellulose are shown for comparison.

5.3.3 Cbh1 secretion

The *cbh1* promoter has been reported to be the strongest in *T. reesei* and CBH1 can make up the majority of the secreted protein mix. Thus CBH1 enzyme secretion was also studied in the presence of glucose, cellobiose, xylose, glycerol and sophorose.

Total protein secretion differed between the different carbon sources but also between earlier (24 h and 36 h) and later time points (72 h) during growth (data not shown). As a general trend, the levels of protein secretion increased the longer mycelia were cultured in TMM. Furthermore, at the earlier time points during growth, total protein concentration was similar at around 2.0 x 10^{-3} g/L except when grown in the presence of sophorose where it was much higher at 5.5 x 10^{-3} g/L. After 72 h growth in TMM, the amount of protein secreted differed widely depending on the carbon source: high amounts of protein (approximately 20.0 x 10^{-3} g/L) were detected in glucose and cellobiose and a very low amount was recorded when grown in xylose and glycerol (3.7 x 10^{-3} g/L and 1.9 x 10^{-3} g/L). Total protein concentration was not measured after 72 h of growth in sophorose media.

CBH1 (4.62 x 10^{-3} g/L) could only be detected in the presence of sophorose and in no other carbon source and comprised 84% of the total protein secreted when *T. reesei* QM6a was grown in sophorose (data not shown).

5.3.4 Protein analysis

To profile secreted proteins from *T. reesei* grown in the different carbon sources, protein gels of the mycelial supernatant were run. The expected protein sizes of cellulases and hemicellulases based on amino acid composition and without glycosylation are as follows: CBH1 65 kDa, CBH2 58 kDa, EGL1 43 kDa, BGL1 75 kDa, XYN1 19 kDa, XYN2 21 kDa and BXL1 80.4 kDa.

As mentioned in section 5.3.3, there was a difference in the variety, number and amount of proteins secreted by QM6a in the presence of different carbon sources. The amount and variety of different proteins was highest in the presence of sophorose (Figure 5.3.5) which was also confirmed by the measurement of total protein concentration (Section 5.3.3). In glucose, the least number of proteins were secreted and the profiles between xylose and glycerol were very similar (Figures 5.3.5).

In order to analyse the nature of some of the secreted protein, several protein bands which were different between carbon sources were analysed by Mass-Spectrometry sequencing (Figure 5.3.5). The results from the sequencing are depicted in red in Figure 5.3.5.

Bands 5 and 6 which were only present when mycelia were grown in sophorose and which turned out to be the same protein due to the large band on gels, reflecting its high secreted levels, corresponded to CBH1 (Figure 5.3.5). This is also in agreement with transcript analysis and the expected band size.

Bands 1 and 2 corresponded to two GHs of families 3 and 55 and these enzymes were present in all carbon sources except for glucose. The GH family 3 comprises mainly β-glucosidases and in this case it was secreted in high amounts in the presence of cellobiose and sophorose. GH family 55 includes enzymes such as glucanases which seemed to be required for growth on cellobiose, xylose and glycerol.

Band 3 corresponded to a protein belonging to the GH 5 family and BLAST analysis confirmed it to be XYN4 (Xylanase 4). XYN4 was secreted in the presence of xylose, glycerol and sophorose but not in glucose and cellobiose.

Sequencing analysis of bands 4 and 8 confirmed these proteins to be esterases. Band 4 corresponded to an α/β hydrolase and band 8 to an acid phosphatase which was only seen in the presence of sophorose. Esterases are involved in the cleavage of ester links and can contribute to the cleavage of various substrates within biopolymers thus helping carbohydrate deconstruction. They use a different mechanism to the GHs (see Chapter 1).

No clear identification was obtained for band 7. A summary of all sequenced bands is given in Table 5.2.
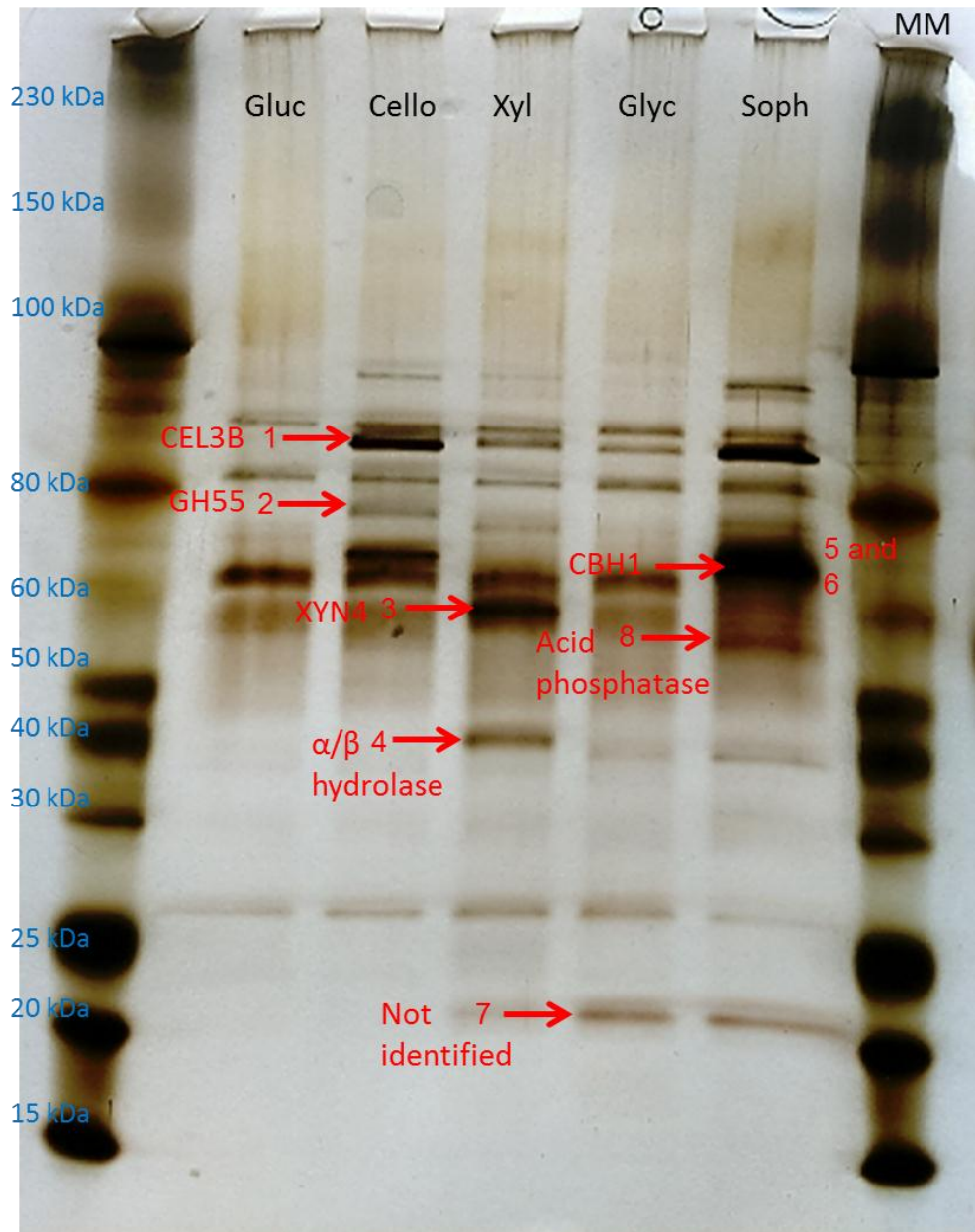
Figure 5.3.5: Silver stained protein gel of TMM supplemented with different carbon sources after 24 h or 36 h growth of *T. reesei* QM6a. A total of 1.5 g/L protein (determined by Nanodrop spectrometry) was loaded into each lane. Red arrows and numbers (CBH1 = cellobiohydrolase, XYN4 = xylanase) within the gel indicate protein bands which have been cut out and send off for Mass-Spectrometry sequencing (MM = molecular marker in kDa = kilodaltons).

Table 5.2: Sequencing details of the 8 bands analysed by Mass-Spectrometry. The type of search and score (Ions score is -10*Log(P), where P is the probability that the observed match is a random event. Individual ions scores > 61 indicate identity or extensive homology (p<0.05). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.), number and sequence coverage of matching peptides to the nearest protein are indicated. Protein masses are also given and their corresponding gene transcript I.Ds.

Band Number	Number of matched peptides	Mass (Da)	Score	Sequence coverage	Matched to nearest *T. reesei* QM6a protein	Transcript I.D. of matched protein	Search type
1	15	93890	493	18.0%	GH3, Cel3b, beta-glucosidase	121735	MASCOT
2	3	83034	77	4.0%	GH55, glucan 1,3-beta-glucosidase	121746	MASCOT
3	4	52812	170	10.0%	GH5, xylanase 4	111849	MASCOT
4	3	unknown	83	10.7%	Alpha/beta hydrolase	104461	manual
5	7	54077	284	15.0%	CBH1, GH7	123989	MASCOT
6	Same as band 5 = CBH1						MASCOT
7	No match due to absence of found peptides as a result of very low signal levels						MASCOT and manual
8	6	47758	221	19.0%	Phosphoesterase	76155	MASCOT

5.3.5 MNase digestion of purified gDNA – control experiment

In order to check whether MNase cuts DNA in the *cbh1* promoter independent of being occupied by nucleosomes or not, extracted gDNA was digested with MNase (Figure 5.3.6), purified and subjected to the same restriction digest as the chromatin samples. Genomic DNA which was only treated with restriction enzymes showed the same banding pattern (Figure 5.3.7) as the chromatin samples which were only subjected to the restriction digest and not to MNase (see Southern blots glucose and sophorose). MNase-treatment of gDNA did not result in any obvious bands (Figure 5.3.7).
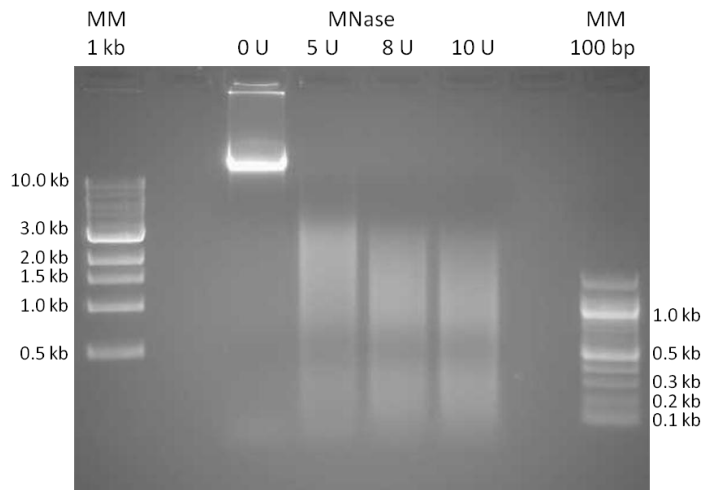
Figure 5.3.6: Agarose gel showing MNase digestion of gDNA. Concentrations of MNase and molecular markers (MM) are indicated.
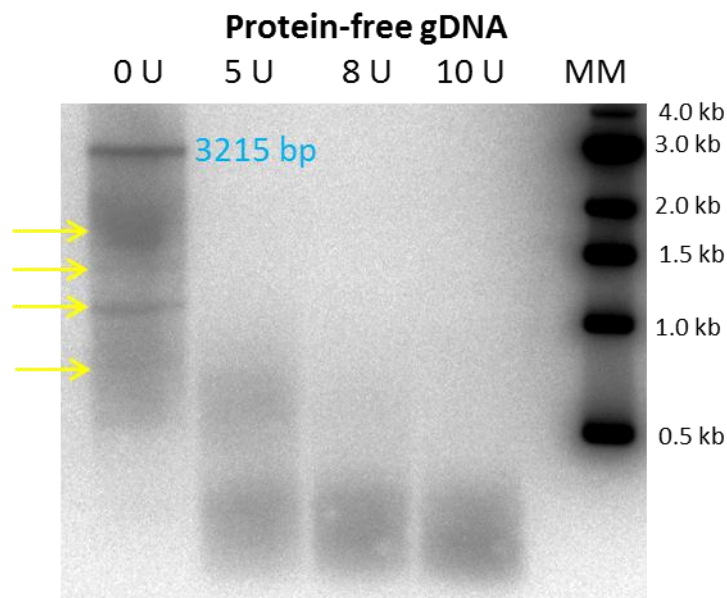
Figure 5.3.7: Southern blot of *T. reesei* gDNA digested with MNase and with restriction enzymes. MNase concentrations are indicated in U (MM = molecular marker). Yellow arrows indicate bands which are also present in the absence of MNase. These bands have the same size as those indicated in the other Southern blot pictures (see below).

5.3.6 Nucleosome positioning in the *cbh1* promoter and coding regions of QM6a and the *cre1* mutant strains Δ*cre1*, *cre1-1* and RUTC30 under repressing conditions

In glucose-rich conditions, digestion of chromatin with MNase resulted in 14 distinct bands for the wild-type strain and 8 for the mutant strains (Figure 5.3.8, differences between the wild-type and mutant strains are indicated by blue arrows). 5 bands were also observed in the absence of MNase for all strains meaning that only 9 of the 14 wild-type and 3 of the 8 mutant bands represented actual MNase cutting sites (yellow arrows in Figure 5.3.8). Thus, there were 9 MNase hypersensitive (MH) sites in the wild-type ORF and 3 MH sites in the ORF for the mutant strains indicating a loss of MH sites in strains carrying a non-functional *cre1* gene. Two nucleosomes could be positioned in the promoter region (between MH sites 1, 2 and 3) of *cbh1* in all strains whereas the coding region of the wild-type contained 4 positioned nucleosomes and the coding region of the mutant strains appeared to be nucleosome-free.

A similar pattern of MNase digested bands was also obtained for chromatin of wild-type mycelia grown in cellobiose and xylose thus nucleosome positioning was the same in all three carbon sources (Figure 5.3.9). The resolution of the cellobiose and xylose blots was less than the glucose one, probably due to low DNA loading.

A duplicate of the MNase digestion of glucose-grown mycelia is in Appendix A9 Figure A.9.F.1.
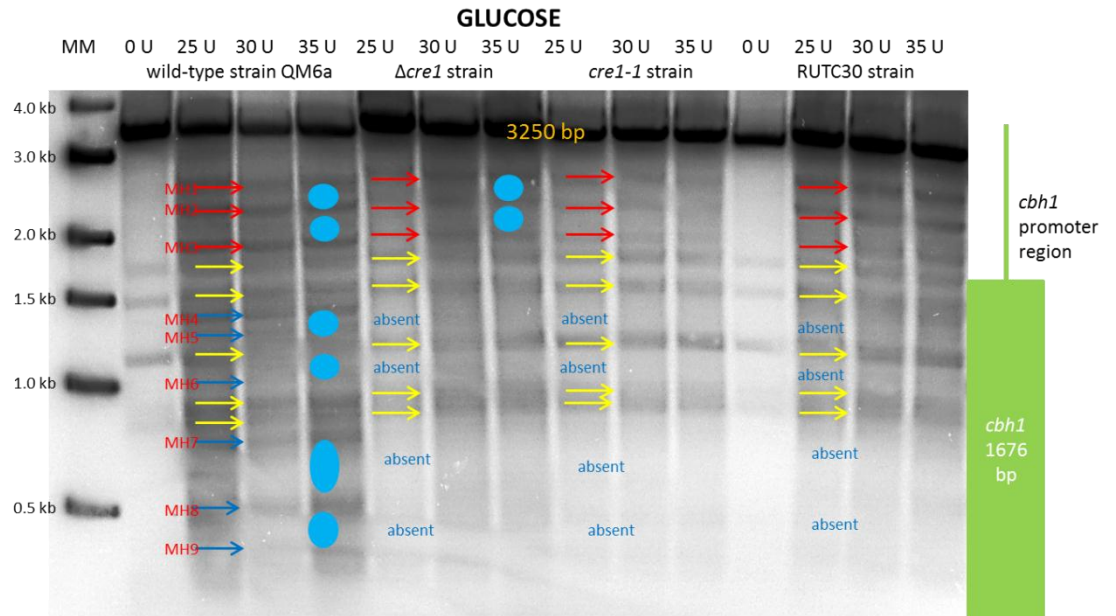
Figure 5.3.8: Southern blot of MNase digested DNA extracted from glucose grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marker; 3250 bp = cut with restriction enzymes but not MNase). Red arrows indicate same MNase hypersensitive (MH) cutting sites for all strains, blue arrows show differences in MH sites between strains and yellow arrows indicate bands which are also present in the absence of MNase. The likely positions of nucleosomes (blue spheres) with reference to the location of the *cbh1* ORF are also indicated.
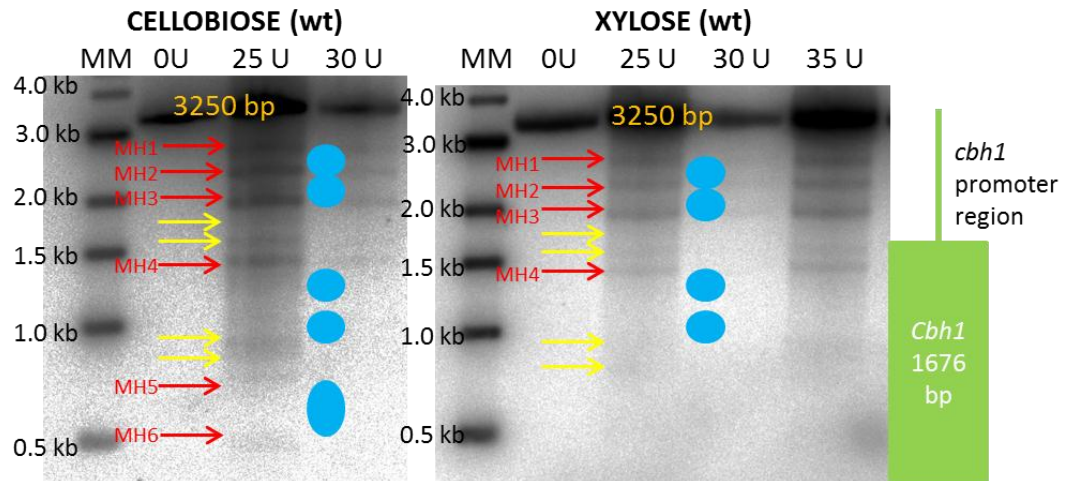
Figure 5.3.9: Southern blot of MNase digested DNA extracted from cellobiose and xylose grown mycelia. MNase concentrations (in U), carbon source and strain are indicated above the blot picture (MM = molecular marker; 3250 bp = cut with restriction enzymes but not MNase). Red arrows indicate MNase hypersensitive (MH) cutting sites and yellow arrows indicate bands which are also seen in the absence of MNase. The likely positions of nucleosomes (blue spheres) with reference to the location of the *cbh1* ORF are also indicated.

5.3.7 Nucleosome positioning in the *cbh1* promoter and coding regions of QM6a and the *cre1* mutant strains Δ*cre1*, *cre1-1* and RUTC30 under inducing conditions

In the presence of sophorose, 12 MNase-generated bands were seen for the wild-type strain and 13 for the *cre1* mutant strains (Figure 5.3.10). In the absence of MNase the same five bands were observed in sophorose-rich conditions as in the presence of glucose. Thus, only 7 MH sites were found within the *cbh1* ORF in the wild-type strain and 8 MH sites for the mutant strains in the presence of sophorose. MH sites 6 to 7/8 were very weak in the wild-type and mutant strains and are likely to not correspond to a positioned nucleosome. In contrast to the situation in glucose, the promoter region of *cbh1* contained 3 positioned nucleosomes (of which two appeared to be "fuzzy" nucleosomes due to weak MH sites 3 and 5) whereas the coding region was nucleosome-free. The MH6 site in the *cre1* deficient strains was also relatively weak and may correspond to a partially positioned nucleosome.

A second duplicate sophorose blot showing clearer bands obtained in the absence of MNase is in Appendix A9 Figure A.9.F.2.

Preliminary analysis of nucleosome positioning in wild-type mycelia grown in cellulose medium revealed a completely nucleosome-free *cbh1* promoter and ORF. Nine bands were observed in each condition but these bands were also present in the absence of MNase (yellow arrows, Figure 5.3.11). No other MNase cutting sites were observed in the *cbh1* promoter or gene region in wild-type mycelia. In the mutant strains, on the other hand, 4 additional weak MNase cutting sites were present which may correspond to two partially or weakly positioned nucleosomes. A duplicate cellulose blot is shown in Appendix A9 Figure A.9.F.3.
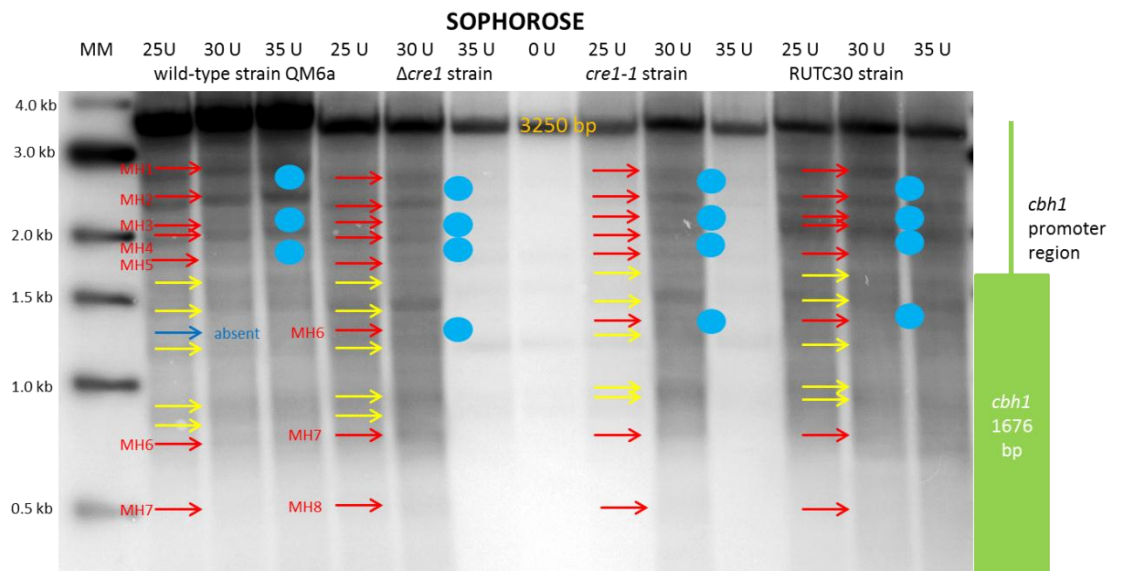
Figure 5.3.10: Southern blot of MNase digested DNA extracted from sophorose-grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marker; 3250 bp = cut with restriction enzymes but not MNase). Red arrows indicate same MNase hypersensitive (MH) cutting sites for all strains, blue arrows show differences in MH sites between strains and yellow arrows indicate bands which are also seen in the absence of MNase. The likely positions of nucleosomes (blue spheres) with reference to the location of the *cbh1* ORF are also indicated.
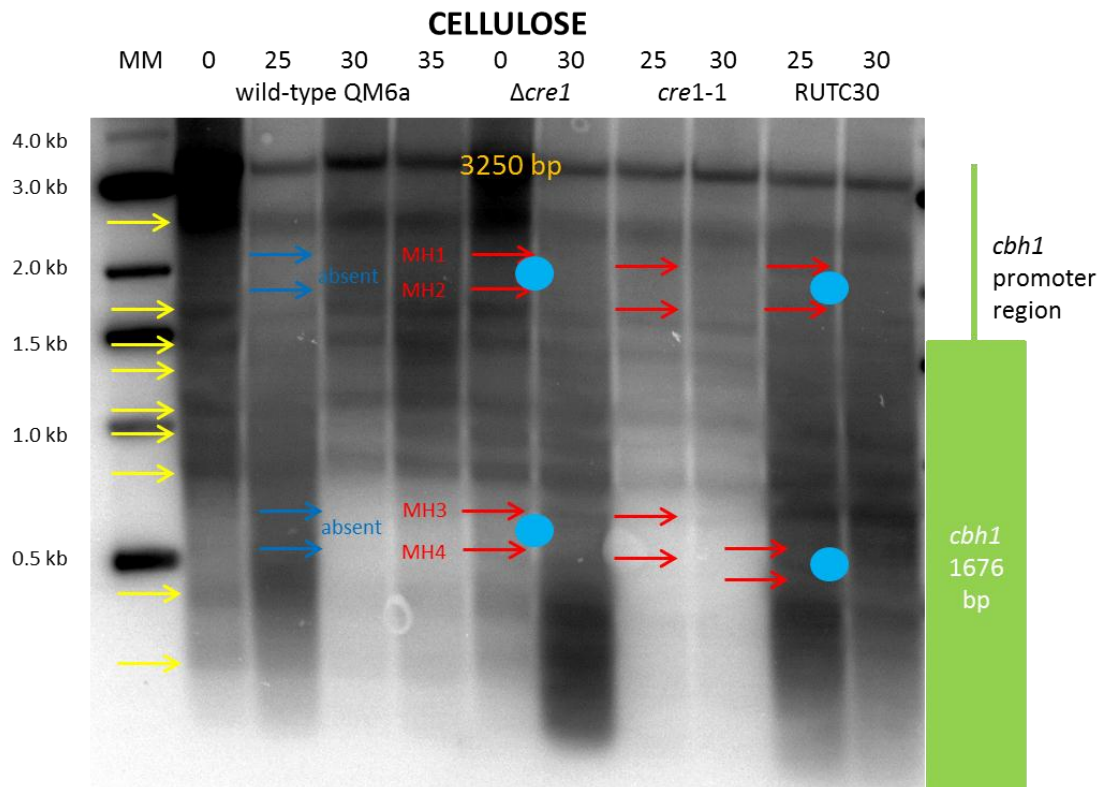
Figure 5.3.11: Southern blot of MNase digested DNA extracted from cellulose-grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marker; 3250 bp = cut with restriction enzymes but not MNase). Red arrows indicate same MNase hypersensitive (MH) cutting sites for all strains, blue arrows show differences in MH sites between strains and yellow arrows indicate bands which are also seen in the absence of MNase. The likely positions of nucleosomes (blue spheres) with reference to the location of the *cbh1* ORF are also indicated.

5.3.8 Summary

Mapping of MH sites revealed the presence of 2 nucleosomes in the *cbh1* promoter region and the presence of 4 nucleosomes in the coding region in glucose-rich conditions (Figure 5.3.12). The first nucleosome (-1) covered the region between approximately -340 bp to -500 bp and spans two ACE1 binding sites. It also sits next to a XYR1 binding site (-320 bp) and an ACE2 binding site (-513 bp). The second nucleosome (-2) spanned the region between -840 bp and -1140 bp and also contains two ACE1 binding sites as

well as being in close proximity to an ACE2 binding site. The four CRE1 binding sites were nucleosome-free in the *cbh1* promoter under glucose-rich conditions. Furthermore, 3 ACE1 binding sites at around -200 bp were also protein-free. The same pattern was observed in the promoter region of the mutant strains. There was a difference in the nucleosome positioning though within the coding region of the gene between the wild-type and the mutant strains. Glucose-grown wild-type mycelia contained 4 positioned nucleosomes which span the most of the coding region whereas the mutant strains lacked these nucleosomes. Thus a non-functional CRE1 protein seems to influence nucleosome positioning within the coding region and two putative binding sequences are found at +797 bp and +1609 bp within the *cbh1* gene.

In the presence of sophorose, the coding region of *cbh1* became nucleosome-free whereas the promoter region saw an increase in partial nucleosome occupancy for all strains (Figure 5.3.13). The -2 nucleosome observed under glucose-rich conditions remained in the same position in the presence of sophorose (-3 nucleosome). The -2 nucleosome in sophorose was roughly in the same position (-300 bp to -450 bp) as nucleosome -1 under *cbh1* repressing conditions. The CRE1 binding sites as well as the second XYR1 site (-733 bp) and the ACE2 binding sites at around -800 bp were protein-free in sophorose-rich conditions. A fuzzy nucleosome (-1 nucleosome) also appeared to cover the region between +4 bp and -150 bp which contains the TATA box, the TSS and one ACE1 binding site. The same MH site was more pronounced in the mutant strains. The *cre1* mutant strains also contained another weak MH site at +550 bp potentially indicating another fuzzy nucleosome. A non-functional CRE1 protein in *cbh1* inducing conditions had a less dramatic effect on nucleosome positioning than under repressing conditions. Thus, CRE1 does not seem to play an equally important role in organising the chromatin structure under sophorose rich conditions than in the presence of glucose.

In the presence of cellulose, the *cbh1* promoter and ORF were completely nucleosome-free. Thus the chromatin architecture in the *cbh1* promoter and gene regions is different from the one observed in sophorose. In the *cre1* mutant strains two "fuzzy" nucleosomes were observed: the -1 nucleosome covers roughly the same region (between -46 bp to -246 bp) containing the TSS, TATA box, one ACE1 and one ACE2 binding site, as the fuzzy -1 nucleosome in sophorose; and the second nucleosome corresponds to the 3' ORF nucleosome (+1154 bp to +1334 bp) which is also observed under glucose and sophorose conditions.
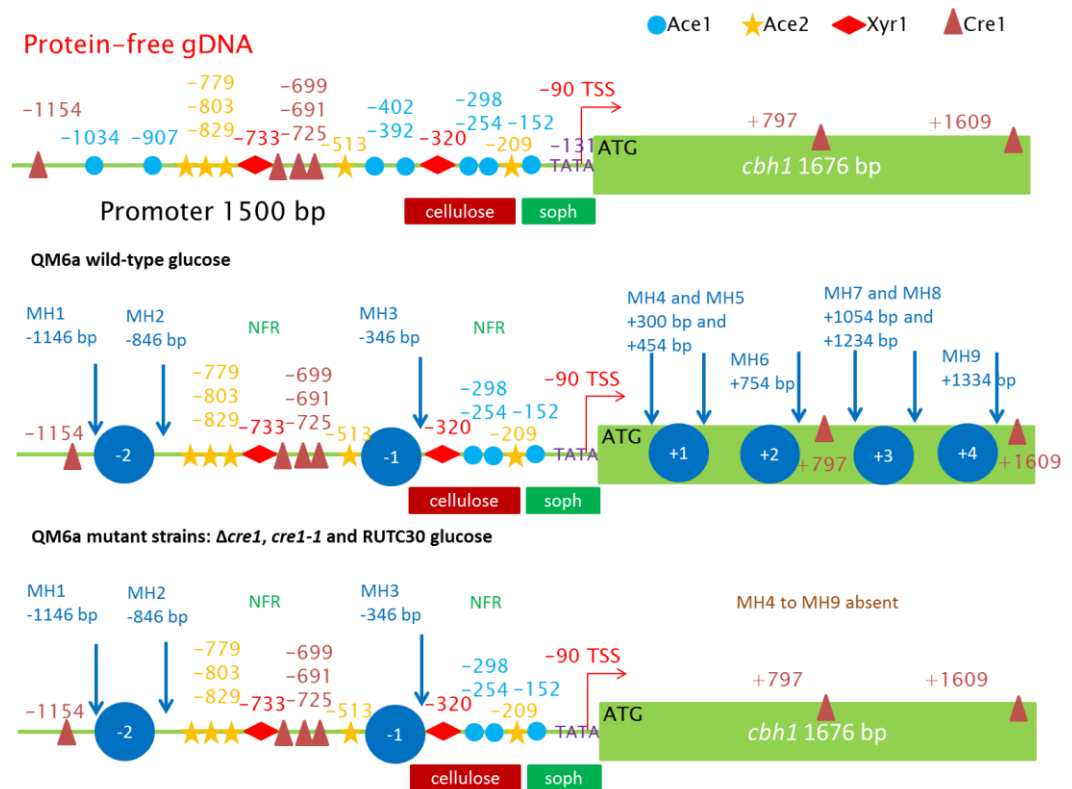
Figure 5.3.12: Diagram of nucleosome positioning (blue spheres) and MNase cutting sites (blue arrow) in the *cbh1* promoter in the presence of glucose in the wild-type strain and the *cre1* mutant strains. As a reference, the promoter region of *cbh1* is also presented when protein free and with all the putative transcription factor binding sites. The locations of DNA sequences thought to mediate induction by sophorose and cellulose are indicated with a green and red box respectively; the TSS (transcription start site), TATA box and nucleosome free regions (NFR) are also shown.
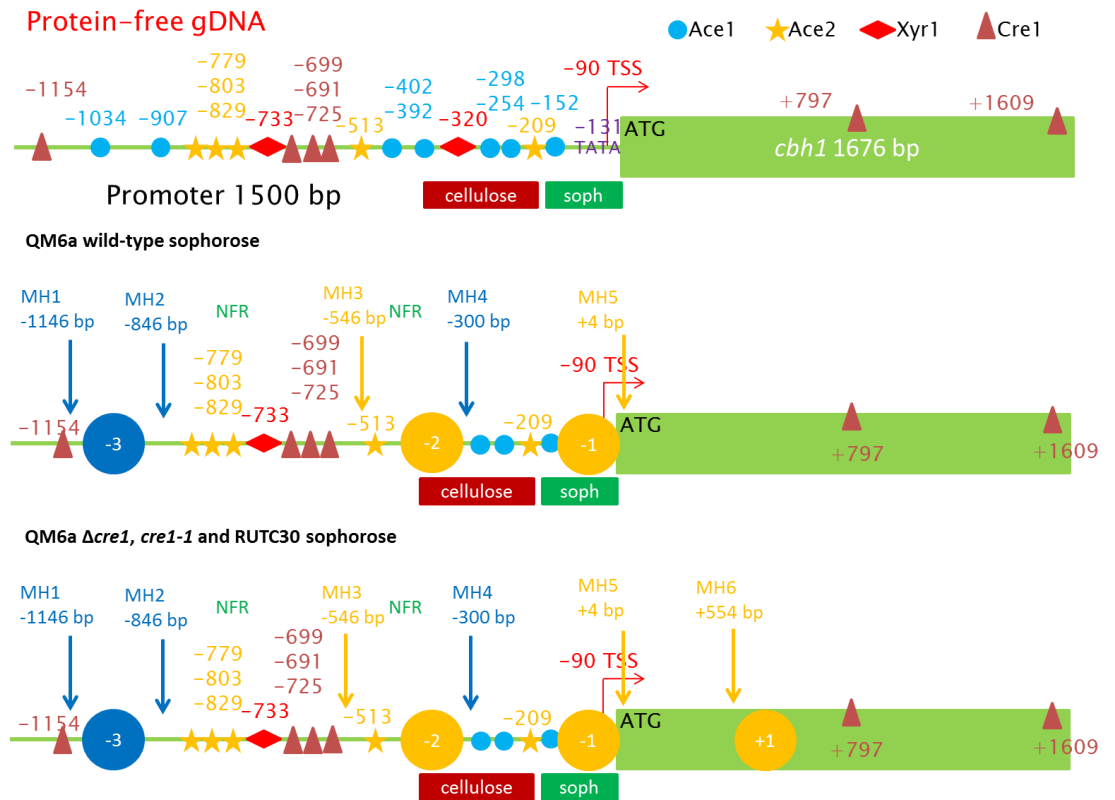
Figure 5.3.13: Diagram of nucleosome positioning (blue and orange spheres) and MNase cutting sites (blue arrows = clear MNase cutting sites and orange arrows = weak MNase cutting sites) in the *cbh1* promoter in the presence of sophorose in the wild-type strain and the *cre1* mutant strains. Blue spheres indicate strongly positioned nucleosomes whereas orange spheres indicate partially positioned nucleosomes. As a reference, the promoter region of *cbh1* is also presented when protein free and with all the putative transcription factor binding sites. The locations of DNA sequences thought to mediate induction by sophorose and cellulose are indicated with a green and red box respectively; the TSS (transcription start site), TATA box and nucleosome free regions (NFR) are also shown.
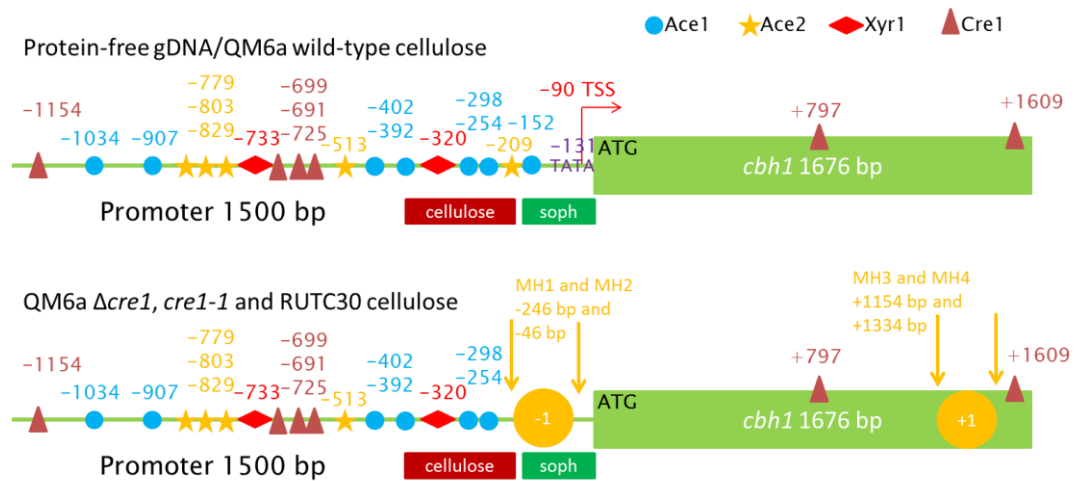
Figure 5.3.14: Diagram of nucleosome positioning (orange spheres) and MNase cutting sites (orange arrows = weak MNase cutting sites) in the *cbh1* promoter in the presence of cellulose in the wild-type strain and the *cre1* mutant strains. Orange spheres indicate partially positioned nucleosomes. As a reference, the promoter region of *cbh1* is also presented when protein free (which also corresponds to the *cbh1* chromatin structure in the wild-type) and with all the putative transcription factor binding sites. The locations of DNA sequences thought to mediate induction by sophorose and cellulose are indicated with a green and red box respectively; the TSS (transcription start site), TATA box and nucleosome free regions (NFR) are also shown.

5.4 Discussion

The expression of different cellulase and hemicellulase-encoding genes by *T. reesei* was studied in the presence of various carbon sources. As discussed in Chapter 3, different carbon sources influence the growth rate of *T. reesei* QM6a and therefore time points were chosen where fungal biomass was similar during growth on the different substrates. QM6a was cultured for 24 h in glucose and cellobiose, 36 h in xylose, glycerol and sophorose; and for 72 h in cellulose. Cellulose was the most complex carbon source used here and more time is needed to deconstruct it. 72 h was chosen as a time point because at this point enough mycelia were present to use in nucleosomal assays.

In the presence of glucose, cellobiose, xylose and glycerol, cellulase gene expression was repressed (Figure 5.3.1). As previously reported, glucose repression of *cbh1*, *cbh2*, *egl1* and *bgl*1 is mediated via CRE1. Xylose also did not induce cellulase gene expression. Xylose was shown to induce cellulolytic and xylanolytic genes at very low concentrations of 0.5mM to 1mM (Mach-Aigner *et al.*, 2010), whereas the xylose concentration used in this project was 132mM (2% w/v - same concentration used as for the other simple carbon sources). In the presence of high concentrations of xylose, ACE1 represses *xyr1* transcription - the main cellulase inducer. Interestingly, cellobiose was found not to promote cellulase gene transcription in this project. Cellobiose was reported to induce cellulases through the conversion of it into sophorose by the transglycosylation activity of BGL1 (Mach *et al.*, 2003). It is possible that the breakdown of cellobiose, as the sole carbon source present, into glucose monomers (cellulase repression) may be energetically more favourable than converting it into sophorose in the conditions studied here. Ilmén *et al.* (1997) reported that cellobiose did not induce *cbh1* in *T. reesei* QM9414 until 62 h after initial inoculation and they speculated that cellulase induction by cellobiose depended on culture conditions, nitrogen source, consumption rate and assimilation of glucose. So it is possible that the concentrations of cellobiose used in this study were too high and incubation time too low to convert this substrate into sophorose. Glycerol was reported to be a neutral carbon source although in this case it seems to be repressive, as no elevated levels of transcription were detected when compared to glucose. Furthermore, no CBH1 activity were detected in the supernatants of mycelia cultured in the different carbon sources confirming that *cbh1* was only expressed in the presence of sophorose. The transcription of hemicellulase-encoding genes in the repressing carbon sources was higher than the expression levels recorded for the cellulase-encoding genes (Figure 5.3.2). In the presence of glucose *xyn1* and *bxl1* transcription was abolished

through CRE1-mediated carbon catabolite repression (Mach *et al.*, 2003). CRE1 was shown to bind to the *xyn1* and *cbh1* promoters although carbon catabolite repression of *cbh1* seems to be more efficient as lower transcript levels of this gene in glucose were detected; *cbh1* is also under the control of another negative regulator, ACE1, which may contribute to these lower levels of transcripts in glucose. In the presence of cellobiose, xylose and glycerol, *xyn1* and *bxl1* transcript levels were slightly higher than when compared to levels in glucose; this indicates alleviation from CRE1-mediated repression rather than induction. Furthermore some alcohols such as glycerol have been shown to mediate concentration-dependent *bxl1* induction. The expression of *xyn2* was highest in glucose, cellobiose, xylose and glycerol and transcript levels did not change much between carbon sources indicating that none of the other carbon sources induced the transcription of this gene when compared to glucose. The expression of *xyn2* is not under the influence of CRE1 which explains the higher transcript levels of this gene in glucose when compared to *xyn1* (Stricker *et al.*, 2007). Xylose was reported to induce *xyn1* and *xyn2* (Stricker *et al.*, 2008) but this was not observed here. Possible reasons for this discrepancy could be the direct transport of xylose into the cell without needing any xylanases to cleave off this substrate from xylan; cultivation time (gene expression may happen much earlier) and the high xylose concentrations used here.

In the presence of sophorose all genes except for *xyn2* were up-regulated (Figures 5.3.1 and 5.3.2). Transcript levels of cellulase-encoding genes, and especially *cbh1* and *egl1* expression levels, were the highest when *T. reesei* was grown in sophorose. The concentration of sophorose used in this project was 0.73 mM (or 0.025% w/v) and that concentration has been reported to induce cellulase-encoding genes efficiently (Suto and Tomita, 2001). CBH1 and EGL1 are the main extracellular cellulases which attack and degrade cellulose fibers. CBH2 is associated with the cell wall and is the main cellulase in conidia (Suto *et al.*, 2001). The transcription of *bgl1* was lowest when compared to the other cellulase-encoding genes. This has been reported before as *bgl1* is subject to CCR in the presence of glucose, a sugar that is released by the enzyme itself. Sophorose also induced *xyn1* and *bxl1* transcript levels but reduced *xyn2* expression (Figure 5.3.2). Expression levels of *xyn1* and *bxl1* were not as high as those for *cbh1*, *cbh2* and *egl1* in the presence of sophorose. This could be because hemicellulases such as the main xylanase *xyn1* are not required for cellulose breakdown. Cellulose is found in lignocellulose in conjunction with hemicellulose and it is possible that this substrate and its resulting inducers (e.g. sophorose) may trigger some hemicellulase gene induction. Furthermore, cellulases were reported to make up the majority of the secretome in *T. reesei* in the presence of sophorose so

it is possible that hemicellulases are generally not induced to the same high levels as cellulases in the wild-type strain. Transcript levels of *bxl1* were the highest of all hemicellulases studied in sophorose and the *T. reesei* BXL1 protein was reported to have some β-glucosidase activity which could be useful for cellulose degradation. It has been reported that sophorose also induces *xyn2* which was not observed here. In contrast, *xyn2* transcript levels decreased in the presence of sophorose in QM6a. A possible explanation for this is that reports were based on the analysis of strain QM9414 which was obtained through gamma-radiation of QM6a (QM6a – QM 9123 mutant 207 – QM 9414) and secretes 5 times more cellulases than QM6a (VTT online culture collection). Druzhinina *et al.* (2006) reported though that carbon utilisation profiles between QM6a and QM9414 were essentially unchanged. Analyses of the nature of the mutations generated in QM9414 may contribute to explaining the results obtained here. Alternatively *xyn2* may be induced at much earlier time points and possibly, later during growth, sophorose actually represses this gene in the wild-type strain. The above discussed results highlight the fundamental differences between cellulase and hemicellulase gene regulation in the wild-type *T. reesei* strain.

Cellulase and hemicellulase gene regulation at the transcriptional level has been studied in great detail and one of the key regulators is CRE1. Hence, expression of the same genes (described above) was monitored in strains expressing a non-functional CRE1 protein and compared to the wild-type gene expression profiles. The strains tested here were QM6a with a deletion of *cre1*, a truncated *cre1* gene (named *cre1-1*) and the cellulase hyper-producing mutant RutC30. Mycelia were cultured for 24 h or 48 h (Δ*cre1*, *cre1-1* and RutC30) in glucose and for 72 h in cellulose. It was reported (and checked in this project) that the three mutant strains grew slower in glucose than the wild-type thus they were incubated for longer (Nakari-Setälä *et al.*, 2007; also see Table 3.3 in Chapter 3 for RutC30 and QM6a growth in glucose). This is in accordance with the findings that CRE1 influences the growth rate (Portnoy *et al.*, 2011).

In glucose, the biggest difference between gene expression levels was observed for the cellulase-encoding genes whereas *bxl1*, *xyn1* and *xyn2* expression levels were similar in all three strains with *bxl1* expressed at the highest levels out of all genes studied (Figure 5.3.4). The transcript levels of *egl1* and *bxl1* were highest in the Δ*cre1* strain; *cbh1* and *cbh2* expression was very low in Δ*cre1* and *cre1-1* but increased in RutC30. RutC30 is a strain which carries many point mutations, large and small gene deletions (Le Crom *et al.*, 2009) and, together, these modifications seem to increase *cbh1* and *cbh2* gene expression when compared to Δ*cre1* or *cre1-1* in the presence of

glucose. The *cbh1* and *xyn1* promoters were shown to bind CRE1 directly *in vivo* (Mach-Aigner *et al.*, 2008) and carbon catabolite de-repression seemed to be enough to cause an increase in *xyn1* and *cbh1* expression (Figure 5.3.2). Furthermore, gene expression in carbon catabolite de-repressed strains was still lower in glucose than when compared to the cellulose-grown wild-type strain. It has been reported that in *cre1* knock-out strains a further inducer is required as this mutation is not enough to increase cellulolytic gene expression (Nakari-Setälä *et al.*, 2007). This was also seen in this project: in the presence of cellulose, cellulase and hemicellulase gene expression in the mutants increased markedly when compared to QM6a. Transcript levels of *cbh1* and *cbh2* were very high in all three strains supporting the necessity for an inducer to achieve full cellobiohydrolase expression. Portnoy *et al.* (2011) also showed higher cellulase secretion by *T. reesei cre1* mutant strains in the presence of another inducer, lactose. In contrast to the situation in glucose, expression of all genes studied here was lower in RutC30 when grown in the presence of cellulose when compared to the other two mutant strains. It is possible that the many mutations in RutC30 may actually be slightly inhibitory for cellulolytic and hemicellulolytic gene expression in the presence of cellulose. The expression of *egl1*, *bgl1* and *bxl1* was very similar in Δ*cre1* and *cre1-1* in cellulose whereas *xyn1* and *xyn2* gene expression was higher in *cre1-1*. In contrast to the wild-type strain, the main xylanolytic genes, e.g. *xyn1* and *xyn2*, were also induced to high levels in the presence of cellulose in the mutant strains, confirming that they are also under CCR and that an inducer is required for their gene expression. Xylanolytic gene expression levels were much lower than those found for the cellulolytic genes. This is probably because cellulose mainly induces cellulase-encoding genes as these enzymes are required for its breakdown. It would be interesting to see if the gene expression profile would be reversed in the presence of xylan, a known hemicellulase inducer.

A general comparison of the wild-type strain with the mutant strains, when grown in sophorose and cellulose, established a slightly different expression profile (Figures 5.3.1 and 5.3.4). Minimal medium was used for growth in sophorose and complete medium for growth in cellulose. Generally speaking, cellulase gene expression levels were lower in cellulose than in sophorose; this is probably due to the fact that cellulose consists of insoluble polymers which require the synergistic action of cellulolytic enzymes for degradation into glucose units. Cellulose degradation is much slower than sophorose uptake and full gene expression is probably achieved at later time points. Cellulose mainly induced *egl1* and *bxl1* transcript levels whereas in the presence of sophorose the expression of *cbh1* and *egl1* were highest. The expression levels of *xyn1* and *xyn2* were similar in sophorose and cellulose

214

whereas *bxl1* transcript levels were higher in the presence of cellulose when compared to sophorose. The BXL1 enzyme possesses both arabinofuranosidase and β-glucosidase activity and seems to be important for cellulose degradation and expression is enhanced by inducers (such as sophorose or various cello-oligosaccharides) derived from cellulose breakdown. The expression of *xyn2* decreased in cellulose as was also observed in the presence of sophorose and *xyn2* transcript levels were lower than those for *xyn1* as has been reported before (Margolles-Clark *et al.*, 1997). Thus the main differences between gene induction in the presence of cellulose and sophorose were seen amongst the cellulolytic genes supporting previous reports that cellulase gene induction and secretion are different and not entirely comparable between cellulose and sophorose (Schuster *et al.*, 2011). The transcriptional profile of *xyn2* found here is somewhat contradictory: in *T. reesei cre1* mutant strains this gene was induced in cellulose and sophorose rich conditions, but not in the wild-type strain where transcript levels were reduced in the presence of these substrates. It is possible that the regulation of *xyn2* is quite different from *xyn1* and is not required at all if cellulose inducers are detected by the cell. Alternatively, expression of this gene may happen at a much earlier time point and is then shut down.

Transcriptional studies confirmed that *cbh1* is expressed in the presence of sophorose and cellulose and repressed in the presence of glucose. As eukaryotic DNA is found in the form of chromatin, nucleosome positioning was studied in the *cbh1* promoter and gene region in the presence of repressing and inducing carbon sources. During growth in glucose-rich media, two positioned nucleosomes were observed in the promoter region, covering the DNA sequences between -350 bp and -510 bp and between -850 bp and -1150 bp, defining two nucleosome-free regions (Figure 5.3.12, all positions are given relative to the translational start codon). Previous studies have found that three binding sites at -700 bp are bound by CRE1 and pivotal to maintaining repression of *cbh1* in the presence of glucose (Mach-Aigner *et al*., 2008). This site was confirmed to be nucleosome-free in this study. ACE1 has also been shown to bind at positions -152 bp, -254 bp and -298 bp in the presence of glucose (Ling *et al*., 2009), i.e. all outside of nucleosomal regions. The human and yeast genome both show a prominent NDR (nucleosome depleted region) at the TSS (transcription start site) (Lee *et al*., 2007) and this also seems to be the case in *T. reesei*. Together, the bound CRE1 and ACE1 transcription factors (TFs) and the two nucleosomes would prevent *cbh1* expression by inhibiting the activator XYR1 from binding to its recognition sequences. It has been reported before that transcription factors co-operate to dictate nucleosome positioning and establish nucleosome-free regions in

highly regulated promoters (Cockerill, 2011). The first XYR1 binding site at -320 bp is flanked by the -1 nucleosome at -350 bp and by the adjacent bound ACE1 at -298 bp. The second XYR1 binding motif at -733 bp is located right next to the 3 CRE1 binding sites (-725 bp, -699 bp and -691 bp). Bound ACE1 TFs and nucleosome -1 would also inhibit binding of the other activator ACE2 at -209 bp and -513 bp. The coding region of *cbh1* contained 4 positioned nucleosomes. TATA-containing and nucleosome-occupied promoters tend to have an increase in nucleosome occupancy downstream of the TSS and active promoters are much less likely to be occupied by nucleosomes than coding regions (Sekinger *et al.*, 2005; Arya *et al.*, 2010; Clark, 2010). Furthermore, Lee *et al.* (2007) found, in a genome-wide yeast study, that highly expressed genes are more occupied by nucleosomes than genes expressed in small amounts. They speculated that the act of transcription requires the formation of ordered nucleosome structures within the coding region.

In the *cbh2* promoter region, in glucose-rich conditions, the CAE (*cbh2* activating element) region between -198 bp and -295 bp was reported to always be nucleosome-free and bound by the HAP2/3/5 complex and another (unknown) transcription factor in the presence of inducing and repressing carbon sources. The CAE was flanked by three upstream positioned nucleosomes and 2 CAE downstream positioned nucleosomes (between -49 bp and -193 bp and between +125 bp and -25 bp) which were displaced or disrupted upon induction in the presence of sophorose thus freeing the TATA binding site and TSS (Zeilinger *et al.*, 2003). The regulation of *cbh1* and *cbh2* is mediated by different transcription factors (e.g CRE1 binds to the *cbh1* promoter but not to *cbh2*, HAP2/3/5 binds to the *cbh2* promoter but not to *cbh1*, see introduction of this chapter) and this seems to also be reflected in the positioning of nucleosomes in both promoters (see above and below).

Nucleosome positioning was also studied in *T. reesei cre1* mutant strains and positioning patterns in the promoter region were the same as in the wild-type strain. There was a major difference though in nucleosome occupancy within the coding region of *cbh1* in the mutant strains when compared to the wild-type strain. The coding region of *cbh1* in the *cre1* mutant strains was completely devoid of nucleosomes. Two putative CRE1 binding sites, which are not occupied by nucleosomes, are present within the coding region of *cbh1* at positions +797 bp and +1609 bp, although their functionality is not verified. A truncated CRE1 protein (same mutation as the one used in this project) also led to a loss of positioned nucleosomes in inducing and repressing conditions within the *cbh2* promoter (Zeilinger *et al.*, 2003). In the *cbh2* promoter, CRE1 as well as the HAP2/3/5 complex and the unknown TF are required for correct nucleosome positioning by probably recruiting and

interacting with chromatin remodelling factors. The truncated CRE1 protein can still recognise its DNA binding sequence with only one zinc finger (Belshaw *et al.*, unpublished) but probably cannot interact with other proteins. Portnoy *et al.* (2011) reported that CRE1 represses SNF2, a component of the SWI/SNF ATP-dependent remodelling complex. Deletion or truncation of CRE1 would therefore prevent protein-protein interactions which could also mediate repression by positioning nucleosomes in the coding region of *cbh1*, even if a truncated CRE1 protein can still bind to the DNA. This indicates that CRE1 DNA binding is alone not sufficient for repression. A loss of nucleosomes within the coding region would facilitate transcription and this would explain elevated levels of *cbh1* in the presence of glucose in the *cre1* mutant strains when compared to the wild-type strain (see Figure 5.3.4). It has been reported before, that TFs are capable of recruiting and interacting with chromatin remodelling complexes as for example in the *A. nidulans alc* regulon (Mathieu *et al.*, 2005; also see introduction for examples). It is possible that CRE1 could bind to the two sites within the *cbh1* coding region and interact with other chromatin remodelling factors (e.g ISWI components) in order to ensure correct nucleosome positioning within the *cbh1* coding region in repressing conditions. These nucleosomes could then provide platforms for remodelling complexes when glucose is depleted and basal expression or induction of *cbh1* occurs which requires nucleosome displacement or eviction. Recent evidence suggests that nucleosome phasing downstream of the +1 nucleosome is related to transcriptional initiation and RNAPII binding although the exact mechanism of how RNAPII interacts with nucleosomes and how it displaces them during transcription is unknown (Arya *et al.*, 2010). The progress of RNAPII results in downstream positive-super coiling, which is linked to nucleosome destabilisation (García *et al.*, 2004).

Thus, mutation of CRE1 impaired nucleosome positioning within the *cbh1* coding region but not within the promoter region. It is likely that the promoter nucleosomes are positioned due to the underlying DNA sequence and that the coding region nucleosomes are positioned by ATP-dependent re-modellers. This is also the case in the yeast *his3-pet56* promoter region, where a near absence of nucleosomes is due to intrinsic DNA sequence preferences of histones (Sekinger *et al.*, 2005).

Sophorose was shown to be capable of inducing high transcript levels of *cbh1* in the wild-type strain (and possibly the mutants too). Analysis of the positioning of nucleosomes in the *cbh1* promoter and coding regions in the presence of sophorose revealed a near identical positioning pattern between the wild-type and the mutant strains. The -2 nucleosome seen under glucose-rich conditions remained in the same position in the presence of sophorose (-

3 nucleosome). The -2 nucleosome was also roughly in the same position as nucleosome -1 under *cbh1*-repressing conditions. The coding region was nucleosome-free. Furthermore there were two weak MNase cutting sites at -546 bp and +4 bp which likely correspond to two partially positioned or "fuzzy" nucleosomes. García *et al*., (2004) reported an increase in "fuzzyness" of MNase-generated bands in inducing conditions in the *A. niger prnD-prnB* bidirectional promoter. They expected this if in a given segment of DNA nucleosomes are present but not transcriptionally positioned. Another explanation for partially positioned nucleosomes is the fact that at a given moment in a cell population, there will be many possible chromatin states including cells in which RNAPII initiates transcription, cells in which elongating RNAPII is present at different places on the gene and cells in which the gene is in a non-transcribed state (Clark, 2010). Sophorose is probably not taken up in equal amounts by all the fungal cells and thus it is likely that there is a heterogenous mix of *cbh1* expression levels in the culture. In some cells, sophorose may have been used up already and the cells return to a non-induced state, thus placing a nucleosome over the TATA and TSS sites to prevent RNAPII binding (Figure 5.3.13). This would explain the positioning of the first "fuzzy" nucleosome. The second "fuzzy" nucleosome occupied a similar position as the one seen in the presence of glucose, thus the XYR1 and ACE2 binding sites at around -700 bp/-800 bp where nucleosome-free. Due to the heterogeneity of mycelial cells, sophorose would continue to induce *cbh1* in some cells by preventing CRE1 binding through XYR1 recruitment. In other cells, most of the sophorose will have been converted to glucose which is detected by the cells and which induces CCR of *cbh1* and allows CRE1 binding. Different nucleosome positioning patterns have also been observed under different physiological conditions in the *prn* gene cluster in *A. nidulans*. The promoter region between the genes *prnB* (proline transporter) and *prnD* (proline oxidase) is occupied by 8 positioned nucleosomes in repressing conditions, and these nucleosomes are lost under inducing conditions and a new nucleosome is positioned (Scazzochio and Ramón, 2008). Furthermore, different nucleosome positioning patterns are also observed in this region under inducing-repressing, amino acid starvation conditions and in the presence of trychostatin A (HDAC inhibitor).

In the presence of cellulose, the *cbh1* promoter and coding regions were nucleosome-free when compared to the positioning pattern in sophorose and glucose. This confirms an earlier observation that sophorose and cellulose induce *cbh1* transcription through different molecular mechanisms (Mach and Zeilinger, 2003) which are also apparent at the chromatin level. In the *cre*1 mutant strains, two weakly positioned nucleosomes were observed, of which the first one covers approximately the same DNA sequences in the promoter

region as the first "fuzzy" nucleosome seen under sophorose conditions. The second weakly positioned nucleosome occupies roughly the same position as the +4 nucleosome in glucose at the 3' of the *cbh1* ORF.

Thus it appears that CRE1 is involved in establishing local chromatin structure under *cbh1*-repressing conditions but not in inducing conditions. CRE1 has been described as a cell-wide regulator of carbon assimilation (Portnoy *et al*., 2011 and Chapter 4) and deletion of this gene affects many pathways which may indirectly influence *cbh1* nucleosome positioning during inducing conditions. This may explain the differences observed in the nucleosome positioning patterns in *cbh1* inducing conditions between wild-type and mutant strains. A thorough investigation needs to be carried out to elucidate the full interaction of CRE1 with factors involved in chromatin remodelling.

The interpretation of data from Southern blots needs to be done very carefully. There is no doubt that there are changes in the chromatin structure of genes when the fungus is incubated in different carbon sources but the above described results may look slightly different. Usually, digestion of protein-free gDNA with MNase gives a clear picture of the preferential cutting sites (independent of the presence of nucleosomes), which can be eliminated in subsequent nucleosome positioning assays (Clark, 2010). This was also done here (160 µg gDNA was used per MNase digestion assay), but resulted in the degradation of most gDNA after digestion with restriction enzymes. As a result no preferential cutting sites were observed but it is possible that they exist in the *cbh1* promoter region. Furthermore, many bands resulting from MNase were faint (although a minimum of 160 µg DNA was loaded in each lane) and could therefore lead to misinterpretation. However, all blots were independently repeated at least two times and deliberately over-exposed to help visualise faint bands (see Appendix A9 Figure A.9.F.4 for an example of a gel from which DNA was blotted and over-exposed). Optimisation of the chromatin digestion procedure also needs to be carried out for solid substrates such as cellulose. Mycelia associate to solid substrates and therefore MNase digestion is carried out in the presence of both the substrate and the mycelia (as both cannot be separated) which can lead to low concentrations of DNA and over-digestion by MNase, resulting in low resolution of bands.

As carbon sources were shown to influence cellulase and hemicellulase gene transcript levels and chromatin structures, this was also investigated at the protein level. Hence, protein profiles of the supernatants of QM6a mycelia grown in glucose, cellobiose, xylose, glycerol and sophorose were established. The protein profiles differed between the various carbon sources but were

very similar between xylose and glycerol. The most protein bands were observed in the presence of sophorose and the least amount of protein bands was seen in glucose (Figures 5.3.5). This was also confirmed by the measurements of total protein concentration of the supernatants from each carbon source: at the time points examined here, mycelia secreted more than twice the total amount of total protein in sophorose as in glucose, cellobiose, xylose and glycerol (data not shown). The total protein concentration is lowest in glucose and glycerol and is similar in xylose and cellobiose.

Polypeptides from 8 bands on polyacrylamide gels, and which appeared to be different between the conditions studied, were then sequenced by mass-spectrometry (Figure 5.3.5). Band1 corresponded to a GH3 and is expressed in all carbon sources except for glucose (Figure 5.3.5). This protein also appeared to be secreted in higher amounts in cellobiose and sophorose than in xylose and glycerol as indicated by a more prominent protein band. Band1 was matched to Cel3b, a β-glucosidase belonging to the GH3 family. These enzymes are involved in the cleavage of cellobiose to individual glucose units or they may be able to form sophorose from cellobiose by transglycosylation activity. Sophorose is a known inducer of cellulases and like Cel3a (encoded by *bgl1*), Cel3b is secreted in the presence of sophorose but also in the presence of cellobiose. The genome of *T. reesei* encodes several GH3 family β-glucosidases and these enzymes may have different functions in the presence of different/more complex substrates. Band2 was seen in the presence of cellobiose, sophorose and glycerol (although bands were weak in the latter two carbon sources) and sequencing revealed it to be a exo-1,3-β-glucanase belonging to family 55 of glycoside hydrolases. This GH family is not very well characterised but enzymes belonging to it cleave off glucose residues at the non-reducing end of glucan chains (www.cazy.org). Hence, this enzyme may also act on cellobiose or at least is induced and secreted when this substrate is detected by the cell. As cellobiose results from cellulose breakdown, this enzyme may be more directly involved in the degradation of cellulose chains. Band3 was present in glycerol, xylose and sophorose where it appeared to be secreted at high levels. This band corresponded to an endo-1,4-β-xylanase termed XYN4 belonging to the GH 5 family. Xylanases cleave randomly within xylan chains, releasing xylo-oligosaccharides. In contrast to XYN1, but like XYN2, XYN4 was induced and secreted in the presence of xylose and also partially in glycerol and sophorose, indicating that the genes encoding different xylanases in *T. reesei* are subject to different regulatory mechanisms. Band4 was identified as an epoxide hydrolase; these enzymes use water to cleave glycosidic bonds between aromatic residues. Epoxide hydrolases act on residues such as cyclic ethers which can be found in carbohydrate and lignin residues. In this study, epoxide hydrolase was secreted mainly in the presence

of xylose and sophorose but also glycerol. In nature, hemicellulose and cellulose are usually found in conjunction and this enzyme is likely to help release the different sugar polymers from lignin by acting on aromatic residues. This enzyme was therefore secreted when inducers such as xylose and sophorose are detected which normally would result from the degradation of plant cell walls. Sequencing of bands 5 and 6 revealed both to be CBH1 and this enzyme was secreted in very high amounts in the presence of sophorose. This enzyme was only secreted in the presence of sophorose and this was also confirmed by transcript and enzyme activity analysis. CBH1 belongs to the GH7 family and is the main *T. reesei* cellulase acting on the reducing ends of cellulose chains and is known to be induced in the presence of cellulose and sophorose. Unfortunately no identification for band 7 was obtained. Band 8 was only present in supernatants from cultures grown in sophorose. This band was identified as an acid phosphatase-like protein which acts on ester bonds. These bonds are found within the hemicellulose residues of plant cell walls. As mentioned before cellulose and hemicellulose are closely associated in plants and enzymes such as esterases also help to deconstruct other linkages found in plant cell wall sugars, thus releasing cellulose and hemicellulose sugars. Thus the presence of sophorose induced a far greater array and amount of enzymes which have a role in releasing the cellulose and/or hemicellulose fractions of plant cell wall polymers than when compared to glucose or cellobiose.

5.5 Conclusion

The regulation of cellulase-encoding genes is very different from the genes encoding hemicellulases in the carbon sources assayed here. Cellulolytic

genes were repressed in the presence of high concentrations of simple sugars such as glucose, xylose and cellobiose, whereas the repression of hemicellulolytic genes in the same carbon sources was less severe. Known inducing substrates such as cellulose and sophorose promoted the induction of all genes examined except for *xyn2*. Expression profiles also differed between individual genes. Thus the response of gene induction is graded and not all genes were expressed at the same magnitude in the presence of various different carbon sources. As shown in *cre1* mutant strains, carbon catabolite de-repression is not enough to trigger full gene induction but an inducer such as cellulose or sophorose is needed. Again, expression profiles of hemicellulase-encoding genes in these mutants differed from cellulolytic gene expression patterns, indicating that the mechanism of repression by CRE1 is slightly different for each gene. Cellulose and sophorose also elicited a different transcriptional response of the genes assayed here, as well as a different nucleosomal positioning pattern within the *cbh1* promoter and gene regions, supporting previous findings that gene expression and regulation are different between these two substrates. Other carbon sources also caused changes in the chromatin structure of the *cbh1* ORF. Under repressing conditions (glucose) the *cbh1* promoter only contained 2 positioned nucleosomes whereas the coding region was occupied by 4 nucleosomes. Two transcriptional regulators, CRE1 and ACE1, are predicted to be bound within the promoter region and were proposed to ensure, together with the nucleosomes, *cbh1* repression. A non-functional CRE1 protein caused a loss of nucleosomes within the coding region but not the promoter, indicating that CRE1 is involved in nucleosome positioning within the gene under repressing conditions by possibly interacting with chromatin remodelling factors. Under inducing conditions (sophorose) the *cbh1* coding region is nucleosome-free and the promoter is occupied by 2 strongly and 2 partially positioned nucleosomes. The latter 2 nucleosomes prevent CRE1 or RNAPII from binding depending on which state of transcription the cell is at. In the presence of cellulose, the *cbh1* ORF was nucleosome-free in all strains tested here. Mutations of CRE1 did not affect chromatin structure under sophorose and cellulose conditions. All the above described results indicate a very tightly controlled and essentially different regulation of the main cellulolytic and hemicellulolytic genes in *T. reesei* in the presence of simple and complex carbohydrates. Although the main hydrolytic genes were not always up-regulated in the presence of simple sugars, several other hydrolytic enzymes such as esterases, additional xylanases and glucosidases were secreted in the presence of the substrates assayed in this study. These additional enzymes help to catalyse the release of the cellulose or hemicellulose fractions from lignocellulose. Thus plant cell wall polymers and the resulting inducers elicit a

much greater response resulting in the transcription and secretion of numerous proteins (as already described in Chapter 4).

Chapter 6: General discussion

Replacing fossil fuels with biofuels will help to reduce global CO_2 emissions, have a favourable greenhouse gas profile, decrease dependence on oil and diminishing oil resources, avoid security issues and promote local economies (Banerjee *et al.*, 2010; Fukuda *et al.*, 2009). The production of biofuels from lignocellulosic plant biomass, that does not compete for food-use, is called second generation (2G) and is an environmentally friendlier alternative to petroleum-based energy sources (Alvira *et al.*, 2010). Lignocellulosic wastes are produced by the forestry, pulp and paper and agriculture industries in addition to municipal and animal wastes and are energy sources for conversion to biofuels (Dashtban *et al.*, 2009). Other potential energy sources are grasses such as Switchgrass and *Miscanthus* (Dodd and Cann, 2009) as well as hardwood and softwood (Zhu *et al.*, 2010).

One of the key steps in the conversion of lignocellulosic plant biomass to biofuels is the saccharification of plant cell wall components. Lignocellulose is the major structural component of plant cell walls and is made of cellulose (40% – 60%) which consists of glucose units, hemicelluloses (20% - 40%) which are composed of a mix of pentose sugars (xylose, arabinose) and hexose sugars (mannose, glucose, galactose) and several acids; and lignin (10% - 25%) which consists of aromatic alcohols (Dashtban *et al.*, 2009; Mathew *et al.*, 2008). The degradation of cellulose and hemicellulose, into mono-, di- or oligosaccharides is catalysed by enzymes secreted by filamentous fungi such as *Trichoderma reesei*.

These enzymes are known as glycoside hydrolases (GHs), polysaccharide lyases (PLs) and carbohydrate esterases (CEs), and can cleave a vast and diverse array of glycosidic bonds present within complex carbohydrates. As a result, hetero-synergy occurs between the enzymes, meaning that different hydrolases, lyases and esterases act synergistically on the main chain and the side chains of carbohydrates in order to degrade the substrate as much as possible. GHs, PLs and CEs are classed into different families based on their primary amino acid sequence in the carbohydrate active enzyme database (CAZy - www.cazy.org). The genome of *T. reesei* encodes 201 GHs, 22 CEs and 5 PLs (Häkkinen *et al.*, 2012).

T. reesei is known for secreting high amounts of cellulases (GHs which degrade cellulose) under biotechnological conditions, which are employed in many industrial applications such as in the pulp and paper (to reduce chlorine utilisation), textile and food industries (to improve digestibility and arabinoxylan degradation) (Knob *et al.*, 2010). Secreted protein levels of up to 100 g/L have been reported for some *T. reesei* strains under production conditions (Banerjee *et al.*, 2009). As a result, a wide selection of genetic tools

is available for this fungus, including random and targeted mutagenesis which has led to the development of cellulase hyper-producing strains, the partial elucidation of regulatory mechanisms and pathways of enzymes involved in the metabolism of various sugars and the engineering of the *T. reesei* secretion system in order to produce higher yields of more efficient and thermostable enzymes (Ayrinhac *et al.*, 2011; Dashtban *et al.*, 2009; Liu *et al.*, 2008; Mathew *et al.*, 2008; Rahman *et al.*, 2009). Due to its industrial importance and the already widespread application of its enzymes, *T. reesei* is an important fungus for use of its enzymes in biofuel production.

Despite the commercial importance of *T. reesei*, the regulation of genes encoding proteins involved in carbohydrate deconstruction is under-explored. The cost of production of these enzymes is a major limiting factor for industrial bioethanol production (Jeoh *et al.*, 2008; Klein-Marcuschamer *et al.*, 2012). The overall aim of this study was therefore to study the regulation of genes encoding proteins important for carbohydrate degradation at the transcriptional and post-transcriptional level when grown on various different substrates, including wheat straw, a potential source of lignocellulosic material for bioethanol production.

Chapter 3 described the influence of different sugars on fungal spore development and mycelial growth. In the presence of glucose and cellobiose (glucose disaccharide which can result from the degradation of cellulose), germination of conidia and subsequent hyphal extension were quicker than when compared to growth in xylose (hemicellulose component) and glycerol. Glucose is the preferred carbon source for many fungi and molecules can enter glycolysis directly after being internalised, where they are used to produce ATP. High internal ATP concentrations have been shown to coincide with high growth rate in *T. reesei* QM9414; i.e. a faster accumulation of ATP supports faster growth (Fuji *et al.*, 2010). Extracellular and intracellular BGLs cleave cellobiose, which can be taken up by the cell, into its glucose components (Suto *et al.*, 2001; Kubicek *et al.*, 2009). This may explain, at least in part, why germination in glucose and cellobiose is faster than in the presence of xylose and glycerol. Xylose needs to be converted to D-xylulose-5-phosphate first, a process which requires ATP hydrolysis, before it can enter the pentose phosphate pathway and generate energy (Mach-Aigner *et al.*, 2011). Thus, xylose utilisation is a less exergonic metabolic process and could be the reason for later spore germination. Glycerol is a polyol which is used by many fungi as a regulator of osmotic pressure within the cell and which requires to be transported into the cell. The later germination and accumulation of biomass in glycerol-rich minimal media when compared to growth in glucose-rich conditions could be due to osmotic pressure regulation

and the expression of a transport system for glycerol uptake. In the presence of sophorose, a known inducer of cellulases in *T. reesei* (Mach *et al.*, 2003), growth was much slower when compared to glycerol and yielded less mycelial biomass. The most likely explanation for this is that the induction and secretion of high levels of cellulases uses ATP which otherwise, e.g. in the presence of glucose and cellobiose, would be exclusively used for mycelial growth. Sophorose uptake has been shown to be inhibited in the presence of glucose (Kubicek *et al.*, 2009).

A direct correlation between fungal biomass production and growth was therefore observed and depended on the type of carbon source present. This was also the case when solid media supplemented with different carbon sources was inoculated with *T. reesei* spores: the most complex substrates (e.g. cellulose and wheat straw) sustained less radial outgrowth of the colony when compared to simple sugars such as glucose. Lignocellulosic substrates triggered the expression of enzymes involved in carbohydrate degradation (Chapters 4 and 5), a process which requires energy. In the presence of glucose, expression of these enzymes was repressed (Chapters 4 and 5) and ATP can be used mainly for colony growth. Thus different carbon sources influence growth rate, biomass production and subsequent enzyme secretion.

The process by which a carbon source such as cellulose signals its presence to conidia, causing spore germination and polarised hyphal growth remains largely unknown. One theory is that enzymes attached to the spore cell wall (cellulases such as CBH2 and EGL2) are thought to cleave off soluble inducers, which are detected by the cell and subsequently cause spore development (Suto *et al.*, 2001). The exact mechanisms for triggering germ tube emergence, maintenance of polarised growth and hyphal branching in filamentous fungi is largely unknown with only a few genes having been shown to be directly involved. It was shown in Chapter 3, that germination activated the UPR (unfolded protein response) by increasing transcript levels of genes which encode the main components of the UPR (e.g. PDI1, BIP1, HAC1 and SEC63). This indicates that germination is accompanied by an increase in protein synthesis and secretion. This has previously been reported during the isotropic expansion phase of spores, a step which is also required to break spore dormancy (Wendland, 2001). After isotropic expansion, a site for germ tube emergence is selected, a process which is mainly unknown in filamentous fungi and differs in many aspects from the well characterised process in *S. cerevisiae* (Harris *et al.*, 2004). Three genes, *rax2* (encodes a protein involved in vesicle transport and exocytosis), *axl2* (encodes a trans-membrane protein involved in establishing cell polarity and bud site selection in animals and yeast respectively) and *bud4* (encodes a protein likely to

interact with G-proteins and protein kinases), characterised in this study, could be potential candidates involved in the process of germ tube emergence although a thorough investigation into the roles of these genes and the proteins they encode needs to be carried out. As many different genes are proposed to be involved in site selection for germ tube emergence and maintenance of polarised growth, these three genes may only play a minor role and their encoded proteins are likely to interact with other components involved in signalling and protein secretion. Once a germination site has been chosen, polarity information is transduced by Rho GTPases to the cells morphogenetic machinery. Transcript levels of two Rho GTPases, termed MOD1 and RAC1, in *T. reesei* increased during the first few hours in conidia which were incubated in the presence of different carbon sources. These enzymes were shown to relay the signal for spore germination and hyphal extension to the cells' morphogenetic machinery in *S. cerevisiae* and *A. nidulans* (Momany 2002, 2005). Comparison of the amino acid sequences of these Rho GTPases encoded by the corresponding genes in *T. reesei*, *A. nidulans* and *S. cerevisiae* revealed sequence identity of over 65%. It is therefore likely that the two Rho GTPases RAC1 and MOD1 fulfil a similar role in *T. reesei*. Polarised growth is maintained by the regulation of the cellular cytoskeleton and the Spitzenkörper (structure which organises the delivery of secretory vesicles at the growing tip and which is required for directional growth) at polarisation sites. In filamentous fungi, the morphogenetic machinery and the Spitzenkörper are controlled by multi-protein complexes such as the polarisome, through downstream signalling by PAKs (Harris *et al.*, 2004; Araujo-Palomares *et al.*, 2009). Transcript levels of 3 main components of the polarisome (SEP1, SPA1 and BUD1) increased throughout the first 6 hours of spore incubation in the presence of different carbon sources indicating that they play a role, as has been suggested in *A. nidulans* (Meyer et al., 2008), in regulating polarised growth in *T. reesei*. The steps connecting the Spitzenkörper to the morphogenetic machinery and the polarisome are largely unknown in *T. reesei* and could provide new areas of research in order to understand fungal growth on various different substrates important for biofuel production.

Chapter 4 described the results obtained from the sequencing of transcriptomes of mycelia grown in three different conditions (glucose, straw and straw + glucose) in order to elucidate the mechanism employed by *T. reesei* for lignocellulose substrate recognition and subsequent degradation. The biggest differences in gene expression were between the glucose and straw conditions and classification of significantly differently regulated genes into KEGG groups revealed that the majority of genes encoded proteins involved in carbohydrate and amino acid transport and metabolism, enzymes

involved in energy production and conversion as well as genes encoding proteins required for transcription, post-translational modifications, protein turnover, signal transduction, lipid metabolism and secondary metabolite biosynthesis. The addition of glucose to straw caused the expression of genes encoding proteins involved in cytoskeleton and extracellular structure remodelling as well as proteins assisting in nucleolar transport and metabolism indicating the induction of a response in the nucleus (maybe by repressing the transcription of many genes) as well as having an effect on cell growth and morphology. This is in agreement with the observed results in Chapter 3 where glucose was proposed to primarily fuel fungal growth.

Genes encoding enzymes involved in carbohydrate transport and metabolism represented one of the most numerous groups. GHs, CEs and PLs involved in the deconstruction of complex carbohydrates were highly up-regulated in the presence of wheat straw (CAZy gene RNA represented 8.3% of total cellular RNA) and were repressed in glucose conditions (0.58% and 0.23% of total cellular RNA for the two glucose conditions). 79 GH-encoding genes (Appendix A3 Table A.3.T.1) representing 37 families and 7 CE-encoding genes representing 5 families were significantly differently regulated between the three conditions. Thus, wheat straw which is composed of cellulose, hemicelluloses and lignin, induced many genes encoding a variety of enzymes. The transcript patterns of a few cellulase (*cbh1*&*2*, *egl1*, *bgl1*), hemicellulase (*xyn1*, *xyn2*, *bxl1*) and GH-encoding genes (families 2, 47, 67, 76 and 92) were also confirmed by RT-PCR and, for some, by qRT-PCR. A 24 h straw time course revealed the presence of transcripts of *cbh1*, *xyn1*, *xyn2* and *bgl1* after 6 h (*cbh1*), 2 h (*xyn1* and *xyn2*) and 1 h (*bgl1*) incubation in straw. Transcript abundance of these genes was very high at 24 h indicating full induction of these genes. Transcripts of these genes were absent in a 24 h time course in media containing no carbon source, indicating that straw is detected and recognised as a potential carbon source.

The same time course experiments were carried out for genes encoding 3 GH61 enzymes. GH61 proteins are copper-dependent oxidoreductases (Quinlan *et al.*, 2011), shown to significantly enhance cellulase activity (Harris *et al.*, 2010) and are thus not glycoside hydrolases but function as accessory enzymes which increase cellulose hydrolysis. Transcript patterns between the three GH61-encoding genes differed in the presence of straw indicating differential regulation of these genes. In the absence of any carbon source, transcript patterns were different to those observed in the presence of straw indicating that these genes are up-regulated as a result of the fungus being able to degrade straw.

The *T. reesei* genome was also screened for accessory enzymes (e.g. swollenin) and non-CAZy-encoding genes which could potentially be involved in carbohydrate deconstruction. Twenty-seven genes, encoding non-CAZy enzymes, had an RPKM greater than 50 and were up-regulated over 20-fold in straw-rich conditions (Appendix A4 Table A.4.T.1). These genes encode enzymes of various functions including lipases, surface interacting proteins (e.g. hydrophobins), enzymes involved in xylose metabolism and MFS transporters. MFS transporters are a large family of proteins involved in importing sugars (e.g. glucose, xylose) into the cell. They may play a role in substrate detection i.e. by importing inducer molecules such as sophorose. Sophorose - a β-1,3 linked glucose disaccharide - is thought to be formed from the transglucosylation activity of BGL1 in the presence of cellobiose and cellulose (Mach *et al.*, 2003). Two genes also encoded fungal specific transcription factors which may play a role in regulating genes involved in carbohydrate utilisation. All these genes and the proteins they encode present interesting areas for further research. The transcript patterns of two genes encoding a hydrophobin (*hfb3*) and a swollenin (*swo1*) as well as two genes encoding GPCRs (one PTH11-type and one GPRK-type which were up-regulated over 10-fold) were confirmed by RT-PCR (and some by qRT-PCR). As observed for the CAZy-encoding genes, transcript abundance of *pth11* and *swo1* was decreased in the absence of wheat straw (no carbon source media) indicating that the PTH11-type GPCR may play a role in sensing/mediating the presence of a carbon source and that SWO1 is also involved in carbohydrate deconstruction. In the absence of any carbon source, *hfb3* and *gprk* transcripts were present at much earlier time points when compared to their patterns in straw-rich conditions. Carbon starvation is thus inductive for these two genes. It is possible that these genes encode "scouting" enzymes which under carbon starvation conditions scavenge the extracellular environment for potential substrates. A similar model was proposed for *A. niger* (Delmas *et al.*, 2012). In *T. reesei* screening for other "scouting" enzymes such as GHs, lipases or transporter is necessary in order to confirm or reject this hypothesis.

In order to investigate the regulatory role of the above mentioned CAZy and non-CAZy-encoding genes, the same experiments were carried out in a *cre1* knock-out strain. CRE1 is the main carbon catabolite repressor (see Chapter 5) which in the presence of glucose represses CAZy-encoding genes but is also involved, directly or in-directly, in the regulation of genes involved in nitrogen uptake, chromatin remodelling and developmental processes (Portnoy *et al.*, 2011). In agreement with other studies, alleviation from CRE1-mediated repression caused an earlier onset of transcript abundance of the CAZy-encoding genes; the only exceptions were genes encoding a GH76 family

protein, a GH92 family protein and the CEL61b protein. The transcript abundance of these genes was either decreased or completely absent indicating that CRE1 is required for induction of these genes. Genes encoding SWO1, HFB3 and GPRK are also under the influence of CRE1 repression as transcript levels of these genes in the Δ*cre1* strain was higher than in the wild-type strain. Transcript levels of the PTH11-encoding gene did not change between the wild-type and the mutant strain and thus this gene is probably not under the regulatory control of CRE1.

Chapter 4 also described the comparison of the transcriptional response to the presence of wheat straw between *T. reesei* and *A. niger*. The genome of *A. niger* encodes 45 more GHs, 3 more PLs and 3 more CEs than the genome of *T. reesei*. The total amount of CAZy RNA (out of the total amount of cellular RNA) was much higher in *A. niger* in all three conditions (2.7% 48 h glucose, 19.1% 24 h straw and 1.4% 24 + 5 h glucose). Both organisms invest a lot in the transcription of CAZy genes considering that these genes only represent ~ 2.5% of the total coding genome. Both genomes also encode members from GH, CE and PL families which are not present in the other organism (see Chapter 4, Section 4.3.5.1). Despite this, GH families with the highest expression levels after 24 h incubation in straw were similar in both organisms (included GH families 3, 6, 7, 11, 61 and 67) indicating that they use a similar set of core enzymes to degrade wheat straw. Difference mainly lay in the expression of CE and PL-encoding genes.

In *T. reesei* GH61 transcripts made up ~6% (4 out of 6 GH61-encoding genes were induced) of the total cellular RNA whereas in *A. niger* they made up ~14% (5 out of 7 GH61-encoding genes were induced) of the total amount of cellular RNA in the presence of straw, further supporting an important role of these accessory enzymes in carbohydrate degradation. Primary amino acid sequence comparison revealed that the 7 *A. niger* GH61 proteins clustered into two groups whereas the 3 *T. reesei* GH61 proteins studied here, had low sequence similarity and could not be grouped together. This suggests that although *T. reesei* GH61 transcript levels were less abundant than those in *A. niger*, they may fulfil different roles whereas an expansion of closely related GH61 proteins in *A. niger* suggest activity on similar substrates.

The expression of non-CAZy-encoding genes (lipases, transporters, hydrophobic surface interacting proteins and enzymes involved in xylose metabolism) was similar between *A. niger* and *T. reesei* suggesting a similar approach of both organisms for lignocellulose degradation. The main differences lay in the expression of accessory enzymes: *A. niger* encodes a non-hydrophobin hydrophobic surface binding protein termed HsbA (up-

regulated more than 20-fold when switching from glucose to straw) which is absent from the genome of *T. reesei*. In *A. oryzae*, HsbA was shown to recruit CUTL1 (CE) to hydrophobic surfaces such as PBSA and promote substrate degradation (Ohtaki *et al.*, 2006). Instead, *T. reesei* induced the transcription of two expansin-like proteins (*swo1* and *eel*) in straw-rich conditions and these genes are absent from the genome of *A. niger*.

There was also a difference in the expression of genes encoding PTH11-type GPCRs: the genome of *A. niger* encodes 70 potential PTH11-type GPCRs which is double the number of genes encoding these types of proteins in the genome of *T. reesei*. Groups of cAMP-receptor like proteins, *H. sapiens* mPR-like proteins (steroid receptors) and *Aspergillus* GPRK (G-protein coupled recepor kinase)-like receptors are expanded in *T. reesei* (Brunner *et al.*, 2008). In *T. reesei*, 23 of PTH11-encoding genes were up-regulated in the presence of straw with 7 genes being expressed more than 10-fold when switching from glucose to straw (Appendix 5 Table A.5.T.3). In *A. niger*, 37 PTH11-encoding genes were up-regulated in the presence of straw when compared to glucose-rich conditions and 8 of these genes were up-regulated more than 10-fold (Appendix 5 Table A.5.T.3). In *M. grisea*, Pth11 GPCRs sense plant cutin monomers and interact with the hydrophobin MPG1 to mediate attachment of a spore to the hydrophobic plant surface (Wilson and Talbot, 2009). A similar role for these proteins could be proposed for *A. niger* and *T. reesei*.

Together, these results suggest that the mechanism for solid substrate detection and subsequent degradation in *T. reesei* is similar in concept, if not in the detail, to the model proposed for *A. niger* (Delmas *et al.*, 2012). Carbon starvation possibly triggers the induction of genes encoding "scouting" proteins (e.g. HFB3 and others, not discovered here) which mediate the attachment of the fungus to the substrate. This could serve as a cue for PTH11 and/or GPRK (and maybe others) activation which could then signal the presence of wheat straw to the cells and this would then trigger the induction of CAZy-encoding genes necessary for polysaccharide degradation. Low basal expression (Kubicek *et al.*, 2009) and the early presence of transcripts (2 h, 6 h after incubation in straw) of genes encoding major hydrolytic enzymes such as CBH1, XYN1 and XYN2 may further cleave off oligosaccharides which are then transformed into inducers such as sophorose by the transglycosylation activity of BGL1 and which serve as an additional signal for CAZy-gene induction. Expression of MFS transporters in the cell membrane would allow for efficient inducer and sugar uptake. This hypothesis proposes a basic model of how fungi can detect and subsequently trigger the degradation of solid lignocellulosic biomass and provides a platform for further studies.

Finally, Chapter 4 also described the detection of genes which contain NATs. NATs are a type of antisense regulatory RNAs which were considered to (partially) overlap open reading frames (ORFs) and which change in direction (sense/antisense to antisense/sense) between different conditions (Faghihi and Wahlestedt, 2009). Gene regulation by these types of NATs was easily identifiable and most interesting. Around 2% of all *T. reesei* genes had NATs which varied in abundance and length hinting at the type of regulatory role exerted by these NATs (see Chapter 4 for examples). Most of these genes encode proteins of unknown function whereas others are involved in carbohydrate and amino acid transport and metabolism and transcription. Thus post-transcriptional regulation also occurs in *T. reesei* and further studies are needed to elucidate the mechanism of regulation for each of these NATs.

Chapter 5 described in more detail the mechanisms of regulation which could govern some cellulase and hemicellulase-encoding genes. The expression of 4 cellulase (*cbh1*, *cbh2*, *egl1*, *bgl1*) and 3 hemicellulase (*xyn1*, *xyn2*, *bxl1*)-encoding genes was studied in the presence of different carbon sources (same as those used in Chapter 3). Glucose, cellobiose, xylose and glycerol did not induce cellulase gene expression. In the presence of glucose cellulase-encoding genes are repressed by CRE1, the main carbon catabolite repressor. Cellobiose was reported to induce cellulase gene expression by being converted to sophorose through the transglycosylation activity of BGL1 but cellobiose concentrations used in this project may have been too high and incubation time too low for this to happen. Xylose concentrations were also too high to induce cellulase gene expression and glycerol, as a neutral carbon source, did not cause cellulase gene induction either. Expression patterns of the hemicellulase-encoding genes differed from those observed for the cellulase-encoding genes. Transcript levels of *xyn2* did not change much between the carbon-sources used here and this gene was reported not to be under the control of CRE1. In the presence of glucose, *xyn1* and *bxl1* were repressed by CRE1 whereas in xylose, cellobiose and glycerol, transcript levels were higher indicating alleviation from CRE1-mediated repression. As mentioned before, xylose concentrations used here were too high (132 mM) to cause induction of hemicellulase-encoding genes. Xylose was shown to induce cellulolytic and xylanolytic genes at very low concentrations of 0.5mM to 1mM (Mach-Aigner *et al.*, 2010).

In the presence of sophorose all genes, except for *xyn2*, were induced. Transcript levels of cellulase-encoding genes, except for *bgl1*, were higher than those observed for the hemicellulase-encoding genes, probably because sophorose is a by-product of cellulose breakdown, a process which mainly requires the action of cellulases. Low levels of *bgl1* transcripts have been

reported before, probably because this gene is subject to CCR in the presence of glucose, a sugar that is released by the enzyme itself (negative feedback loop). Transcript levels of *bxl1* were also very high in the presence of sophorose and the enzyme encoded by this gene, BXL1, has been reported to have BGL and ABF activity. This is a good example of a GH with multiple specificities and BXL1 is probably involved in cellulose and hemicellulose breakdown. No good explanation could be found for the low transcript levels of *xyn2* and this may have to be investigated further. A different expression profile for the cellulase and hemicellulase-encoding genes was also observed in the presence of cellulose. Transcript levels of genes were lower (with the exception of *xyn1* and *xyn2* which were the same) in cellulose than when compared to sophorose probably because cellulose degradation (complex, insoluble carbon source) is much slower than sophorose uptake and full gene expression may be achieved at later time points. Transcript levels of *egl1* and *bxl1* were highest out of all genes in the presence of cellulose probably highlighting the necessity for EGL1 to randomly cleave cellulose chains and provide free chain ends for CBH1/CBH2; and the requirement for poly-specific enzymes such as BXL1 involved in degrading cellulose polymers. Transcript levels of *xyn1* and *xyn2* did not differ much between sophorose and cellulose conditions. The expression of *xyn2* decreased in cellulose as was also observed in the presence of sophorose and *xyn2* transcript levels were lower than those for *xyn1* as has been reported before (Margolles-Clark *et al.*, 1997). The xylanases encoded by *xyn1* and *xyn2* cannot degrade cellulose, but as cellulose is found in conjunction with hemicellulose in plant cell walls, induction of *xyn1* (and maybe *xyn2*, remains to be investigated) together with cellulase-encoding genes may present an advantage.

The role of CRE1 in the regulation of cellulase and hemicellulase-encoding genes was also investigated as this transcription factor was already shown to play a role in the regulation of quite a few genes in Chapter 4. Transcript levels of cellulase and hemicellulase-encoding genes in 3 *cre1*-deficient strains were increased when compared to the wild-type strain in the presence of glucose. It is possible that deletion of *cre1* (Δ*cre1* strain), truncation of *cre1* (*cre1-1* strain; retains one zinc finger) and all the mutations present in RUTC30 (Le Crom *et al.*, 2009) influence cellulase gene expression differently probably because CRE1 is a cell-wide regulator influencing many pathways involved in different cellular functions (Portnoy *et al.*, 2011). In the presence of cellulose, transcript levels increased markedly in all the mutant strains when compared to those in the wild-type strain. Alleviation from CRE1-mediated repression does not cause induction of these genes in the presence of repressive carbon sources such as glucose. In agreement with other studies (Nakari-Setälä *et al.*,

2007), the presence of an inducer (e.g. cello-oligosaccharides cleaved from cellulose chains) is required for full gene induction.

The results described in Chapter 5 show that the differences in regulation of genes involved in carbohydrate deconstruction, are extremely fine-tuned and dependent on factors such as culture conditions, consumption rate, assimilation of glucose as well as the presence of transcription factors. In eukaryotes such as *T. reesei*, transcriptional regulation of a gene can be mediated by the binding of different factors (activators and repressors) to multiple binding sites within the promoter region. The synergistic action of these regulators can either modulate the levels of gene expression or completely switch them on or off. In *T. reesei*, cellulase and hemicellulase gene expression is governed by the balance of positive and negative regulators such as ACE1, ACE2, the HAP2/3/5 complex and CRE1; and that they also interact with the essential, positive regulator XYR1. Recently, two new transcription factor-encoding genes termed *clr-1* and *clr-2* were described in *N. crassa* and which are thought to mediate cellulose degradation by up-regulating genes encoding 16 cellulases, 6 hemicellulases, 15 enzymes with activity on polysaccharides as well as proteins involved in the secretion pathway and the UPR (Coradetti *et al*., 2012). Orthologues of *clr-1* and *clr-2* in *T. reesei* (transcript I.Ds 27600 and 26163) were up-regulated over 5-fold when switching from glucose to straw and these transcription factors (and maybe others, see Appendix A4 Table A.4.T.1) may also play a role in regulating cellulase-encoding genes in addition to other known *T. reesei* regulators. The effect of the interplay of all these factors is graded so that they are individually or together only partially responsible for complete regulation of cellulolytic and xylanolytic genes (Aro *et al*., 2003).

This conclusion was also supported by the analysis of the profiles of secreted proteins in the presence of different carbon sources. Protein band patterns differed between each carbon source (glucose, cellobiose, xylose, glycerol and sophorose) highlighting the regulatory differences caused by the presence of different substrates. In the presence of glucose the least amount of proteins was secreted whereas the highest levels of protein bands were observed in the presence of sophorose. Sequencing of some protein bands from the sophorose conditions confirmed the presence of cellulases such as CBH1. This also supports the proposed model described in Chapter 3 in which glucose fuels faster growth and sophorose does not, due to the cells investing energy in the secretion of proteins.

Chapter 5 also described in detail the pattern of nucleosome positioning in the *cbh1* promoter and gene. Chromatin is a feature of all eukaryotes,

representing an additional level of gene regulation and a feature which is closely associated to the interaction with numerous transcription factors. Nucleosome positioning was different between *cbh1* repressing (glucose) and inducing (sophorose and cellulose) conditions. In the presence of glucose, the promoter region of *cbh1* contained two nucleosomes and CRE1 was reported to be bound at ~ -700 bp (Mach-Aigner *et al*., 2008) and ACE1 at ~ -200 bp (Ling *et al*., 2009). Together, the bound CRE1 and ACE1 transcription factors (TFs) and the two nucleosomes would prevent *cbh1* expression by inhibiting the activator XYR1 from binding to its recognition sequences. It has been reported before that transcription factors co-operate to dictate nucleosome positioning and establish nucleosome-free regions in highly regulated promoters (Cockerill, 2011). The coding region of *cbh1* contained 4 positioned nucleosomes. TATA-containing and nucleosome-occupied promoters tend to have an increase in nucleosome occupancy downstream of the TSS and active promoters are much less likely to be occupied by nucleosomes than coding regions (Sekinger *et al*., 2005; Arya *et al*., 2010; Clark, 2010).

In the *cre1*-deficient strains, the four positioned nucleosomes within the coding region were lost whereas those within the promoter region remained bound. This interesting observation led to the suggestion that CRE1 may play a role in nucleosome positioning by possibly recruiting chromatin remodelling factors. Portnoy *et al*. (2011) reported that CRE1 represses SNF2, a component of the SWI/SNF ATP-dependent remodelling complex. Loss of nucleosome positioning was also observed in the coding region of a wild-type strain which encodes a truncated CRE1 protein (strain *cre*1-1). Truncated CRE1 contains one zinc finger with which it can still recognise its DNA binding sequence (Belshaw *et al*., unpublished). Deletion and truncation of CRE1 probably prevent protein-protein interactions which may be involved in mediating *cbh1* repression. This indicates that CRE1 DNA binding is alone not sufficient for repression. Two putative CRE1 binding sites, which are not occupied by nucleosomes, are present within the coding region of *cbh1* at positions +797 bp and +1609 bp, although their functionality is not verified. It is possible that CRE1 could bind to the two sites within the *cbh1* coding region and interact with other chromatin remodelling factors (e.g ISWI components) in order to ensure correct nucleosome positioning within the *cbh1* coding region in repressing conditions. These nucleosomes could then provide platforms for remodelling complexes when glucose is depleted and basal expression or induction of *cbh1* occurs which requires nucleosome displacement or eviction. This model also explains why, in the absence of CRE1, *cbh1* transcript levels in the presence of glucose, sophorose, cellulose and wheat straw (Chapter 4) were increased: *cre1* deletion abolishes the requirement for nucleosome removal prior to transcription and *cbh1* gene

expression can readily occur. Furthermore, it has been reported that disrupting the coding region of *cbh1* diminished expression suggesting that an intact gene is required for transcriptional regulation (Rahman *et al.*, 2009). The exact roles of CRE1 and the coding region in nucleosome positioning are topics for future investigations.

CRE1 did not seem to play a role in nucleosome positioning under *cbh1* inducing conditions. In the presence of sophorose, two nucleosomes were present in the promoter region whereas the coding region was mainly devoid of nucleosomes (gene induction). One of the promoter nucleosomes was displaced by approximately 100 bp downstream when compared to the position of this nucleosome under glucose conditions. Displacement of this nucleosome prevents binding of the repressor ACE1 at around ~ -200 bp. Two weakly positioned nucleosomes were also observed in the *cbh1* promoter region: one nucleosome covered the TSS and TATA box whereas the other was placed over the CRE1 binding sites. An explanation for this could be a heterogenous mix of *cbh1* expression in a population of cells. Alternatively, mis-interpretation of the data, due to faint bands generated by MNase may provide another explanation.

In all strains tested here, the *cbh1* promoter and coding regions were mainly nucleosome-free under cellulose conditions. Nucleosomal assays in the presence of cellulose would need to be repeated though as experimental design needs to be optimised for analysis of the chromatin structure in the presence of solid substrates such as cellulose. Despite this, it was observed that nucleosome positioning pattern as well as *cbh1* transcript levels (described above) differed between sophorose and cellulose. This confirms an earlier observation that sophorose and cellulose induce *cbh1* transcription through different molecular mechanisms (Mach and Zeilinger, 2003).

In conclusion, the results presented in this thesis describe expression patterns and regulation of genes involved in plant cell wall degradation. Furthermore, the *T. reesei* genome was also screened for new, non-CAZy-encoding genes involved in carbohydrate deconstruction and which could be useful for strain improvement for biofuel production. It was shown that different carbon sources influence growth rate, biomass production, gene expression patterns and subsequent enzyme secretion. Growth in the presence of complex lignocellulosic substrates such as wheat straw and cellulose was slow and these substrates caused the up-regulation of many CAZy and non-CAZy-encoding genes; whereas growth in simple carbohydrates was very quick and substrates such as glucose caused the repression (mediated by CRE1) of these genes. Genes encoding new, potential non-CAZy proteins involved in solid

substrate recognition and degradation were described for spores and mycelia. A model for solid substrate recognition in *T. reesei* was described, based on the comparison with the one proposed for *A. niger*. Mechanisms of regulation at the transcriptional (chromatin) and post-transcriptional (regulatory RNAs) levels of genes involved in carbohydrate deconstruction were also described in *T. reesei*. CRE1 was shown to be one of the key factors for the regulation of these genes and several other regulatory roles were attributed to this transcription factor.

Further studies

Further studies characterising the roles of genes, and the proteins they encode, during the process of carbohydrate degradation would focus on the following:

- Creating knock-out mutants of genes encoding potential markers for bud site selection in germinating spores and analysing the resulting phenotype.
- Analysing the expression patterns of genes encoding MFS transporters, TFs, GPCRs and potential "scouting enzymes" (all of which were found to be highly up-regulated in the presence of straw) at different time points in the presence and absence of lignocellulosic substrates in order to find additional candidates involved in solid substrate recognition and which would support the model proposed in Chapter 4.
- Characterising NATs of genes involved in carbohydrate metabolism in order to elucidate their regulatory role.
- Deleting the *cbh1* coding region and analysing the resulting chromatin promoter structure.
- Carrying out chromatin immunoprecipitation and protein-protein interaction experiments to see whether CRE1 binds to the *cbh1* coding region and with which other proteins CRE1 may interact directly.

Appendix A0

Figure A.0.F.1: Microscopic analysis of spore development over a time period of 12 h when inoculated in TMM supplemented with 2% of the respective carbon source. Scale bars (Ⱶ) indicate size.

	0 h	2 h	4 h	6 h	8 h	10 h	12 h

Xylose

Glycerol
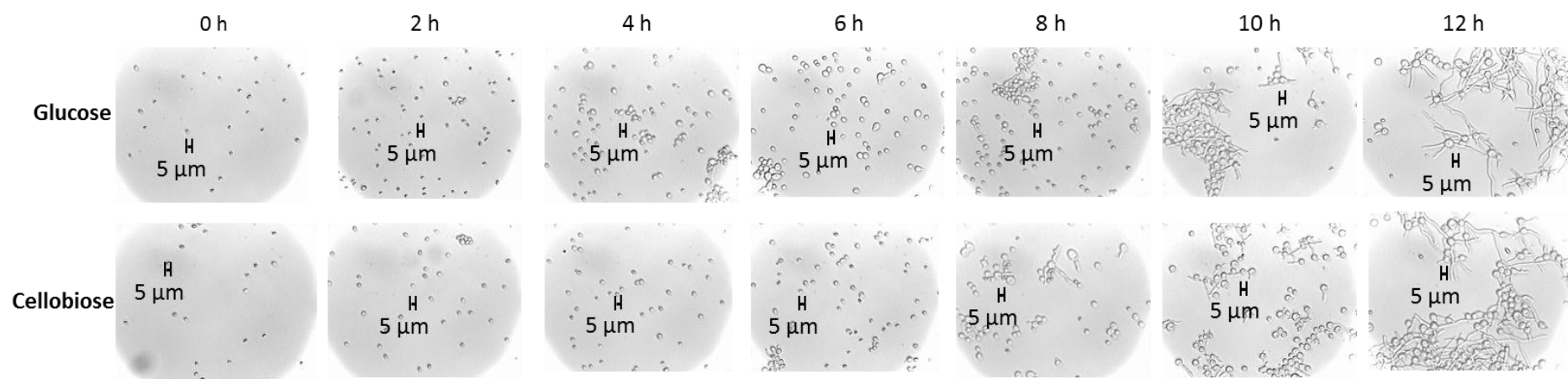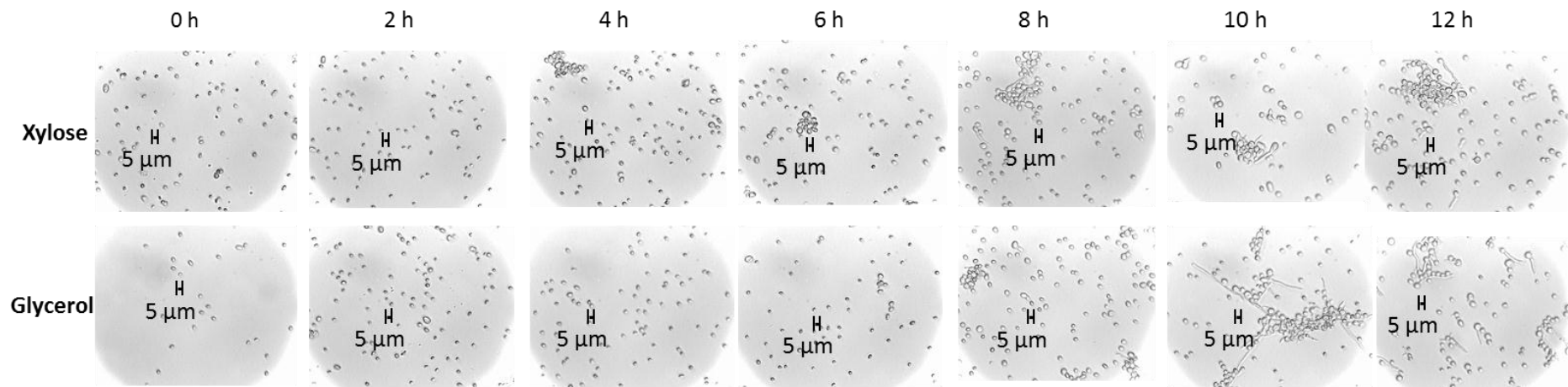
A.0.F.2: Semi-logarithmic graph of produced fungal biomass (g/L; average of three independent experiments) over a time period of 96 h in TMM supplemented with 2% (w/v) cellobiose. The Δy-value indicates the doubling of biomass whereas the Δx-value indicates how long it takes for the biomass to double. Specific growth rate and doubling time (Δx) were calculated from these values.

$$\Delta y = (0.2 \rightarrow 0.4)\, g/L$$
$$\Delta x = 4\, h = dt$$
$$\mu m = \ln 2 / dt = 0.17$$
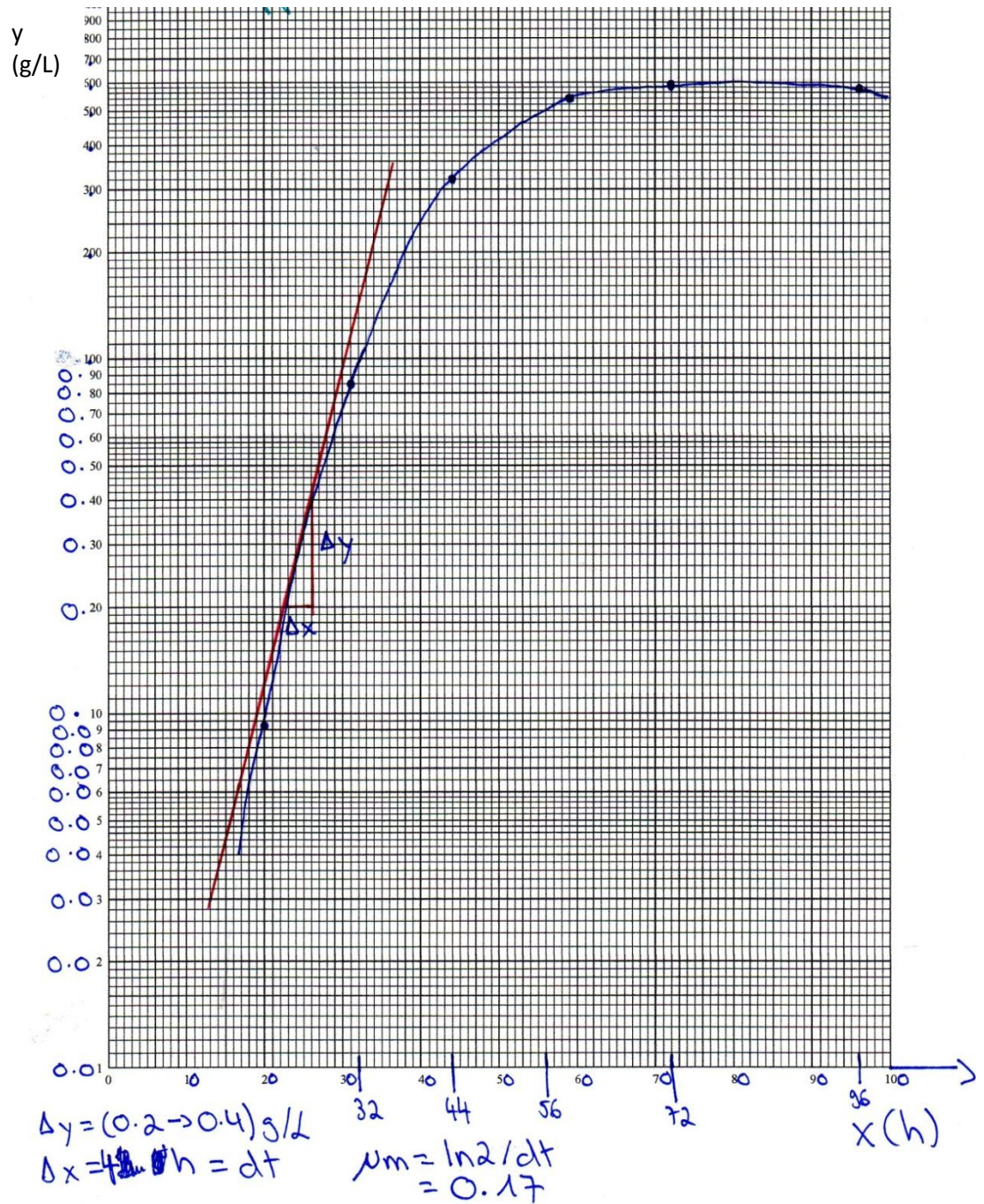
Appendix A1

Table A.1.T.1: Transcript I.Ds, names and annotations of *T. reesei* genes which were studied in the different chapters in this thesis.

Name	Transcript I.D.	Annotation
Chapter 3 Spore development		
Name	Transcript I.D.	Annotation
bud4	76029	Hypothetical protein
axl2	110910	Protein and calcium ion binding, one transmembrane domain, dystroglycan-type cadherin-like
rax2	76515	DNA binding, protein transporter and oxidoreductase activity, regulation of transcription, protein secretion; had HMG1/2 box, Sec1-like, amine oxidase
rac1	47055	Ras small GTPase
mod1	50335	Cdc42, small Rho type GTPase
sep	77031	Actin binding, cytokinesis protein sepA
spa1	108829	Class-II aminotransferase protein, involved in metabolism
bud1	35386	Hypothetical protein
pdi1	122415	Protein disulfide isomerase
bip1	122920	HSP70 family ER chaperone
hac1	46902	Transcription factor with bZIP
sec63	121754	DnaJ superfamily molecular chaperone, part of protein translocation complex
Chapter 4 Transcriptomics		
Name	Transcript I.D.	Annotation
hfb3	123967	Hydrophobin, class II
swo1	123992	Swollenin
gprk	37525	GprK-type GPCR
pth11	27992	PTH11-type GPCR
gh61	27554	Candidate endoglucanase GH61, *gh61*
gh61a	73643	GH61 cel61a, *gh61a*
gh61b	120961	Endoglucanase GH61 cel61b, *gh61b*
gh2	69245	Predicted beta-mannosidase, GH2
gh47	79960	Alpha-1,2-mannosidase, GH47
gh67	72526	Alpha-glucuronidase, GH67
gh76	55802	Endo-1,6 beta-mannosidase, GH76
gh92	79921	Putative alpha-1,2-mannosidase, GH92
aa1	81442	Amino acid transporter
bgluc	76852	Beta-Glucuronidase GUSB (GH 2) ATP coupled synthesis to proton transport, membrane bound
gh16	70542	Candidate glucan endo-1,3(4)-β-D-glucosidase
tf1	121412	Fungal specific TF
Chapter 5 Chromatin structure analysis		
Name	Transcript I.D.	Annotation
cbh1	123989	Cellobiohydrolase I, GH7

cbh2	72567	Cellobiohydrolase II, GH6
egl1	122081	Endoglucanase EG-1, GH7, CBM1 module
bgl1	76672	Beta-glucosidase 1, GH3
xyn1	74223	Xylanase 1, GH11
xyn2	123818	Endoxylanase 2, GH 11
bxl1	121127	Beta-xylosidase, GH3
act1	44504	Actin

Appendix A2
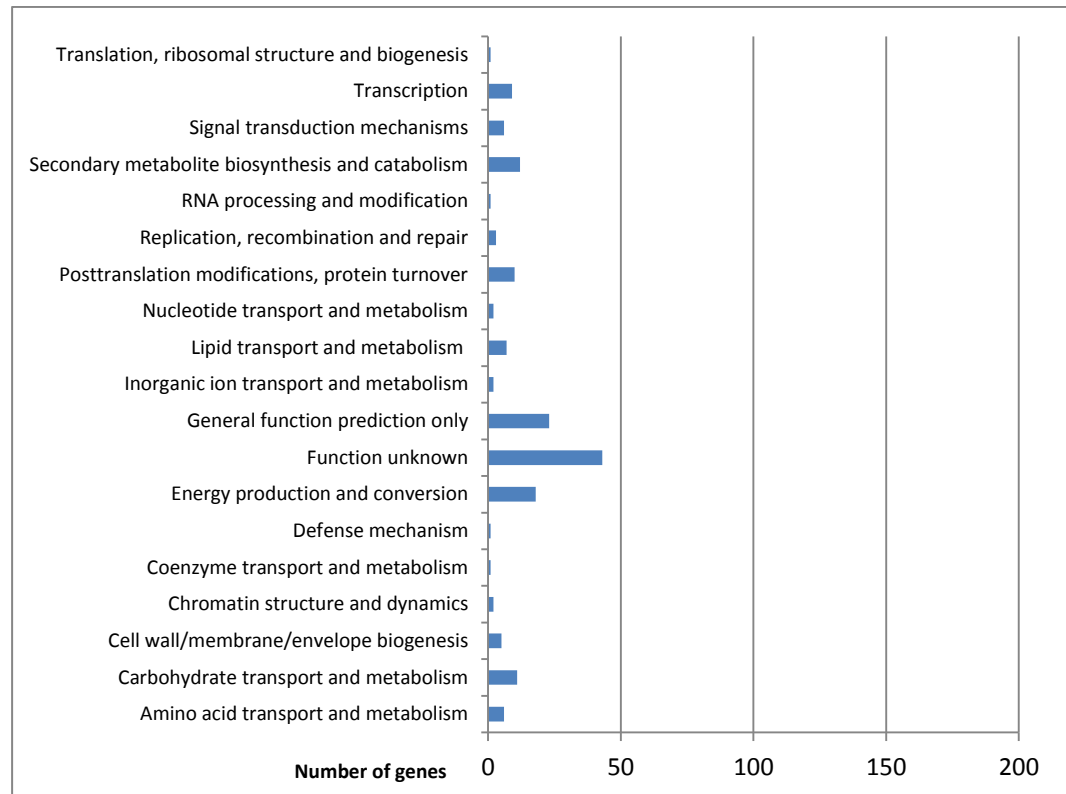
Figure A.2.F.1: Classification of genes which were significantly differently expressed between 48 h glucose and 24 h straw into their individual KEGG groups. To find all significantly differently expressed genes, three statistical significance tests were applied to changes in gene expression: the Likelihood Ratio Test (LRT; [Marioni *et al*., 2008]), Fisher's Exact Test (FET; [Bloom *et al*., 2009]) and an MA-plot-based method with Random Sampling model (MARS; [Wang *et al*., 2010]). All significantly differently expressed genes had a p-value of <0.001 for all three statistical tests.
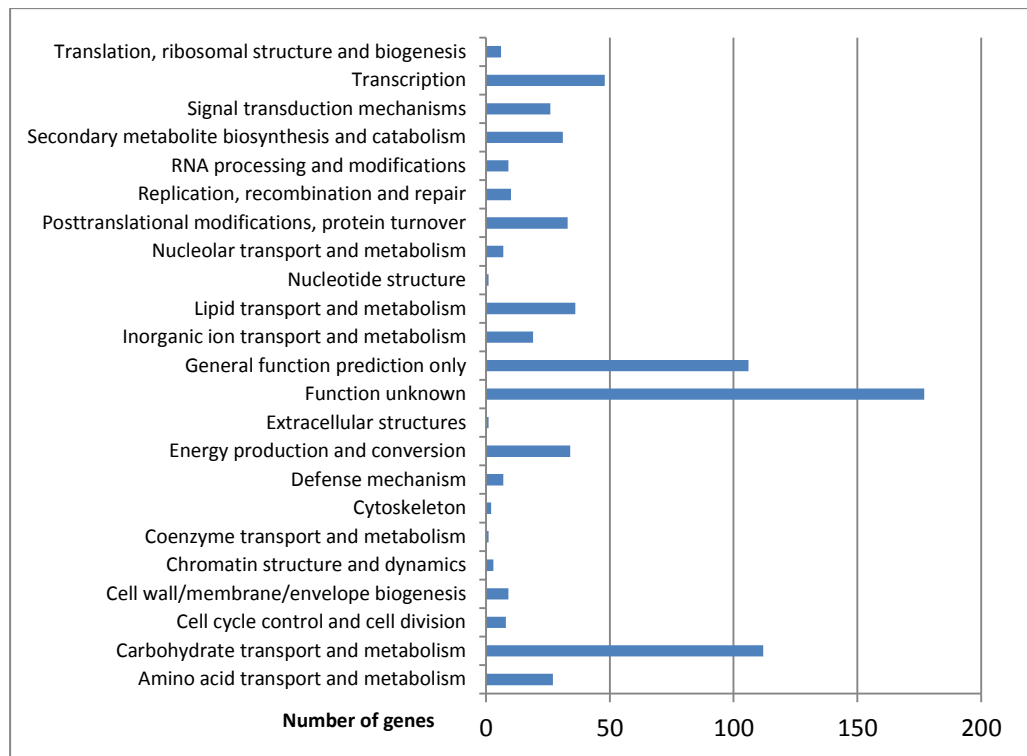
Figure A.2.F.2: Classification of genes which were significantly differently expressed between 24 h straw and 5 h glucose into their individual KEGG groups. To find all significantly differently expressed genes, three statistical significance tests were applied to changes in gene expression: the Likelihood Ratio Test (LRT; [Marioni *et al*., 2008]), Fisher's Exact Test (FET; [Bloom *et al*., 2009]) and an MA-plot-based method with Random Sampling model (MARS; [Wang *et al*., 2010]). All significantly differently expressed genes had a p-value of <0.001 for all three statistical tests.
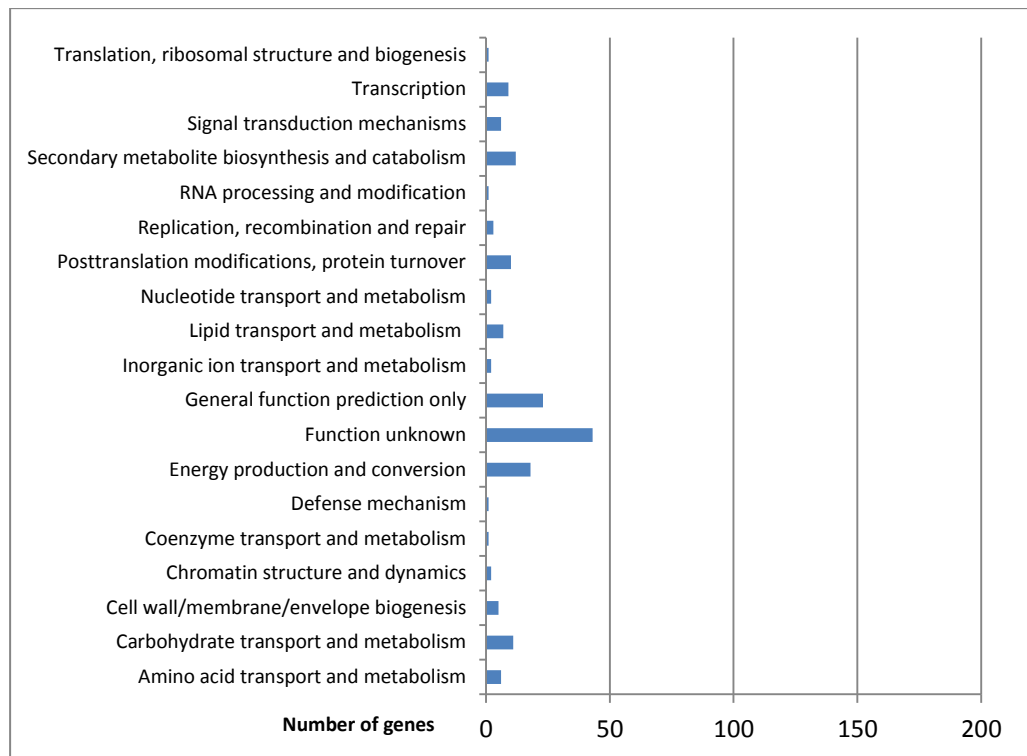
Figure A.2.F.3: Classification of genes which were significantly differently expressed between 48 h glucose and 5 h glucose into their individual KEGG groups. To find all significantly differently expressed genes, three statistical significance tests were applied to changes in gene expression: the Likelihood Ratio Test (LRT; [Marioni *et al.*, 2008]), Fisher's Exact Test (FET; [Bloom *et al.*, 2009]) and an MA-plot-based method with Random Sampling model (MARS; [Wang *et al.*, 2010]). All significantly differently expressed genes had a p-value of <0.001 for all three statistical tests.

Appendix A3

Table A.3.T.1: Expression values (RPKM) of all *T. reesei* CAZy-encoding genes in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. Gene transcript I.Ds, annotations and corresponding protein CAZy family number are also indicated.

Transcript I.D.	Annotation	CAZy Family	RPKM Glucose 48 h	RPKM Straw 24 h	RPKM Glucose 5 h
107268	Candidate esterase	CE1	4.077	22.668	6.653
107850	Candidate esterase	CE1	1.532	7.818	0.525
72072	Candidate esterase	CE1	0.045	0.028	0.000
44366	Candidate esterase	CE3	0.105	17.596	0.091
41248	Candidate acetyl xylan esterase	CE3	1.017	4.983	0.499
107488	Candidate esterase	CE3	0.000	0.000	0.000
31227	Candidate esterase	CE3	1.542	5.843	0.640
70021	Candidate acetyl xylan esterase	CE3	0.000	0.035	0.015
67678	Candidate chitin deacetylase, CBM18	CE4	0.031	0.000	0.000
69490	Candidate chitin deacetylase, CBM18	CE4	0.013	0.000	0.000
65215	Candidate imidase	CE4	0.373	7.714	0.124
105072	Candidate esterase	CE4	0.872	2.815	0.876
73632	Acetyl xylan esterase with CBM1, *axe1*	CE5	0.433	369.785	0.229
54219	Acetyl xylan esterase	CE5	0.022	114.922	0.062
60489	Cutinase	CE5	0.045	23.243	0.000
44214	Acetyl xylan esterase, *axe2*	CE5	6.031	15.890	6.924
79671	N-acetyl-glucosamine-6-phosphate deacetylase	CE9	2.022	14.718	0.752
3101	Candidate N-acetylglucosaminylphosphatidylinositol de-N-acetylase	CE14	1.968	2.058	1.578
58550	Candidate N-acetylglucosaminylphosphatidylinositol de-N-acetylase	CE14	1.732	2.595	0.566
123940	Glucuronoyl esterase, *cip2*	CE15	0.030	48.634	0.058
103825	Candidate acetyl esterase	CE16	0.202	3.703	0.058
121418	Acetyl esterase, *aes1*	CE16	0.148	188.581	0.107
120749	β-glucosidase CEL1A, *bgl2*	GH1	0.862	258.303	3.931
22197	candidate β-glucosidase CEL1B	GH1	0.620	51.848	0.592
120229	Endo-β-1,4-xylanase, *xyn3*	GH10	0.000	2.084	0.000
123818	Endo-β-1,4-xylanase, *xyn2*	GH11	1.793	2092.381	1.440
74223	Endo-β-1,4-xylanase, *xyn1*	GH11	0.174	278.194	0.028
112392	Endo-β-1,4-xylanase, *xyn5*	GH11	0.021	36.221	0.030
123232	Endoglucanase with expansin-like activity CEL12A, *egl3*	GH12	0.000	13.136	0.049
77284	Candidate endo-β-1,4-glucanase	GH12	13.636	8.351	6.575

123368	Candidate 1,4-α-glucan branching enzyme	GH13	8.466	9.038	6.106
105956	Candidate α-amylase	GH13	2.243	5.055	5.137
59578	Candidate α-glucosidase	GH13	0.054	1.168	0.011
108477	Candidate α-glucosidase	GH13	0.661	12.531	2.338
57128	Candidate glycogen debranching enzyme	GH13	13.578	10.365	3.638
65333	Candidate α-glycosidase	GH15	21.242	20.956	1.142
1885	Glucoamylase, CBM20, *gla*	GH15	1.074	16.215	3.723
76266	Glycoside hydrolase associated to cellulase signal transduction, CBM18	GH16	10.816	24.317	2.842
58239	Candidate glucanosyltransferase	GH16	9.050	18.798	2.544
122511	β-1,3/1,4-glucanase	GH16	7.359	14.275	18.952
123726	Candidate glucan endo-1,3(4)-β-D-glucosidase	GH16	0.000	11.616	0.098
39755	Candidate glucan endo-1,3(4)-β-D-glucosidase	GH16	16.065	9.499	2.596
121294	β-1,3-1,4-glucanase	GH16	14.680	6.626	4.120
38536	Candidate glucan endo-1,3(4)-β-D-glucosidase	GH16	0.111	0.386	0.000
71399	β-glycosidase	GH16	7.339	5.632	2.665
65406	Candidate glucanosyltransferase	GH16	2.842	5.482	5.317
66843	Candidate cell wall glucanosyltransferase	GH16	2.090	2.199	1.750
50215	Candidate β-glycosidase	GH16	3.599	1.529	11.029
49274	β-1,3/4-glucanase	GH16	0.410	0.676	0.835
70542	Candidate glucan endo-1,3(4)-β-D-glucosidase	GH16	0.676	0.642	0.168
73101	Candidate glucan endo-1,3(4)-β-D-glucosidase	GH16	0.078	0.290	0.028
41768	Candidate cell wall glucanosyltransferase	GH16	0.000	0.096	0.000
55886	β-glycosidase (GPI anchor)	GH16	0.000	0.000	0.000
49193	Candidate glucan 1,3-β-glucosidase	GH16	2.016	5.021	4.120
66792	Candidate glucan endo-1,3-β-glucosidase	GH16	3.980	12.960	8.558
76700	Candidate glucan 1,3-β-glucosidase	GH17	5.603	9.473	2.721
39942	β-1,3-endoglucanase	GH17	10.907	3.222	9.053
81598	Candidate chitinase, *chi18-7*	GH18	44.952	24.150	11.896
80833	Candidate chitinase, *chi46*	GH18	0.246	11.512	0.066
65162	Endo-N-acetyl-β-D-glucosaminidase, *endoT*	GH18	0.000	5.685	0.116
62645	Candidate chitinase, *chi18-4*	GH18	1.373	2.719	0.186
2735	Candidate chitinase, *chi18-6*	GH18	0.619	1.945	0.447
121355	Candidate endo-N-acetyl-β-D-glucosaminidase, *chi18-rel2*	GH18	0.742	1.859	0.340
66041	Candidate chitinase, *chi18-18*	GH18	0.957	1.668	2.579
62704	Candidate chitinase, *chi18-3*	GH18	0.556	1.532	0.205
43873	Candidate chitinase, *chi18-12*	GH18	0.000	1.213	0.034
59082	Candidate chitinase, *chi18-2*	GH18	4.287	1.153	3.188
59791	Candidate chitinase, *chi18-15*	GH18	0.000	0.678	0.000
72339	Candidate chitinase, CBM18, *chi19-9*	GH18	0.085	0.587	0.074
119859	Candidate chitinase, *chi18-13*	GH18	0.093	0.387	0.075
124043	Candidate chitinase, CBM1, *chi18-14*	GH18	0.000	0.043	0.000

56448	Candidate chitinase, *chi18-11*	GH18	0.012	0.034	0.000
68347	Candidate chitinase, CBM1, *chi18-16*	GH18	0.000	0.021	0.000
110317	Candidate chitinase, CBM1, *chi18-17*	GH18	0.000	0.011	0.000
56894	Candidate chitinase, CBM18, *chi18-10*	GH18	0.000	0.005	0.000
53949	Candidate chitinase, CBM18, *chi18-1*	GH18	0.000	0.000	0.006
108346	Candidate chitinase, CBM18, *chi18-8*	GH18	0.000	0.000	0.010
62166	Candidate β-mannosidase	GH2	1.044	54.345	1.601
76852	Candidate β-galcosidase	GH2	0.007	7.548	0.000
69245	β-mannosidase	GH2	0.010	6.928	0.021
77299	Exo-β-D-glucosaminidase	GH2	0.019	3.603	0.014
5836	β-mannosidase	GH2	0.020	2.928	0.030
102909	Candidate GH2 protein	GH2	0.531	2.532	0.511
57857	Candidate β-mannosidase	GH2	0.506	2.051	0.194
59689	Predicted β-mannosidase	GH2	0.000	0.000	0.000
23346	Candidate exochitinase	GH20	36.860	15.223	11.458
21725	N-acetyl-β-D-glucosaminidase, *nag1*	GH20	0.204	4.120	0.534
105931	β-N-acetylhexosaminidase	GH20	0.000	0.025	0.020
109278	Candidate lysozyme	GH24	0.000	0.000	0.000
103458	Diacetyl-muraminidase	GH25	0.000	1.056	0.000
72632	α-galactosidase, *agl1*	GH27	0.042	18.150	0.036
55999	Candidate α-galactosidase	GH27	0.037	6.024	0.148
59391	Candidate α-galactosidase	GH27	0.305	1.632	0.011
72704	α-galactosidase, *agl3*	GH27	0.021	1.337	0.041
65986	Candidate α-galactosidase	GH27	0.009	0.156	0.000
27259	Candidate α-galactosidase	GH27	0.000	0.039	0.015
75015	Candidate α-galactosidase	GH27	0.000	0.036	0.000
27219	Candidate α-galactosidase	GH27	0.091	0.064	0.000
103049	Candidate polygalacturonase	GH28	0.183	130.553	0.105
122780	Polygalacturonase, *rgx1*	GH28	0.460	70.909	0.823
112140	Exo-polygalacturonase, *pgx1*	GH28	0.000	2.433	0.000
70186	Polygalacturonase	GH28	0.011	0.011	0.000
121127	β-xylosidase, *bxl1*	GH3	0.471	644.458	0.870
76672	β-glucosidase CEL3A, *bgl1*	GH3	0.022	56.036	0.096
79669	Candidate β-N-acetylglucosaminidase	GH3	13.227	38.569	2.301
121735	β-glucosidase CEL3B	GH3	0.162	20.259	1.061
108671	β-glycosidase, *bgl3f*	GH3	0.025	16.033	0.022
58450	Candidate β-xylosidase, *xyl3b*	GH3	0.012	11.046	0.017
76227	Candidate β-glucosidase, CEL3E	GH3	0.543	11.037	0.253
104797	β-glucosidase, *bgl3j*	GH3	0.195	6.878	0.145
47268	β-glucosidase, *bgl3i*	GH3	0.414	6.257	0.106
82227	β-glucosidase CEL3C	GH3	0.094	5.293	0.072
46816	β-glucosidase CEL3D	GH3	0.050	1.819	0.024
69557	Candidate β-N-acetylglucosaminidase	GH3	1.378	1.164	0.150

66832	Candidate β-glucosidase	GH3	0.856	0.834	0.555
111849	Endo-β-1,4-xylanase, *xyn4*	GH30	0.654	663.939	0.423
69276	β-glycoside hydrolase	GH30	0.010	14.744	0.029
3094	Glucuronoxylanase	GH30	1.253	0.114	0.054
69736	β-1,6-glucanase	GH30	0.039	0.054	0.014
110894	Candidate endo-β-1,6-galactanase	GH30	0.328	32.192	0.083
69944	Candidate α-xylosidase/glucosidase	GH31	0.145	184.109	0.070
121351	Glucosidase II α-subunit, *gls2*	GH31	13.090	14.314	3.422
82235	Candidate α-glucosidase	GH31	0.117	10.351	2.718
60085	Candidate α-glucosidase	GH31	0.000	2.183	0.039
80240	β-galactosidase, *bga1*	GH35	0.339	49.342	0.065
64827	candidate raffinose synthase domain protein	GH36	0.419	1.092	0.097
124016	α-galactosidase, *agl2*	GH36	0.426	46.762	0.233
120676	Candidate α,α-trehalase	GH37	7.304	7.761	1.953
123226	Candidate α,α-trehalase	GH37	0.061	4.282	0.097
3196	Candidate α-mannosidase	GH38	1.433	4.777	0.310
73102	Candidate β-xylosidase	GH39	0.157	0.304	0.015
68064	Candidate β-xylosidase/arabinosidase	GH43	4.711	5.205	0.892
3739	Candidate β-xylosidase/arabinosidase	GH43	0.000	3.907	0.000
49976	Endo-β-1,4-glucanase, *egl5*	GH45	1.475	20.711	0.234
79044	Candidate α-1,2 mannosidase	GH47	15.543	24.826	1.690
65380	Candidate α-1,2 mannosidase	GH47	4.838	19.642	3.155
22252	Candidate α-1,2 mannosidase	GH47	3.712	9.510	1.436
79960	Candidate α-1,2 mannosidase	GH47	0.033	6.344	0.305
2662	Candidate α-1,2 mannosidase	GH47	2.780	6.273	1.243
45717	α-1,2 mannosidase, *mds1*	GH47	0.774	6.201	0.360
64285	Candidate α-mannosidase	GH47	2.030	2.924	0.748
111953	Candidate α-1,2 mannosidase	GH47	0.256	0.508	0.202
120312	Endoglucanase CEL5A, CBM1	GH5	0.197	252.026	0.099
56996	Mannan endo-1,4-b-mannosidase, CBM1	GH5	0.050	60.116	0.054
82616	Endoglucanase CEL5B	GH5	9.877	7.782	7.518
77506	Candidate β-glycosidase CEL5D	GH5	2.153	3.809	0.463
53731	Endo-β-1,4-glucanase	GH5	11.370	9.265	3.338
71554	Candidate β-1,3-mannanase/endo-β-1,4-mannosidase	GH5	0.000	0.018	0.000
123283	α-L-arabinofuranosidase, CBM42, *abf1*	GH54	0.273	52.499	0.131
55319	Candidate α-L-arabinofuranosidase, CBM 42, *abf3*	GH54	0.154	8.395	0.144
121746	Exo-ß-1,3-glucosidase	GH55	0.270	6.743	0.226
108776	Candidate β-1,3-glucanase	GH55	0.024	0.045	0.018
70845	β-1,3-glucanase	GH55	0.315	3.969	0.874
56418	Candidate β-1,3-glucanase	GH55	0.006	1.079	0.035
73248	Exo-1,3-β-glucanase	GH55	0.139	0.733	0.033
54242	Candidate β-1,3-glucanase	GH55	0.193	0.158	0.410

72567	Cellobiohydrolase CEL6A, CBM1, *cbh2*	GH6	0.322	462.383	0.372
73643	Copper-dependent polysaccharide monooxygenase CEL61A, CBM1	GH61	0.396	372.057	0.371
120961	Copper-dependent polysaccharide monooxygenase CEL61B	GH61	0.257	260.890	0.156
22129	Copper-dependent polysaccharide monooxygenase CEL61	GH61	0.504	2.471	0.727
31447	Candidate copper-dependent polysaccharide monooxygenase	GH61	0.000	0.000	0.000
76065	Candidate copper-dependent polysaccharide monooxygenase	GH61	0.371	0.418	0.707
27554	Candidate copper-dependent polysaccharide monooxygenase	GH61	0.099	0.066	0.000
76210	α-L-arabinofuranosidase, *abf2*	GH62	0.200	126.776	0.021
75036	Candidate α-glucosidase	GH63	25.585	16.778	7.328
22072	Candidate α-glucosidase	GH63	10.086	11.027	1.810
124175	Candidate β-1,3-glucanase	GH64	0.020	9.946	0.250
123639	β-1,3-glucanase	GH64	0.010	0.098	0.000
65137	β-1,3-glucanase	GH64	0.011	0.050	0.000
123456	Candidate α,α-trehalase	GH65	0.183	10.177	0.724
25224	Candidate α,α-trehalase	GH65	0.064	0.673	0.021
72526	α-glucuronidase, *glr1*	GH67	0.473	560.700	0.281
123989	Cellobiohydrolase CEL7A, CBM1, *cbh1*	GH7	0.865	1271.183	0.942
122081	Endoglucanase CEL7B, CBM1, *egl1*	GH7	0.019	47.720	0.054
71532	Candidate α-1,3-glucanase	GH71	0.094	8.958	1.172
73179	Candidate α-1,3-glucanase	GH71	0.172	0.995	0.047
120873	Candidate α-1,3-glucanase, CBM24	GH71	0.000	0.954	0.057
108672	Candidate α-1,3-glucanase, CBM24	GH71	0.054	0.043	0.044
123538	Candidate beta-1,3-glucanosyltransferase	GH72	0.330	6.298	2.825
22914	Candidate beta-1,3-glucanosyltransferase, CBM43	GH72	11.062	22.060	12.148
82633	Candidate beta-1,3-glucanosyltransferase	GH72	110.070	40.366	22.662
77942	Candidate beta-1,3-glucanosyltransferase	GH72	0.367	2.525	1.129
78713	Candidate beta-1,3-glucanosyltransferase	GH72	0.329	2.361	0.213
49081	Xyloglucanase, CBM1, CEL74A	GH74	0.679	120.710	0.262
70341	Chitosanase	GH75	0.000	2.251	0.000
42152	Chitosanase	GH75	0.000	0.132	0.058
66789	Chitosanase	GH75	0.000	0.102	0.000
53542	Candidate α-1,6-mannase	GH76	4.967	12.630	1.793
49409	Candidate α-1,6-mannase	GH76	5.952	10.489	2.230
55802	Candidate α-1,6-mannase	GH76	0.359	6.647	0.364
69123	Candidate α-1,6-mannase	GH76	1.512	5.498	1.199
27395	Candidate α-1,6-mannase	GH76	0.100	4.523	0.013
74807	Candidate α-1,6-mannase	GH76	0.185	3.077	0.074
67844	Candidate α-1,6-mannase	GH76	1.162	2.916	1.077
122495	Candidate α-1,6-mannase	GH76	0.000	0.072	0.000
58887	Candidate α-L-rhamnosidase	GH78	0.000	0.012	0.010
106575	β-glycosidase	GH79	0.030	0.504	0.029

71394	Candidate β-glycosidase	GH79	0.038	0.035	0.027
73005	Candidate β-glucuronidase	GH79	0.009	0.000	0.013
72568	Candidate β-glucuronidase	GH79	0.000	0.037	0.043
79602	Candidate endo-1,3-β-glucanase	GH81	0.133	0.420	0.008
73256	Candidate endo-1,3-β-glucanase	GH81	0.005	0.040	0.008
69700	α-N-acetylglucosaminidase	GH89	0.085	0.221	0.026
58117	α-N-acetylglucosaminidase	GH89	0.396	0.196	0.232
79921	Candidate α-1,2-mannosidase	GH92	0.116	15.799	0.476
111733	Candidate α-1,2-mannosidase	GH92	0.000	0.005	0.000
57098	Candidate alpha-1,2-mannosidase	GH92	1.550	6.621	0.720
74198	Candidate α-1,2-mannosidase	GH92	0.069	40.312	0.017
55733	Candidate α-1,2-mannosidase	GH92	0.169	3.254	0.177
60635	Candidate α-1,2-mannosidase	GH92	0.006	0.537	0.049
69493	Candidate α-1,2-mannosidase	GH92	0.006	0.072	0.009
5807	Candidate α-L-fucosidase	GH95	0.037	2.320	0.009
58802	Candidate α-L-fucosidase	GH95	0.000	1.201	0.042
72488	Candidate α-L-fucosidase	GH95	0.227	0.304	0.117
111138	Candidate α-L-fucosidase	GH95	0.039	5.736	0.028
4221	Candidate rhamnogalacturonyl hydrolase	GH105	0.048	0.011	0.000
57179	Candidate rhamnogalacturonyl hydrolase	GH105	0.000	0.359	0.000
79606	Candidate xylan-α-1,2-glucuronidase	GH115	0.427	11.393	0.136
108348	Unknown GH	Unknown	0.000	0.007	0.000
105288	Unknown GH	Unknown	0.034	0.063	0.016
121136	Unknown GH	Unknown	0.053	0.000	0.000
110259	Candidate alginate lyase	PL7	0.000	0.444	0.000
103033	Candidate alginate lyase	PL7	0.037	0.035	0.000
111245	Candidate chondroitin lyase	PL8	0.047	0.718	0.009
53186	Glucuronan hydrolase	PL20	0.000	0.017	0.000
69189	Candidate endo-β-1,4-glucuronan lyase	PL20	0.000	0.050	0.000

Table A.3.T.2: Expression values (RPKM) in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose of all *T. reesei* CAZy-encoding genes which were significantly differently expressed between the three conditions. Gene transcript I.Ds, annotations, the presence of a secretion signal and corresponding protein CAZy family number are also indicated.

Transcript I.D.	Annotation	CAZy family	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose	Secretion signal
12749	β-glucosidase, CEL1A, *bgl2*	GH1	0.862	258.303	3.931	no
22197	β-glucosidase, CEL1B	GH1	0.620	51.848	0.592	no
76852	GusB β-glucuronidase	GH2	0.007	7.548	0.000	yes
69245	Predicted β-mannosidase	GH2	0.010	6.928	0.021	yes
5836	Predicted β-mannosidase	GH2	0.020	2.928	0.030	no
62166	Predicted β-mannosidase	GH2	1.044	54.345	1.601	no
77299	Predicted β-mannosidase	GH2	0.019	3.603	0.014	yes
121127	B-xylosidase, *bxl1*	GH3	0.471	644.458	0.870	yes
58450	Potential β-xylosidase, *xyl3b*	GH3	0.012	11.046	0.017	yes
108671	β-glucosidase, *bgl3f*	GH3	0.025	16.033	0.022	yes
104797	β-glucosidase, *bgl3j*	GH3	0.195	6.878	0.145	yes
46816	β-glucosidase, CEL3D	GH3	0.050	1.819	0.024	no
76227	β-glucosidase, CEL3E	GH3	0.543	11.037	0.253	yes
76672	B-glucosidase, *bgl1*	GH3	0.022	56.036	0.096	yes
121735	β-glucosidase, CEL3B	GH3	0.162	20.259	1.061	yes
82227	Candidate β-glucosidase, CEL3C	GH3	0.094	5.293	0.072	yes
47268	β-glucosidase, *bgl3i*	GH3	0.414	6.257	0.106	no
79669	Candidate β-N-acetylglucosaminidase	GH3	13.227	38.569	2.301	no
56996	Endo-β-1,4-mannosidase, CBM1	GH5	0.050	60.116	0.054	yes
120312	Endoglucanase, CBM1, *egl2*	GH5	0.197	252.026	0.099	yes
72567	Cellobiohydrolase, CBM1, *cbh2*	GH6	0.322	462.383	0.372	yes
122081	Endoglucanase, CBM1, *egl1*	GH7	0.019	47.720	0.054	yes
123989	Cellobiohydrolase, CBM1, *cbh1*	GH7	0.865	1271.183	0.942	yes
74223	Xylanase, *xyn1*	GH11	0.174	278.194	0.028	yes
123818	Xylanase, *xyn2*	GH11	1.793	2092.381	1.440	yes
112392	Xylanase, *xyn5*	GH11	0.021	36.221	0.030	no
123232	Endo β-1,4 glucanase, *egl3*	GH12	0.000	13.136	0.049	yes
59578	α-glucosidase	GH13	0.054	1.168	0.011	no
65333	Putative α-glycosidase	GH15	21.242	20.956	1.142	no
123726	Candidate β-1,4,1,3-glucanase	GH16	0.000	11.616	0.098	yes
80833	Chitinase, *chi46*	GH18	0.246	11.512	0.066	yes
65162	Endo-N-acetyl-β-D-glucosaminidase, *endoT*	GH18	0.000	5.685	0.116	no
62645	Chitinase, *chi18-4*	GH18	1.373	2.719	0.186	yes
21725	β-N-acetylhexosaminidase	GH20	0.204	4.120	0.534	yes

72632	α-D-galactosidase, *agl1*	GH27	0.042	18.150	0.036	yes
55999	α-D-galactosidase	GH27	0.037	6.024	0.148	yes
72704	α-D-galactosidase, *agl3*	GH27	0.021	1.337	0.041	yes
59391	α-D-galactosidase	GH27	0.305	1.632	0.011	yes
103049	Potential polygalacturonase	GH28	0.183	130.553	0.105	yes
122780	Polygalacturonase, *rgx1*	GH28	0.460	70.909	0.823	yes
110894	Endo-β-1,6-galactanase	GH30	0.328	32.192	0.083	yes
69276	Candidate endo-β-1,4-xylanase	GH30	0.010	14.744	0.029	yes
111849	Endo-β-1,4-xylanase, *xyn4*	GH30	0.654	663.939	0.423	yes
69944	Candidate α-xylosidase	GH31	0.145	184.109	0.070	no
82235	Candidate α-glucosidase	GH31	0.117	10.351	2.718	yes
80240	β-galactosidase	GH35	0.339	49.342	0.065	yes
124016	α-galactosidase, *agl2*	GH36	0.426	46.762	0.233	yes
123226	Candidate α,α-trehalase	GH37	0.061	4.282	0.097	yes
3196	α-mannosidase	GH38	1.433	4.777	0.310	no
73102	Candidate β-xylosidase	GH39	0.157	0.304	0.015	yes
49976	Endoglucanase, CBM1, *egl5*	GH45	1.475	20.711	0.234	yes
79960	α-1,2-mannosidase	GH47	0.033	6.344	0.305	yes
45717	1,2-α-mannosidase	GH47	0.774	6.201	0.360	yes
79044	α-1,2-mannosidase	GH47	15.543	24.826	1.690	yes
55319	α-L-arabinofuranosidase, CBM42, *abf3*	GH54	0.154	8.395	0.144	yes
123283	α-L-arabinofuranosidase, CBM42, *abf1*	GH54	0.273	52.499	0.131	yes
56418	Candidate β-1,3-glucanase	GH55	0.006	1.079	0.035	yes
121746	Exo-1,3-β-glucanase	GH55	0.270	6.743	0.226	yes
73248	Candidate exo-1,3-β glucanase	GH55	0.139	0.733	0.033	yes
27554	Copper-dependent monooxygenase GH61	GH61	0.099	0.066	0.000	no
73643	Copper-dependent monooxygenase CEL61a, CBM1	GH61	0.396	372.057	0.371	yes
120961	Copper-dependent monooxygenase GEL61b	GH61	0.257	260.890	0.156	yes
76210	α-L-arabinofuranosidase, *abf2*	GH62	0.200	126.776	0.021	yes
124175	β-1,3-glucanase	GH64	0.020	9.946	0.250	no
123456	Acid trehalase	GH65	0.183	10.177	0.724	yes
25224	Acid trehalase	GH65	0.064	0.673	0.021	no
72526	α-glucuronidase, *glr1*	GH67	0.473	560.700	0.281	yes
73179	Endo-1,3-α glucosidase	GH71	0.172	0.995	0.047	yes
120873	Endo-1,3-α glucosidase, CBM24	GH71	0.000	0.954	0.057	yes
27395	α-1,6-mannanase	GH76	0.100	4.523	0.013	no
55802	α-1,6-mannanase	GH76	0.359	6.647	0.364	no
106575	Candidate β-glucuronidase	GH79	0.030	0.504	0.029	yes
79602	Predicted endo-1,3-β glucanase	GH81	0.133	0.420	0.008	no
79921	Putative α-1,2-mannosidase	GH92	0.116	15.799	0.476	yes
60635	α-1,2-mannosidase	GH92	0.006	0.537	0.049	yes

74198	Putative α-1,2-mannosidase	GH92	0.069	40.312	0.017	yes
55733	Putative α-1,2-mannosidase	GH92	0.169	3.254	0.177	yes
5807	Candidate α-L-fucosidase	GH95	0.037	2.320	0.009	yes
58802	Candidate α-L-fucosidase	GH95	0.000	1.201	0.042	yes
44366	Candidate esterase	CE3	0.105	17.596	0.091	yes
65215	Candidate imidase	CE4	0.373	7.714	0.124	no
73632	Acetyl xylan esterase, CBM1, *axe1*	CE5	0.433	369.785	0.229	yes
54219	Acetyl xylan esterase	CE5	0.022	114.922	0.062	no
60489	Cutinase	CE5	0.045	23.243	0.000	no
123940	Glucuronoyl esterase	CE15	0.030	48.634	0.058	yes
121418	Acetyl esterase	CE16	0.148	188.581	0.107	yes

Appendix A4

Table A.4.T.1: Expression values (RPKM) in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose of all *T. reesei* non-CAZy-encoding genes with an over 20-fold expression between 48 h glucose and 24 h straw and an RPKM value higher than 50 in 24 h straw. Gene transcript I.Ds and annotations are also indicated.

Transcript I.D.	Annotation	RPKM Glucose 48 h	RPKM Straw 24 h	RPKM Glucose 5 h
Lipases				
121418	Lipase, active site G-D-S-L	0.15	188.58	0.11
64397	Ceramidase family protein, associated to cellulase signal transduction	1.34	52.95	0.42
Surface interacting proteins				
119989	HFB2, class II hydrophobin	0.44	275.21	96.17
74282	QI74 orthologue, cell wall protein	0.18	169.54	5.77
123967	HFB3, class II hydrophobin	0.31	99.71	13.19
120784	Cell wall mannoprotein	0.79	59.26	0.73
Enzymes of carbon and nitrogen metabolism				
107776	Xylose reductase, XYL1	1.55	161.02	1.61
123009	Glutamine synthetase	4.44	117.02	72.37
Transporters				
3405	MSF (major facilitator superfamily) permease	0.63	906.19	1.18
50894	MSF permease	0.21	251.43	0.28
104072	Xylose transporter	0.15	182.60	0.12
69957	Maltose permease	0.13	174.58	0.19
82309	Predicted transporter (MFS)	0.26	87.44	0.07
120017	Oligopeptide transporter	1.02	77.15	3.29
106330	Major facilitator superfamily	0.29	50.18	0.26
Gene regulation				
44747	SNF2 family helicase	3.45	124.96	9.33
80980	Peptidyl arginine deiminase (PAD)	0.05	65.21	4.06
108357	C_2H_2 transcriptional regulator	0.34	59.43	0.28
Oxidation-related				
80659	Alcohol oxidase AOX1	0.53	890.78	0.39
56840	GFO_IDH_MocA dehydrogenase	1.06	219.64	0.59
76696	Flavin-containing monooxygenase	0.37	60.77	0.19
123827	Bifunctional catalase/peroxidase	2.33	58.01	2.73
Unknown				
55272	Unknown protein with high similarity to a C6 transcription factor from *Metarhizium acridum*	4.97	408.73	23.80

120031	Unknown protein	1.05	126.10	10.35
42752	Unique protein, some similarity to GT51 family protein from *Exiguobacterium sibiricum*	1.56	101.28	10.31
107715	Unknown protein	2.76	66.26	3.07
74580	Unknown protein with high similarity to a Regulator of G-protein signalling from *Cordyceps militaris*	0.41	51.33	4.73

Appendix A5

A.5.T.1: Expression values (RPKM) of all GH61 protein-encoding genes in *T. reesei* and *A. niger* in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. Gene names used in this study, gene transcript I.Ds and fold-inductions between 48 h glucose and 24 h straw are also indicated.

Name used in this study	Transcript I.D.	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose	Fold-induction
T. reesei					
gh61	22129	6.703	13.677	3.868	2.040
gh61a	73643	1.475	2.493	1.119	1.690
gh61b	120961	0.444	1.030	0.273	2.318
not applicable	27554	0.099	0.066	0.000	repressed
not applicable	76065	0.371	0.417	0.706	1.125
not applicable	31447	0.000	0.000	0.000	no change
A. niger					
An12g04610	211595	3.287	3389.868	54.290	1031.308
An08g05230	52688	1.608	4.277	1.946	2.659
An15g05470	210233	3.454	3.207	1.337	repressed
An12g02540	43784	0.000	2.221	0.130	up-regulated
An14g02670	53797	0.598	1.975	0.662	3.302
An15g04900	182430	0.064	1.193	0.496	18.767
An04g08550	194765	0.062	0.000	0.024	repressed

A.5.T.2: Expression values (RPKM) of all hydrophobin, swollenin, expansin and hydrophobic surface binding protein (HsbA)-encoding genes in *T. reesei* and *A. niger* in 48 h glucose, 24 h straw and 24 h + 5 h glucose. Gene transcript I.Ds, annotations and fold-inductions between 48 h glucose and 24 h straw are also indicated.

T. reesei					
Transcript I.D.	Annotation	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose	Fold-induction
106538	HFB5, putative classII hydrophobin, related to Hfb2	0.000	0.686	0.000	up-regulated
119989	HFB2, class II	0.439	275.208	96.170	627.015
123967	HFB3, class II	0.305	99.707	13.187	326.735
104293	Putative class II hydrophobin, similar to HFBI	0.000	0.000	0.000	0.000
105869	Hydrophobin 2 with a class II C-terminal hydrophobin domain	n/a (not applicable)	n/a	n/a	n/a
73173	HFB1, class II	1.761	0.023	0.732	repressed
123992	Swollenin, swo1	2.181	315.259	0.906	144.523
111874	Swollenin 2, swo2	0.426	0.227	0.088	repressed
23115	Distantly related to plant expansins, EXPN	1.349	3.050	8.052	2.261
71390	Distantly related to plant expansins, EXPN	0.000	0.000	0.000	no change
104079	Distantly related to plant expansins, EXPN	0.000	0.000	0.000	no change
73638	CIP1, candidate cellulose binding protein, contains CBM1	0.140	176.860	0.140	1263.286
A. niger					
Transcript I.D.	Annotation	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose	Fold induction
n/a	Hyp1/HfbC	0.000	138.500	29.300	up-regulated
53462	The nucleotide sequence of the ORF shows strong similarity to two ESTs of *A. niger* (an_3228 and	0.261	118.604	0.984	454.501

	an_3303, unknown protein products).				
188224	HsbA	0.300	118.600	1.000	395.333
45685	HfbB	0.287	33.921	84.953	118.334
128530	HfbD	3.459	380.396	423.200	109.957
45683	HfbA, strong similarity to spore cell wall fungal hydrophobin *dewA* from *Aspergillus nidulans*.	n/a	n/a	n/a	n/a
n/a	HfbE, similarity to the secreted hydrophobin HFBI from *Trichoderma reesei*.	9.594	5.488	7.201	repressed
194815	Strong similarity to the small, moderately hydrophobic *rodA* from *Aspergillus nidulans*; similarity to 3 cell wall proteins from *Schizophyllum commune* which are important in the formation and function of aerial structures.	0.042	0.000	2.661	repressed

A.5.T.3: Expression values (RPKM) of all Pth11-type GPCR-encoding genes in *T. reesei* and *A. niger* in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. *T.reesei* gene transcript I.Ds, *A. niger* gene names, gene annotations and fold-inductions between 48 h glucose and 24 h straw are also indicated.

T. reesei					
Transcript I.D.	**Annotation**	**RPKM glucose 48 h**	**RPKM straw 24 h**	**RPKM glucose 5 h**	**Fold-Induction**
27983	PTH11 GPCR	0.00	0.03	0.00	up-regulated
69500	PTH11 GPCR	0.00	1.14	0.04	up-regulated
111861	PTH11 GPCR	0.00	0.21	0.09	up-regulated
70967	Candidate PTH11 GPCR	0.00	0.39	0.00	up-regulated
27992	PTH11 GPCR	0.47	12.34	0.18	26.52
67334	Putative MFS transporter, PTH11-type?	0.27	4.48	0.19	16.89
69904	PTH11 GPCR	0.41	4.39	0.10	10.73
45573	PTH11 GPCR	1.98	15.96	5.55	8.07
103694	PTH11 GPCR	0.75	5.61	0.37	7.44
40156	Candidate PTH11 GPCR	0.23	1.17	0.26	5.14
58767	Candidate PTH11 GPCR	0.29	1.04	0.33	3.64
106082	Candidate PTH11 GPCR	0.76	2.76	0.18	3.60
110744	PTH11 GPCR	1.52	4.59	0.59	3.02
41425	Candidate PTH11 GPCR	0.04	0.12	0.03	2.80
107042	PTH11 GPCR	0.03	0.08	0.00	2.80
78499	PTH11 GPCR	4.41	11.58	3.03	2.63
66786	Candidate PTH11 GPCR	0.25	0.57	0.13	2.29
121990	PTH11 GPCR	2.44	5.37	0.91	2.20
61354	Candidate PTH11 GPCR	0.24	0.46	0.27	1.93
124113	PTH11 GPCR	5.90	9.03	3.18	1.53
122795	PTH11 GPCR	4.40	6.38	0.57	1.45
57101	PTH11 GPCR	0.00	0.00	0.00	no change
109146	PTH11 GPCR	0.00	0.00	0.01	no change
53452	Candidate PTH11 GPCR	0.00	0.00	0.00	no change
5647	PTH11 GPCR	0.89	0.17	0.14	repressed
82041	PTH11 GPCR	0.55	0.37	0.52	repressed
62462	PTH11 GPCR	2.83	1.28	2.55	repressed
76763	PTH11 GPCR	12.03	1.60	1.65	repressed
41260	PTH11 GPCR	0.03	0.00	0.00	repressed
55561	PTH11 GPCR	4.99	3.33	4.12	repressed
105224	PTH11 GPCR	0.17	0.03	4.28	repressed
122824	PTH11 GPCR	1.85	0.25	0.48	repressed

39587	Unknown protein	1.41	0.27	0.12	repressed

A. niger					
Gene name	Annotation	RPKM Glucose 48 h	RPKM Straw 24 h	RPKM Glucose 5 h	Fold Induction
An02g00750	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.000	0.374	0.270	up-regulated
An07g00220	similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.000	0.029	0.000	up-regulated
An11g00060	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.095	26.417	4.144	278.383
An11g07030	Similarity to plasma membrane protein Pth11 -*Magnaporthe grisea*	0.017	1.522	0.042	91.429
An18g02730	Similarity to transmembrane protein PTH11 -*Magnaporthe grisea*	1.420	53.559	3.509	37.712
An11g03090	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.092	2.794	0.669	30.280
An11g06430	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.047	0.917	0.067	19.334
An05g02420	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.048	0.754	0.175	15.730
An09g03230	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.048	0.650	0.883	13.436
An11g03960	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.018	0.190	0.015	10.814
An01g01220	Similarity to surface recognition protein PTH11 - *Magnaporthe grisea*	0.015	0.118	0.025	7.865
An14g05870	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.114	0.840	1.008	7.373
An01g11690	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.261	1.891	1.540	7.244
An14g05850	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.054	0.388	0.126	7.128
An02g00470	Similarity to transmembrane protein PTH11 -*Magnaporthe grisea*	0.017	0.115	0.056	6.882
An18g00430	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.016	0.093	0.027	5.899
An16g04840	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.017	0.098	0.028	5.899
An02g00160	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.042	0.235	0.616	5.571
An16g05730	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.065	0.333	0.272	5.161
An18g01570	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.096	0.471	0.417	4.916
An03g01430	Weak similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.031	0.138	0.052	4.424
An09g03930	Similarity to integral membrane	0.016	0.062	0.000	3.932

	protein PTH11 - *Magnaporthe grisea*				
An16g01210	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.092	0.306	0.046	3.343
An18g00980	Weak similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	17.577	58.121	18.982	3.307
An15g01930	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.134	0.431	0.226	3.217
An09g06200	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.860	2.747	1.242	3.195
An04g10010	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.070	0.206	0.250	2.949
An01g11190	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.193	0.537	0.041	2.785
An15g02150	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	2.659	5.596	1.536	2.105
An06g00490	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.608	1.228	0.678	2.019
An04g00950	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.103	0.202	0.104	1.966
An03g05810	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.016	0.031	0.013	1.966
An12g02600	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea.*	0.028	0.056	0.000	1.966
An09g03180	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.021	0.041	0.018	1.966
An03g01000	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.065	0.127	0.014	1.966
An16g08520	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea.*	0.217	0.402	0.101	1.857
An04g02930	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.030	0.044	0.038	1.475
An14g03530	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.110	0.139	0.185	1.264
An08g06190	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.305	0.381	0.210	1.251
An03g00290	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.081	0.080	0.287	repressed
An02g00390	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.031	0.031	0.066	repressed
An14g05730	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	1.016	0.981	6.949	repressed
An10g00340	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	1.307	1.080	0.576	repressed
An18g01230	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.421	0.260	0.720	repressed
An17g00940	Similarity to integral membrane protein PTH11 - *Magnaporthe*	1.880	1.131	0.887	repressed

	grisea.				
An02g00200	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.269	0.153	0.477	repressed
An03g01300	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.178	0.100	0.064	repressed
An15g04130	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	5.379	2.635	1.584	repressed
An09g04830	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	2.919	1.165	2.147	repressed
An15g07470	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.255	0.100	0.021	repressed
An02g09500	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	6.563	2.571	6.730	repressed
An01g01150	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.035	0.011	0.000	repressed
An07g00570	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	1.256	0.401	0.921	repressed
An07g02370	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	1.465	0.422	1.572	repressed
An16g01270	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.401	0.109	0.058	repressed
An01g02360	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.237	0.062	0.120	repressed
An01g13320	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	2.212	0.560	3.672	repressed
An01g09240	Weak Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.795	0.195	0.521	repressed
An11g08020	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.614	0.130	0.224	repressed
An16g01320	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	5.495	0.969	0.545	repressed
An04g06980	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	2.801	0.484	1.258	repressed
An09g02320	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	19.056	1.651	3.271	repressed
An08g09720	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.660	0.056	0.157	repressed
An17g00590	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.956	0.063	0.046	repressed
An16g07200	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	1.167	0.064	0.911	repressed
An16g01170	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	8.592	0.312	0.226	repressed
An02g09530	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea* membrane	16.877	0.172	2.269	repressed
An01g01780	Similarity to integral membrane protein PTH11 - *Magnaporthe*	0.016	0.000	0.014	repressed

	grisea				
An12g10780	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.019	0.000	0.008	repressed
An18g01530	Similarity to integral membrane protein PTH11 - *Magnaporthe grisea*	0.188	0.000	0.110	repressed

Appendix A6

Table A.6.T.1: Expression values (RPKM) of all *T. reesei* genes containing NATs (natural antisense transcripts) in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. Gene transcript I.Ds, annotations and KEGG group classification are also indicated.

Transcript I.D.	Annotation	KEGG Group	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose
120176	Cysteine dioxygenase CDO1	Amino acid transport and metabolism	0.437	7.732	0.563
57015	Amino acid transporter	Amino acid transport and metabolism	1.129	0.146	0.191
75414	Anthranilate phosphoribosyltransferase	Amino acid transport and metabolism	3.124	1.799	1.535
58282	Amidohydrolase, peptidase M14, carboxypeptidase A	Amino acid transport and metabolism	0.244	12.682	0.739
106064	Zn-finger, nucleic acid binding	Amino acid transport and metabolism	0.775	3.328	0.180
77647	Threonine aldolase	Amino acid transport and metabolism	0.248	0.147	0.114
62172	Amino acid transporter	Amino acid transport and metabolism	0.000	0.378	0.000
62463	Flavoprotein/aromatic ring monooxygenase/hydroxylase	Amino acid transport and metabolism	8.871	0.192	0.047
49112	Acetylornithine aminotransferase	Amino acid transport and metabolism	0.827	0.294	0.384
65410	Phosphoadenosine phosphosulfate reductase	Amino acid transport and metabolism	8.297	1.593	3.749
51110	Amino acid permease	Amino acid transport and metabolism	0.134	22.905	3.882
67806	Amino acid transporter	Amino acid transport and metabolism	3.011	1.019	2.577
81442	Amino acid transporter	Amino acid transport and metabolism	1.370	0.310	0.551
23298	Chorismate synthase	Amino acid transport and metabolism	13.132	4.320	5.220
43961	Predicted phosphoglycerate mutase	Carbohydrate transport and metabolism	3.163	5.573	1.222
54632	Membrane transporter (major facilitator superfamily)	Carbohydrate transport and metabolism	0.252	6.673	0.276
72632	Alpha-D-galactosidase, GH27	Carbohydrate transport and metabolism	0.042	18.150	0.036
56341	Sucrose transporter and related proteins	Carbohydrate transport and metabolism	1.379	4.232	2.436
76151	Galactosyltransferases	Carbohydrate transport and metabolism	3.241	0.201	0.059
105260	Predicted sugar transporter (major facilitator superfamily)	Carbohydrate transport and metabolism	0.016	5.040	0.023
58584	Aquaporin (major intrinsic protein family)	Carbohydrate transport and metabolism	0.000	4.553	0.127
76852	Beta-Glucuronidase GUSB (GH 2), membrane bound, ATP coupled synthesis to proton transport	Carbohydrate transport and metabolism	0.007	7.548	0.000

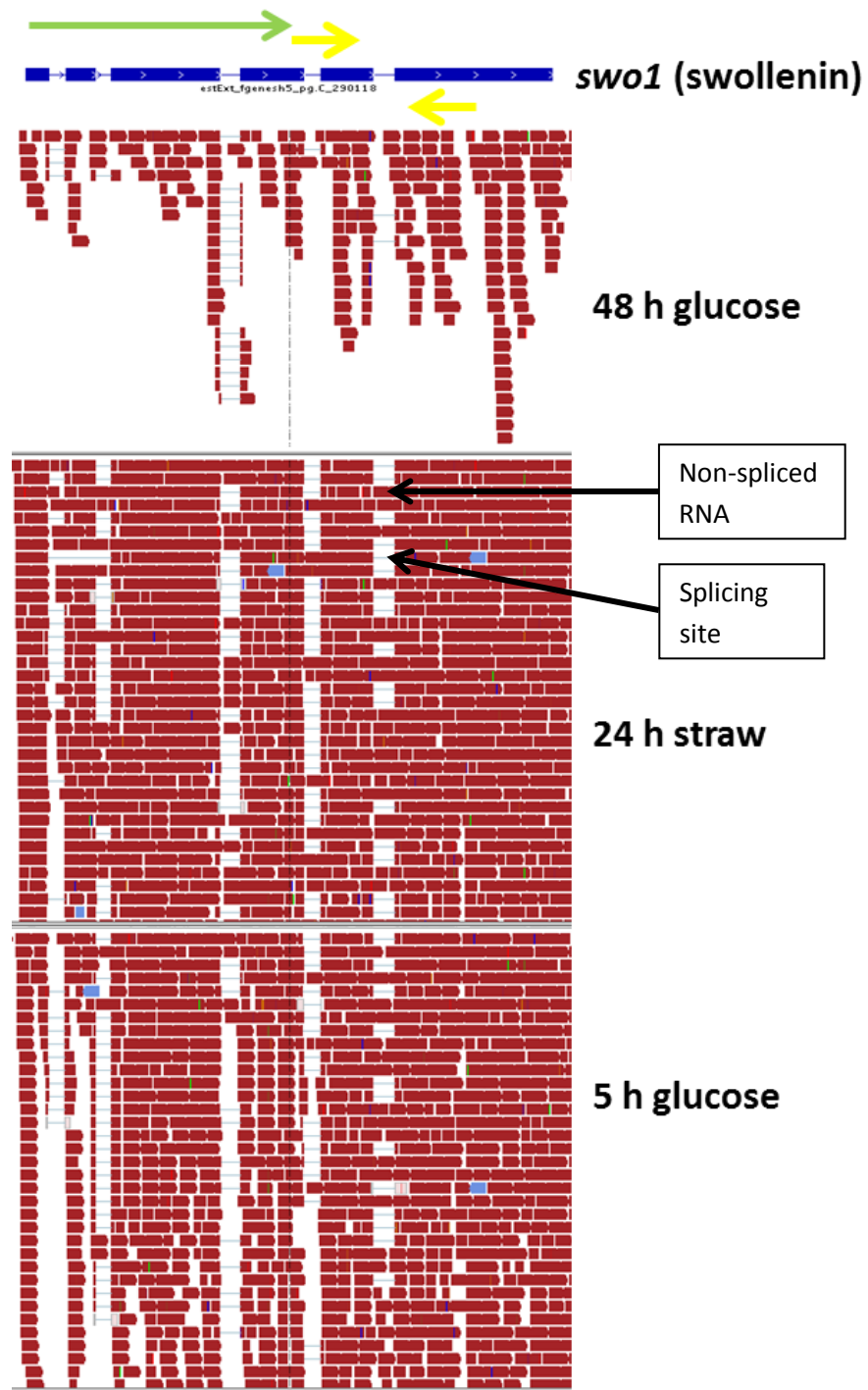|					
106330	Sugar transporter (major facilitator superfamily)	Carbohydrate transport and metabolism	0.287	50.183	0.264
60329	Permease of the major facilitator superfamily	Carbohydrate transport and metabolism	0.000	0.612	0.000
122079	Glucose/Ribitol dehydrogenase (Reductases with broad range of substrate specificities)	Carbohydrate transport and metabolism	0.090	4.478	0.114
22426	Dimeric dihydrodiol dehydrogenase	Carbohydrate transport and metabolism	1.743	2.397	0.909
71817	Synaptic vesicle transporter SVOP and related transporters (major facilitator superfamily) and polysaccharide deacetylase	Carbohydrate transport and metabolism	1.879	0.111	0.771
79960	Mannosyl-oligosaccharide alpha-1,2-mannosidase and related glycosyl hydrolases	Carbohydrate transport and metabolism	0.033	6.344	0.305
80058	Sugar transporter (major facilitator superfamily)	Carbohydrate transport and metabolism	0.014	2.040	0.000
123550	Indoleamine 2,3-dioxygenase-like protein	Carbohydrate transport and metabolism	0.066	0.788	0.047
68606	Triosephosphate isomerase	Carbohydrate transport and metabolism	1.089	2.803	0.414
81670	Sugar transporter (major facilitator superfamily)	Carbohydrate transport and metabolism	0.537	11.480	0.560
70341	Fungal chitosanase, GH75	Carbohydrate transport and metabolism	0.000	2.251	0.000
124134	Sugar/hydrogen symporter (phosphotransferase system)	Carbohydrate transport and metabolism	0.666	3.209	0.330
70542	Candidate glucan endo-1,3(4)-β-D-glucosidase, GH16	Carbohydrate transport and metabolism	0.676	0.642	0.168
68821	Candidate β-glycosyltransferase	Carbohydrate transport and metabolism	0.353	0.918	0.096
102947	Speckle-type POZ protein SPOP and related proteins with TRAF, MATH and BTB/POZ domains	Cell cycle control, cell division, chromosome partitioning	1.985	2.515	0.598
33387	GCN5-related N-acetyltransferase	Chromatin structure and dynamics	0.029	6.989	0.000
124283	Predicted histone tail methylase containing SET domain	Chromatin structure and dynamics	3.372	14.075	2.076
58837	Molybdenum cofactor biosynthesis pathway protein	Coenzyme transport and metabolism	0.673	1.420	0.273
61988	Myosin regulatory light chain, EF-Hand protein superfamily	Cytoskeleton	1.736	1.158	1.116
108401	Kinesin (SMY1 subfamily)	Cytoskeleton	0.502	2.435	0.307
49205	Haem peroxidase	Defence mechanisms	9.821	1.542	5.159
56432	Glycolate oxidase	Energy production and conversion	1.715	4.718	0.670
104606	Alcohol dehydrogenase, class IV	Energy production and conversion	0.485	10.356	0.404
77547	UDP-glucuronosyl and UDP-glucosyl transferase	Energy production and conversion	7.867	0.599	1.310
60847	Mitochondrial F1F0-ATP synthase, subunit c/ATP9/proteolipid	Energy production and conversion	178.319	10.617	44.996
47930	Mitochondrial oxoglutarate/malate carrier proteins	Energy production and conversion	0.308	36.373	0.626
66662	Enolase/mandelate racemase/muconate lactonizing enzyme	Energy production and conversion	0.011	0.616	0.000

67008	Kynurenine 3-monooxygenase and related flavoprotein monooxygenases	Energy production and conversion	2.704	0.629	0.494
111932	Aldehyde dehydrogenase	Energy production and conversion	0.037	1.678	0.242
111341	Unknown	Function unknown	0.125	0.021	0.000
122874	Unknown	Function unknown	0.139	9.336	0.456
32261	Predicted membrane protein	Function unknown	13.786	3.935	2.381
54260	Predicted indoleamine-2,3-dioxygenase	Function unknown	0.206	2.641	0.016
55242	Unknown	Function unknown	0.433	0.880	0.625
103660	Unknown	Function unknown	0.209	0.049	0.264
103898	Unknown	Function unknown	0.478	0.511	0.132
120248	Unknown	Function unknown	0.256	0.489	0.192
104684	Unknown	Function unknown	0.182	0.009	0.014
57558	Unknown	Function unknown	0.061	3.954	0.025
58244	Unknown	Function unknown	1.022	4.390	0.216
120929	Unknown	Function unknown	10.618	1.707	1.096
3267	Unknown	Function unknown	0.574	0.315	0.146
59152	Unknown	Function unknown	1.047	7.336	0.168
121189	Unknown	Function unknown	1.010	3.830	0.645
3488	Unknown	Function unknown	13.379	2.648	0.327
73248	Unknown	Function unknown	0.139	0.733	0.033
60482	Unknown	Function unknown	1.272	37.448	11.809
102468	Unknown	Function unknown	0.089	0.344	0.059
102892	Unknown	Function unknown	2.123	0.648	0.607
103130	Unknown	Function unknown	0.119	0.287	0.123
2100	Unknown	Function unknown	5.677	2.354	2.679
120311	Unknown	Function unknown	37.352	35.438	31.968
104277	Unknown	Function unknown	0.079	48.403	0.152
77593	Protein containing a leucine-rich repeat, ribonuclease inhibitor type	Function unknown	1.519	1.003	0.255
22257	Unknown	Function unknown	0.020	2.236	0.118
77767	Unknown	Function unknown	3.217	0.387	0.649
61504	Unknown	Function unknown	0.179	0.030	0.047
107386	Unknown	Function unknown	0.159	17.876	7.137
77998	Unknown	Function unknown	0.206	4.527	0.023
121883	Unknown	Function unknown	0.000	7.489	0.156
107667	Unknown	Function unknown	2.849	0.361	0.145
122324	Unknown	Function unknown	0.252	1.442	0.227
63733	Unknown	Function unknown	0.143	0.950	0.269
122795	Unknown	Function unknown	4.403	6.382	0.570
109287	Unknown	Function unknown	6.228	2.794	0.579
66353	Unknown	Function unknown	1.416	2.151	1.728
110173	Unknown	Function unknown	2.075	0.222	0.655
67579	Unknown	Function unknown	0.000	1.565	0.144
110877	Unknown	Function unknown	0.858	1.525	0.453
110878	Unknown	Function unknown	0.795	1.313	0.265

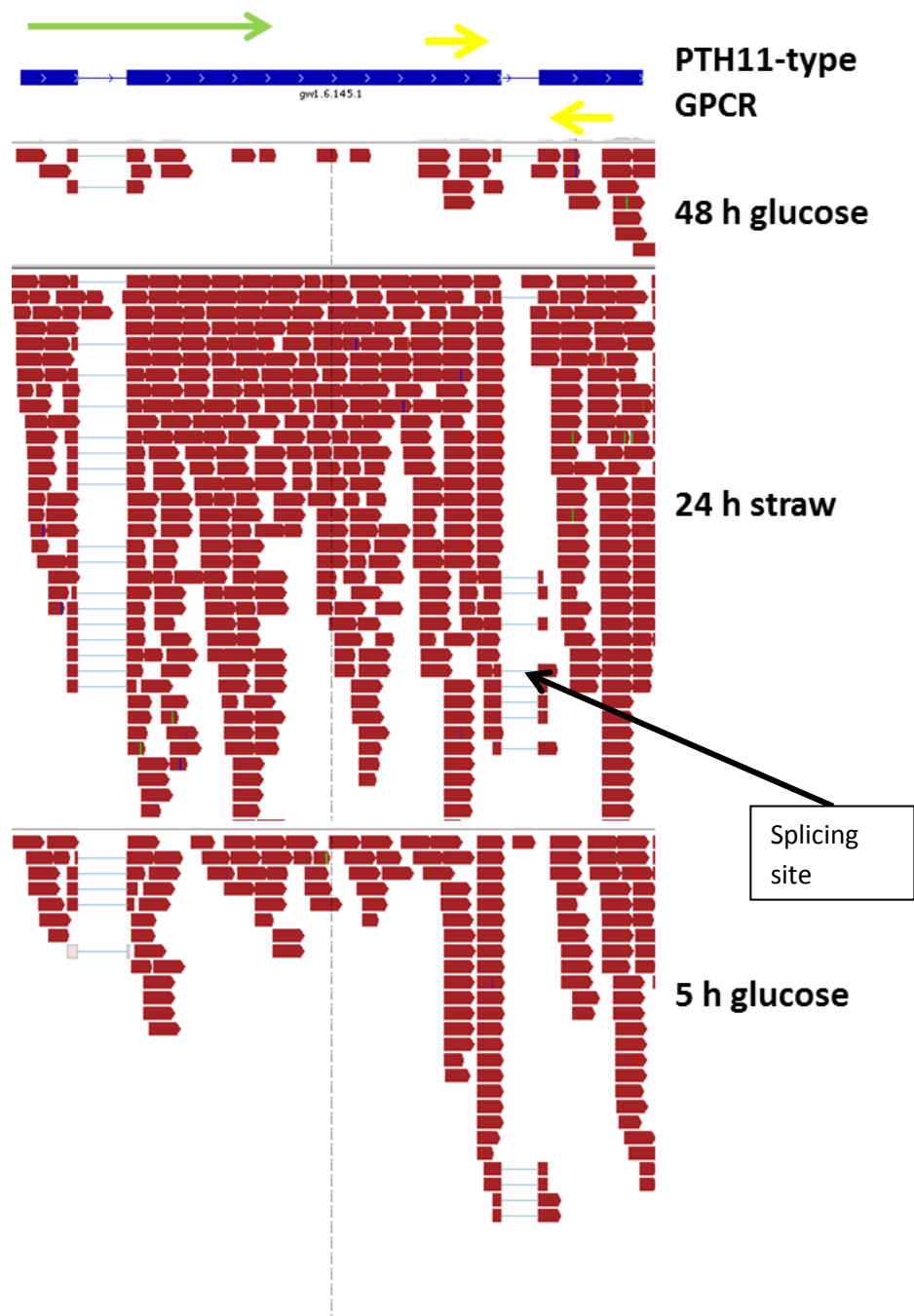| 81328	Unknown	Function unknown	1.000	1.073	0.891
111137	Unknown	Function unknown	3.829	6.409	5.254
68574	Unknown	Function unknown	0.812	0.329	0.000
123797	Unknown	Function unknown	0.571	14.878	0.155
69483	Unknown	Function unknown	0.219	0.626	0.889
111915	Unknown	Function unknown	0.032	1.050	0.000
111937	Unknown	Function unknown	3.292	6.773	2.518
52521	Uncharacterised conserved Urotein	Function unknown	1.073	1.405	0.457
5924	Unknown	Function unknown	0.162	1.601	0.233
70806	Unknown	Function unknown	1.244	1.075	0.576
112502	Unknown	Function unknown	0.119	4.172	0.086
119790	Aldehyde reductase	General function prediction only	0.013	3.043	0.000
119981	Peroxisomal membrane protein MPV17 and related proteins	General function prediction only	7.260	11.449	2.088
55561	Predicted protein with N-terminal TonB box	General function prediction only	4.990	3.329	4.116
3001	RNA binding protein FOG: RRM domain	General function prediction only	2.594	4.064	1.041
105565	Synaptic vesicle transporter SVOP and related transporters (major facilitator superfamily)	General function prediction only	0.280	0.698	0.049
121584	Zn-finger, nucleic acid binding (TF?)	General function prediction only	3.236	4.254	2.034
64758	Synaptic vesicle transporter SVOP and related transporters (major facilitator superfamily)	General function prediction only	3.334	4.086	2.562
79694	FAD-dependent oxidoreductase	General function prediction only	1.862	0.555	0.229
64608	WD40 protein DMR-N9	General function prediction only	1.147	4.299	0.882
65782	Alpha/beta hydrolase, esterase/lipase/thioesterase	General function prediction only	0.430	2.783	0.278
5007	Metallophosphoesterase	General function prediction only	2.458	1.086	1.373
50793	Tyrosinase	General function prediction only	7.533	0.633	0.225
67024	FOG: Zn-finger	General function prediction only	0.010	0.387	0.070
110910	Dystroglycan-type cadherin-like protein	General function prediction only	0.016	0.450	0.022
23228	Protein of unknown function DUF1479	General function prediction only	0.108	5.243	0.065
69840	Glucose/Ribitol dehydrogenase (Reductases with broad range of substrate specificities)	General function prediction only	0.138	0.746	0.155
70520	Glucose/Ribitol dehydrogenase (Reductases with broad range of substrate specificities)	General function prediction only	1.106	0.085	0.105
28409	Synaptic vesicle transporter SVOP and related transporters (major facilitator superfamily)	General function prediction only	0.000	2.068	0.018
22402	Fe2+/Zn2+ regulated transporter	Inorganic ion transport and metabolism	1.326	2.867	0.223
79398	Ca2+/H+ antiporter VCX1 and related proteins	Inorganic ion transport and metabolism	1.007	2.323	1.490
80672	Mitochondrial Fe2+ transporter MMT1 and related transporters (cation diffusion facilitator superfamily)	Inorganic ion transport and metabolism	6.188	11.692	6.316
81430	Ca2+ transporting ATPase	Inorganic ion transport and metabolism	0.243	22.293	0.642
54166	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	Lipid transport and metabolism	0.435	2.919	0.179

103215	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	Lipid transport and metabolism	0.147	2.099	0.187
56236	SAM-dependent methyltransferases	Lipid transport and metabolism	0.000	0.962	3.511
57370	Acyl-CoA synthetase	Lipid transport and metabolism	0.165	8.213	0.409
65760	Hormone-sensitive lipase HSL	Lipid transport and metabolism	0.997	4.213	0.755
5270	Peroxisomal 3-ketoacyl-CoA-thiolase P-44/SCP2	Lipid transport and metabolism	6.024	5.997	0.855
69647	Enoyl-CoA hydratase	Lipid transport and metabolism	1.084	2.120	0.538
123795	Dihydroorotase and related enzymes	Nucleotide transport and metabolism	1.463	0.637	1.107
76571	Predicted pyroglutamyl peptidase	Posttranslational modification, protein turnover, chaperones	2.114	5.444	1.617
31798	Aspartic endopeptidase	Posttranslational modification, protein turnover, chaperones	1.794	3.556	1.957
122363	Molecular chaperone (small heat-shock protein Hsp26/Hsp42)	Posttranslational modification, protein turnover, chaperones	0.620	1.049	0.073
49099	Glutaredoxin and related proteins	Posttranslational modification, protein turnover, chaperones	0.934	1.225	0.517
122870	Beta-1,6-N-acetylglucosaminyltransferase, contains WSC domain	Posttranslational modification, protein turnover, chaperones	0.977	1.314	0.577
65819	Molecular chaperones HSP70/HSC70, HSP70 superfamily	Posttranslational modification, protein turnover, chaperones	1.120	0.591	0.197
66521	Protein kinase	Posttranslational modification, protein turnover, chaperones	2.013	4.676	1.189
112328	Peptidase M14, carboxypeptidase A	Posttranslational modification, protein turnover, chaperones	0.393	0.171	0.053
55902	Single-stranded DNA-binding replication protein A (RPA), medium (30 kD) subunit	Replication, recombination and repair	2.279	3.788	2.056
59142	Apurinic/apyrimidinic endonuclease and related enzymes	Replication, recombination and repair	5.366	12.017	1.305
108157	Serine/threonine protein kinase of the CDC7 subfamily involved in DNA synthesis, repair and recombination	Replication, recombination and repair	0.031	2.111	0.106
64672	DNA ligase	Replication, recombination and repair	0.909	1.471	0.466
80159	DNA polymerase B (WD40 repeat)	Replication, recombination and repair	2.702	4.397	1.456
21198	Messenger RNA cleavage factor I subunit	RNA processing and modification	1.025	1.413	0.775
107385	Messenger RNA splicing factor ATP-dependent RNA helicase	RNA processing and modification	0.541	0.877	0.682
55566	Multicopper oxidases	Secondary metabolites biosynthesis, transport and catabolism	0.000	0.768	0.203
120696	Alcohol dehydrogenase, class V	Secondary metabolites biosynthesis, transport and catabolism	0.434	1.258	0.177
62263	Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	Secondary metabolites biosynthesis, transport and catabolism	0.265	0.172	0.058
49753	Sorbitol dehydrogenase	Secondary metabolites biosynthesis, transport and catabolism	0.039	31.066	0.139

109811	Cytochrome P450 CYP2 subfamily	Secondary metabolites biosynthesis, transport and catabolism	0.044	1.310	0.051
110499	Multidrug/pheromone exporter, ABC superfamily	Secondary metabolites biosynthesis, transport and catabolism	1.972	0.454	0.250
67964	Cytochrome P450 CYP2 subfamily	Secondary metabolites biosynthesis, transport and catabolism	0.145	9.013	0.406
48541	Serine/threonine protein kinase	Signal transduction mechanisms	3.154	3.793	1.602
62377	Exostosin EXT1L	Signal transduction mechanisms	5.205	2.381	4.385
64167	Sexual differentiation process protein ISP4	Signal transduction mechanisms	1.482	0.264	0.968
111103	Ca2+-modulated nonselective cation channel polycystin	Signal transduction mechanisms	1.055	2.655	0.937
123806	G protein-coupled receptor	Signal transduction mechanisms	0.030	1.818	0.000
72800	Serine/threonine protein kinase	Signal transduction mechanisms	0.121	0.806	0.137
103230	Fungal specific TF	Transcription	4.976	7.200	1.703
105263	Fungal transcription regulatory protein, zinc DNA binding cluster	Transcription	1.676	5.957	0.705
58853	Transcriptional regulator, binds DNA	Transcription	12.293	8.946	5.427
121164	Fungal specific TF	Transcription	0.311	1.947	0.487
121412	Fungal specific TF	Transcription	0.653	20.868	0.564
77513	Fungal specific TF	Transcription	0.802	15.902	1.516
78445	Fungal transcription regulatory protein, zinc DNA binding cluster	Transcription	2.003	1.741	0.153
72076	Fungal specific TF	Transcription	0.300	1.686	0.053
81517	Sirtuin 5 and related class III sirtuins (SIR2 family)	Transcription	0.306	4.299	0.861
68624	RNA polymerase II, large subunit	Transcription	0.424	0.818	0.408
81865	Polycomb enhancer protein, EPC	Transcription	1.268	4.434	1.314
55981	Transfer RNA cytosine-5-methylases and related enzymes of the NOL1/NOP2/sun superfamily	Translation, ribosomal structure and biogenesis	7.006	7.934	3.282
78843	Translation initiation factor 2B, gamma subunit (eIF-2Bgamma/GCD1)	Translation, ribosomal structure and biogenesis	6.833	8.980	2.035
63400	Exosomal 3'-5' exoribonuclease complex, subunit Rrp44/Dis3	Translation, ribosomal structure and biogenesis	2.247	1.911	0.324
82380	Translation initiation factor 2B, beta subunit (eIF-2Bbeta/GCD7)	Translation, ribosomal structure and biogenesis	1.756	2.488	2.074

Appendix A7

swo1 (swollenin)

estExt_fgenesh5_pg.C_290118

48 h glucose

Non-spliced RNA

Splicing site

24 h straw

5 h glucose

Figure A.7.F.1: IGV output of all the sense (red) and antisense (blue) transcripts mapped to the *swo1* and *pth11* genes of *T. reesei* QM6a mycelia grown in glucose for 48 h then transferred into straw for 24 h and with the addition of glucose to straw for 5 h. A diagram of the gene itself is shown in dark blue, the direction of the gene is indicated by a green arrow and yellow arrows mark the location of the primers used in subsequent RT-PCR reactions. Intron splicing sites are portrayed by a light blue line between two blocks of RNA transcripts. Splice sites and non-spliced RNAs are indicated by black arrows.
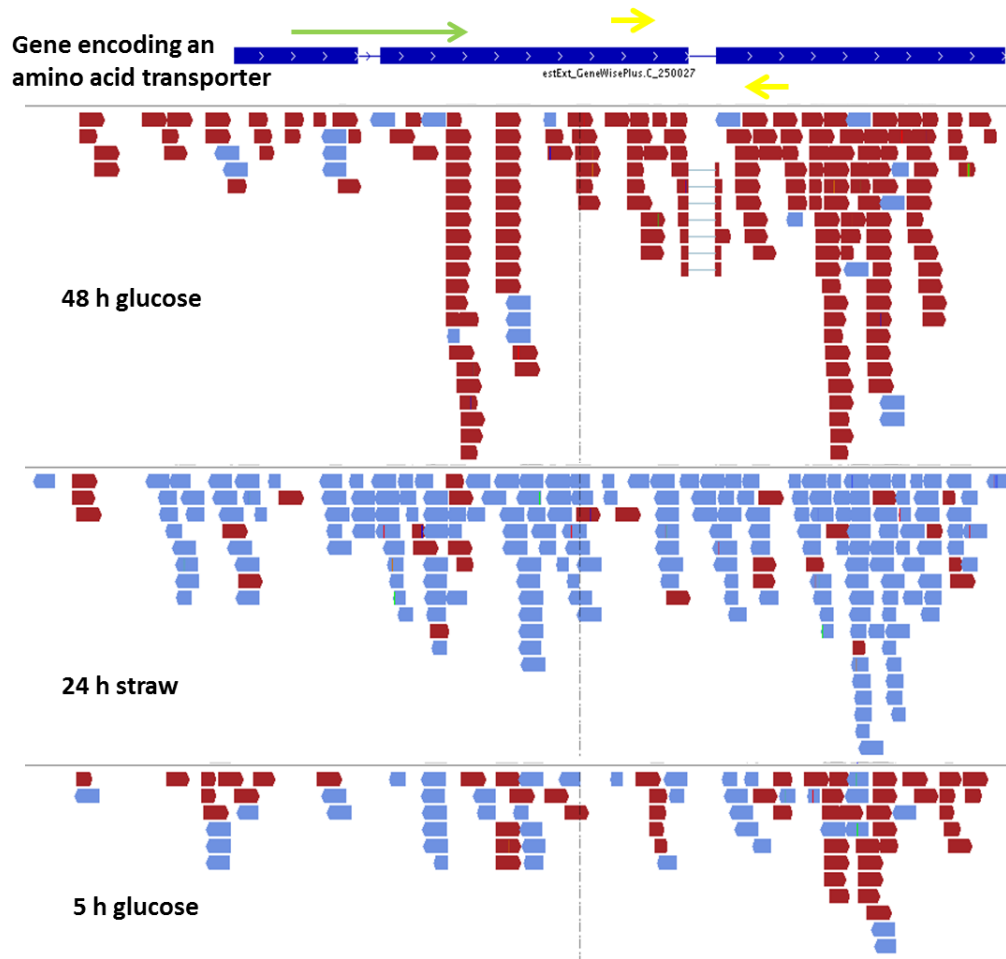
Figure A.7.F.2: IGV output of all the sense (red) and antisense (blue) transcripts mapped to a gene encoding an amino acid transporter in *T. reesei* QM6a mycelia grown in glucose for 48 h then transferred into straw for 24 h and with the addition of glucose to straw for 5 h. A diagram of the gene is shown in dark blue, the direction of the gene is indicated by a green arrow and yellow primers indicate the location of the primers used for further studies. Intron splicing sites are portrayed by light blue line between two blocks of RNA transcripts.
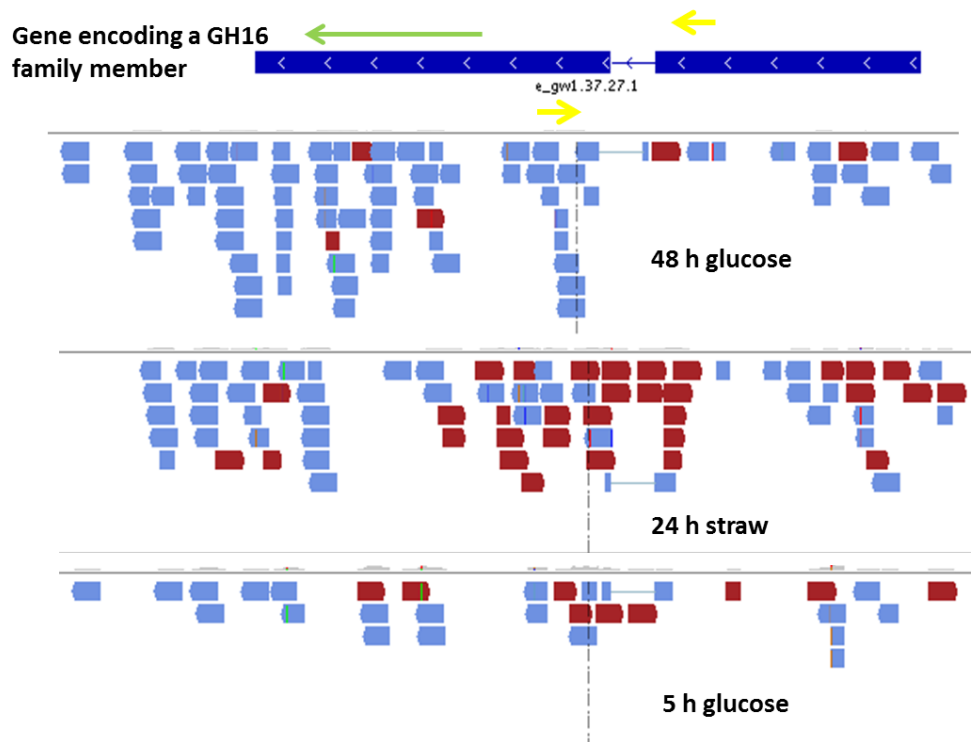
Figure A.7.F.3: IGV output of all the sense (blue) and antisense (red) transcripts mapped to a gene encoding a GH16 family protein in *T.reesei* QM6a mycelia grown in glucose for 48 h then transferred into straw for 24 h and with the addition of glucose to straw for 5 h. A diagram of the gene is shown in dark blue, the direction of the gene is indicated by a green arrow and yellow primers indicate the location of the primers used for further studies. Intron splicing sites are portrayed by light blue line between two blocks of RNA transcripts.

Gene encoding a fungal specific transcription factor

48 h glucose

24 h straw
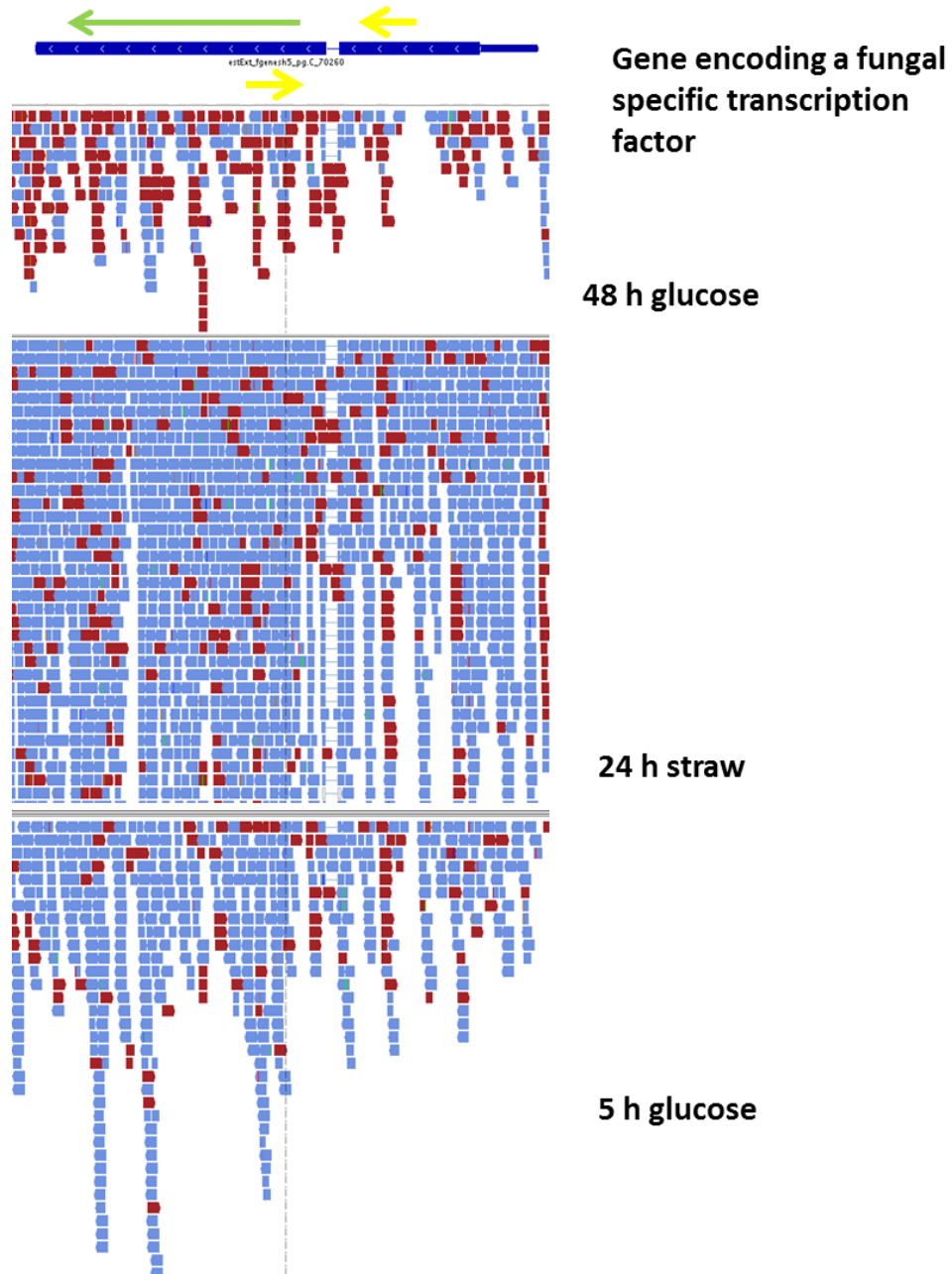
5 h glucose

Figure A.7.F.4: IGV output of all the sense (blue) and antisense (red) transcripts mapped to a gene encoding a transcription factor in *T. reesei* QM6a mycelia grown in glucose for 48 h then transferred into straw for 24 h and with the addition of glucose to 24 h straw for 5 h. A diagram of the gene is shown in dark blue, the direction of the gene is indicated by a green arrow and yellow primers indicate the location of the primers used for further studies. Intron splicing sites are portrayed by light blue line between two blocks of RNA transcripts.

Appendix A8

Table A.8.T.1: Expression values (RPKM) of all *T.reesei* histone protein, ATP-dependent chromatin re-modelling factor, enzymes mediating histone post-translational modifications and nucleosome assembly factor-encoding genes in 48 h glucose, 24 h straw and 24 h straw + 5 h glucose. Gene transcript I.Ds and annotations are also indicated (Note: Further genes may exist which could encode other chromatin-associated proteins due to one third of the *T. reesei* genome not being annotated).

Transcript I.D.	Annotation	RPKM 48 h glucose	RPKM 24 h straw	RPKM 5 h glucose
Histone proteins				
34402	Histone H1	7.568	19.322	8.619
1857	Histone H4	0.250	0.913	0.251
55201	Histone H4	5.298	13.195	3.945
80167	Histone H4 variant, *hh4.2*	58.183	95.470	106.647
82510	Histone H4 variant, *hh4.1*	110.163	103.552	129.199
57870	Histone H3, possible variant	2.917	4.021	5.495
124210	Histone H3, *hh3*	22.651	24.270	22.238
124052	Histone H2A, possible variant	5.076	6.864	4.815
121522	Histone H2A, *hh2A*	24.253	25.753	30.125
121516	Histone H2B	26.391	26.808	34.543
ATP-dependent chromatin re-modellers				
121196	SWI-SNF chromatin-remodelling complex protein	0.823	3.361	0.874
122943	SWI-SNF chromatin-remodelling complex protein	0.855	2.418	1.143
103804	SWI-SNF chromatin-remodelling complex protein	0.556	1.370	0.393
122283	SWI-SNF chromatin-remodelling complex protein	3.006	7.219	1.987
123327	SWI-SNF chromatin-remodelling complex protein	4.007	8.147	8.216
80835	SWI-SNF chromatin-remodelling complex protein	3.123	5.430	1.756
54670	Putative SWI-SNF chromatin-remodelling complex subunit (Snf5)	1.130	1.863	0.694
2826	SNF2-like family helicase	0.383	0.770	0.180
44747	SNF2 family helicase	3.450	124.959	9.333
63568	SNF2-like family helicase, ATPase domain	1.893	4.658	0.407
53381	SNF2 family ATPase	2.695	6.597	1.154
57608	Putative SNF2 family helicase/ATPase	2.500	4.931	1.564
2369	SNF2 family domain-containing protein	0.660	0.817	0.189

109526	SNF2 family helicase/ATPase	1.231	1.190	0.402
106138	SNF2 family DNA-dependent ATPase	0.025	0.098	0.018
22783	SNF2 family DNA-dependent ATPase	0.709	1.975	0.620
57935	Possible homolog of the *S. pombe* Snf2 chromatin remodelling protein	6.454	13.472	2.649
54598	Snf5/Smarcb1/Ini1	1.352	1.955	0.846
2269	Condensin complex component SMC2	4.805	9.068	1.888
25703	Condensin complex component CND2	1.195	2.999	0.734
61692	Chromosome condensation protein 3	4.720	7.111	4.104
57720	Putative Swr1p complex component (SWC5)	1.297	2.194	0.333
63484	Chromodomain-helicase-DNA-binding protein	1.196	2.523	0.458
50539	INO80 putative SNF2 helicase	3.672	7.257	1.569
58928	SNF2 family chromodomain-helicase DNA-binding protein, putative CHD1	4.483	6.746	1.389
77685	Putative SAGA complex bromodomain subunit Spt7	1.357	3.870	0.556
Proteins mediating post-translational histone modifications				
79414	Hypothetically part of transcriptional adaptor and HAT (histone acetyltransferase) complexes	3.300	10.888	1.912
123619	Transcription coactivator, component of the ADA and SAGA transcriptional adaptor/HAT complexes	2.840	5.320	0.789
64680	HAT, catalytic subunit of the ADA and SAGA complexes	0.617	1.756	0.201
110943	HAT SAGA/ADA, catalytic subunit PCAF/GCN5 and related proteins	0.008	0.051	0.011
21542	HAT ESA1 (GNAT family)	7.503	12.831	2.987
103311	HAT type B catalytic subunit	3.366	5.415	0.909
120339	HAT (MYST family)	2.989	4.797	0.526
120737	HAT type B catalytic subunit	1.897	2.598	1.339
33085	Subunit of TFIID and SAGA complexes	13.020	19.591	11.312
66433	GCN5 N-acetyltransferase (GNAT family)	5.545	1.628	1.119
103103	GCN5-N-Acetyltransferase (GNAT family)	0.170	0.114	0.035
111236	N-acetyltransferase of bacterial origin	0.104	3.009	0.600
4989	HAT	2.171	4.644	0.676
62071	Dosage compensation regulatory complex/HAT complex, subunit MSL-3/MRG15/EAF3, and related CHROMO domain-containing proteins	5.334	11.091	4.854
48386	RPD3-type class I HDAC (histone deacetyltransferase); regulates transcriptional silencing	0.741	1.689	0.600
79441	Subunit of HDAC complex	3.486	5.387	1.410
57220	Class I HDAC	0.698	0.904	0.269
80797	Class II, HDA-1 type HDAC	4.619	6.652	0.901
65533	Class II, HDA-1 type HDAC	0.271	1.618	0.404

67057	Sirtuin HDAC, Sir2 (silencing information regulator) family	2.968	3.793	0.963
79919	Sirtuin-type HDAC	5.129	11.147	2.102
120404	Possible sirtuin-type HDAC	1.599	3.608	0.675
50268	Sir2 family HDAC	11.516	9.098	3.332
81517	Possible sirtuin-type HDAC	0.306	4.299	0.861
54330	Possible member of Sir2 proteins	0.654	1.064	0.300
22492	Putative COMPASS complex protein	0.344	1.158	0.520
124283	Unknown protein, SET and MYND domains found in histone lysine methyltransferases	3.372	14.075	2.076
51287	SET domain protein, methyltransferase?	3.634	6.979	1.000
122500	Predicted histone tail methylase containing SET domain	2.472	4.075	0.806
32755	Predicted histone tail methylase containing SET domain	0.685	0.160	0.031
46209	SAM(S-adenosyl methionine)-methyltransferase, only found in Hypocreaceae	12.001	4.539	11.008
67931	Methyltransferase domain-containing protein	0.522	0.219	0.104
58521	Protein with SAM (and some other nucleotide) binding motif	0.043	0.020	0.000
53025	SAM-dependent methyltransferase	0.011	0.010	0.016
102855	Possibly a putative methyltransferase	1.696	2.113	2.682
64138	Protein involved in histone methylation	0.864	1.596	0.383
2591	UbiE-like methylase, SAM methyltransferase	0.774	1.737	0.903
65711	SAM-dependent methyltransferase	11.291	50.078	3.353
21557	Related to chromatin remodelling factors (contains SWIRM and DNA-binding Myb-domains), histone lysine de-methylase?	16.356	20.048	6.758
76515	Putative histone de-methylase, contains HMG1/2 box, Sec-1 like domain, amino oxidase	7.029	8.393	1.394
80997	Predicted RNA methylase, homologous to *A. nidulans* RmtA	1.126	2.403	1.534
107383	Arginine N-methyltransferase, homologous to *A. nidulans* RmtC	8.026	10.437	1.793
75646	Predicted RNA methylase only, homologous to *A. nidulans* RmtB	5.992	10.263	3.774
53402	Putative Serine/threonine protein kinase, homologue in *A. nidulans* is involved in H3-S10 phosphorylation	4.529	4.096	1.647
68364	Protein kinase, equivalent to *A. nidulans* nimA involved in H3-S10 phosphorylation	2.111	2.233	1.091
77795	LIMPET, E3-ubiquitin ligase SCON-2	3.943	6.107	5.839
27384	Putative E3 ubiquitin ligase complex F-box protein, *grr1*	0.730	2.306	0.615
119856	E3 ubiquitin-protein ligase/Putative upstream regulatory element binding protein	3.685	6.603	1.060
80553	ADP-ribosylation factor-like protein	0.583	2.506	0.450

57208	Putative ubiquitin-like modifier SUMO, similar to *sumO* in *A. nidulans*	19.286	35.251	5.875
Proteins involved in nucleosome assembly				
61508	Putative chromatin assembly protein, possible chaperone?	2.178	3.052	0.763
47838	Anti-silencing function protein 1, homologous to histone chaperone ASF1 in yeast	5.609	9.723	2.024
57295	Transcription elongation factor SPT6, chaperone for nucleosome assembly	4.441	7.630	2.186
57600	FACT complex protein (Facilitates chromatin transcription complex subunit SPT16)	4.723	7.133	2.632
80498	Nucleosome assembly protein NAP-1	23.759	32.779	12.400
51228	Putative Histone H1-binding protein	15.449	7.714	10.414
64684	Putative chromatin assembly factor 1 subunit B (CAF-1 B), other *A. nidulans* CAF-1 complex subunit not found	1.136	1.524	0.935

Appendix A9
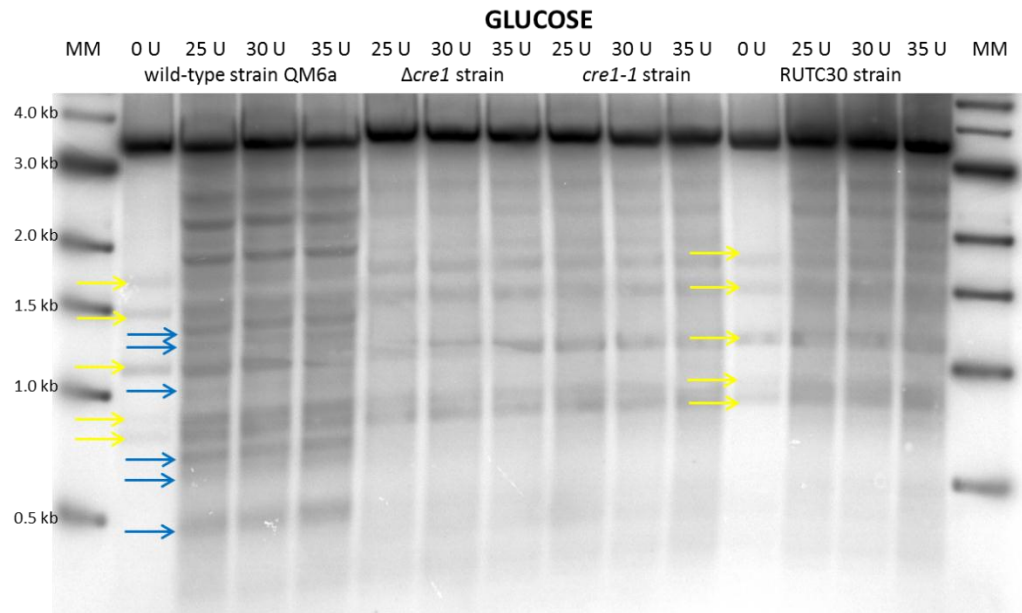
Figure A.9.F.1: Southern blot of MNase digested DNA extracted from glucose-grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marker). Blue arrows show differences in MH sites between strains and yellow arrows indicate bands which are also present in the absence of MNase.
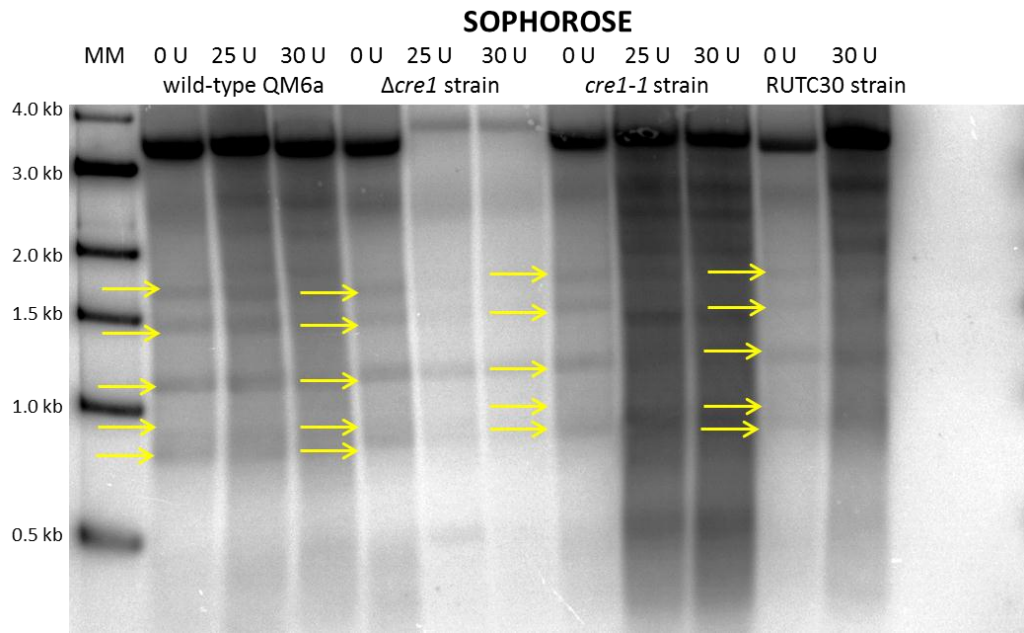
Figure A.9.F.2: Southern blot of MNase digested DNA extracted from sophorose-grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marker). Yellow arrows indicate bands which are also present in the absence of MNase.
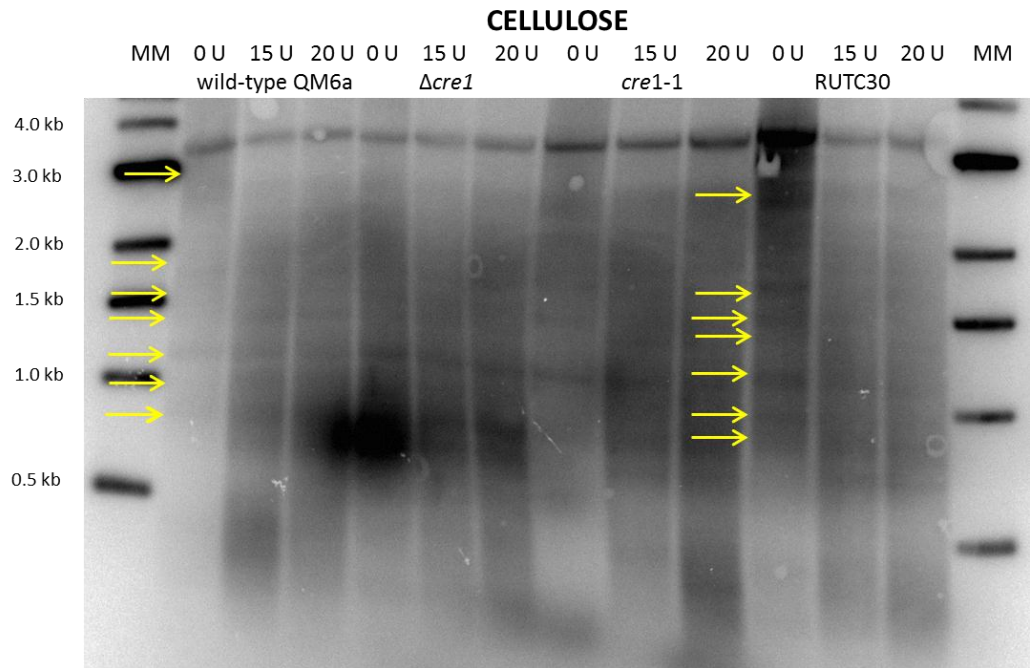
Figure A.9.F.3: Southern blot of MNase digested DNA extracted from cellulose-grown mycelia. MNase concentrations (in U) and strain are indicated above the blot picture (MM = molecular marke). Yellow arrows indicate bands which are also present in the absence of MNase.
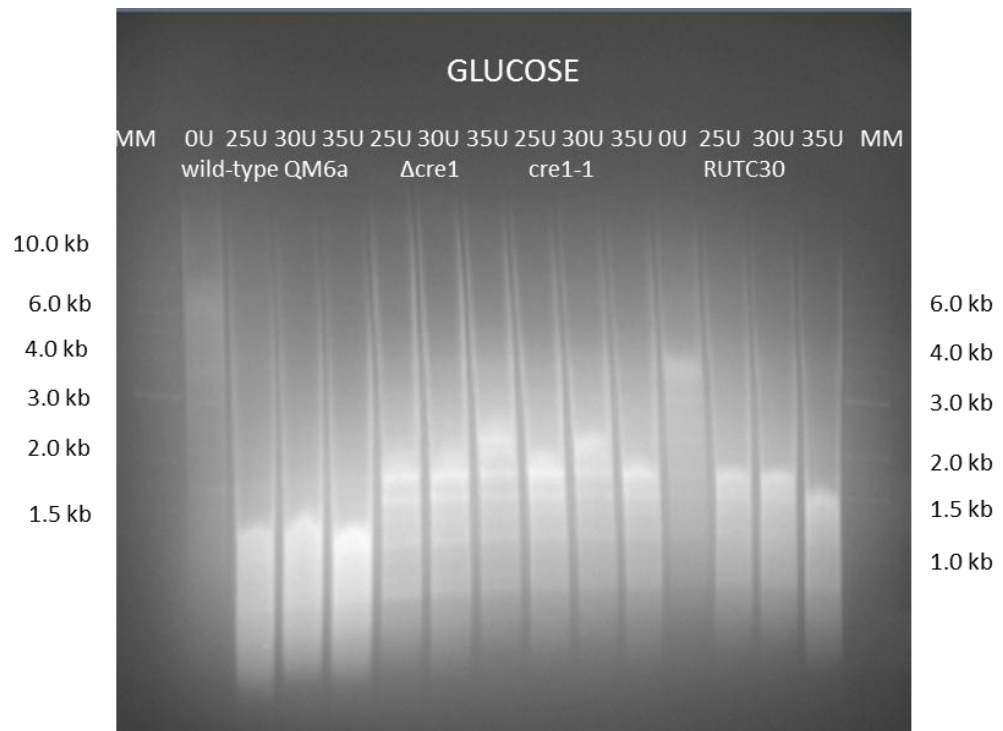
Figure A.9.F.4: Picture of an agarose gel containing gDNA extracted from glucose-grown mycelia after they were digested with MNAse and DNA was purified and subjected to a restriction enzyme digest. After gel electrophoresis, DNA was transferred onto a positively charged membrane which was hybridised with labelled probe and ladder. Excess probe was removed and blot was over-exposed. Results from this blot are shown Figure 5.3.8 of Chapter 5. MNase concentrations (in U) and strain are indicated in the gel picture (MM = molecular marke).

References

Abe, K., Furukawa, K., Fujioka, T., Hagiwara, D., Maeda, H., Marui, J., Mizutani, O., Takahashi, T., Yoshimi, A., Yamagat, Y., Gomi, K. and Hasegawa, F. (2010) Aspergillus: Molecular Biology and Genomics. Editor: Masayuki Machida and Katsuya Gomi. Caister Academic Press.

Akel, E., Metz, B., Seiboth, B. and Kubicek, C. P. (2009) Molecular regulation of arabinan and L-arabinose metabolism in *Hypocrea jecorina* (*Trichoderma reesei*). *Eukaryotic Cell* **8**: 1837-1844.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Molecular Biology of the Cell, 4th Edition. Editor: Sarah Gibbs. Garland Science, Taylor & Francis Group.

Alvira, P., Tomas-Pejo, E., Ballesteros, M. and Negro, M. J. (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresource Technology* **101**: 4851-4861.

Araujo-Palomares, C., Riquelme, M. and Castro-Longoria, E. (2009) The polarisome component SPA-2 localises at the apex of *Neurospora crassa* and partially colocalises with the Spitzenkörper. *Fungal Genetics and Biology* **46**: 551-563.

Aro, N., Ilmén, M., Saloheimo, A. and Penttilä, M. (2003) ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Applied and Environmental Microbiology* **69**: 56-65.

Aro, N., Saloheimo, A., Ilmén, M. and Penttilä, M. (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase Genes of *Trichoderma reesei*. *The Journal of Biological Chemistry* **276**: 24309-24314.

Arvas, M. Haiminen, N., Smit, B., Rautio, J., Vitikainen, M., Wiebe, M., Martinez, D., Chee, C., Kunkel, J., Sanchez, C., Nelson, M. A., Pakula, T., Saloheimo, M., Penttilä, M. and Kivioja, T. (2010) Detecting novel genes with sparse arrays. *Gene* **467**: 41-51.

Arvas, M., Pakula, T., Lanthaler, K., Saloheimo, M., Valkonen, M., Suortti, T., Robson, G. and Penttilä, M. (2006) Common features and interesting differences in transcriptional responses to secretion stress in the fungi

Trichoderma reesei and *Saccharomyces cerevisiae*. *BMC Genomics* **7**: 32-49.

Arya, G., Maitra, A., Grigoryev, S. A. (2010) A structural perspective on the where, how, why and what of nucleosome positioning. *Journal of Biomolecular Structure and Dynamics* **27**: 803-819.

Aslam, N., A Shiekh, M., Ashraf, M. and Jamil, A. (2010) Expression pattern of *Trichoderma* cellulases under different carbon sources. *Pakistan Journal of Botany* **42**: 2895-2902.

Ayrinhac, C., Margeot, A., Ferreira, N. L., Chaabane, F. B., Monot, F., Ravot, G., Sonet, J.-M. and Fourage, L. (2011) Improved saccharification of wheat straw for biofuel production using an engineered secretome of *Trichoderma reesei*. *Organic Process Research & Development* **15**: 275-278.

Baldrian, P., Voriskova, J, Dobiasova, P., Merhautova, V., Lisa, L. and Valaskova, V. (2011) Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi from the upper layers of forest soil. *Plant Soil* **338**: 111-125.

Banerjee, G., Scott-Craig, J. S. and Walton, J. D. (2010) Improving enzymes for biomass conversion: a basic research perspective. *Bioenergy Research* **3**: 82-92.

Becker, P. B. (2002) Nucleosome sliding: facts and fiction. *The EMBO Journal* **21**: 4749-4753.

Beisel, C. and Paro, R. (2011) Silencing chromatin: comparing modes and mechanisms. *Nature Reviews Genetics* **12**: 123-145.

Belshaw, N. J., Ilmén, M., Penttilä, M. E. and Archer, D. B. (2001) Nucleosome positioning and the role of CRE1 in the regulation of cellobiohydrolase I gene expression in *Trichoderma reesei*. Unpublished manuscript.

Bloom, J. S., Khan, Z., Kruglyak, L., Singh, M. and Caudy, A. A. (2009) Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel expression microarrays. *BMC Genomics* **10**: 221-230.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemistry Journal* **382**: 769-781.

Börjesson, J., Engqvist, M., Sipos, B. and Tjerneld, F. (2007) Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocelluloses. *Enzyme and Microbial Technology* **41**: 186-195.

Borkovich, K. A., Alex, L. A., Yarden, O., Freitag, M., Turner, G. E., Read, N. D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., Plamann, M., Goodrich-Tanrikulu, M., Schulte, U., Mannhaupt, G., Nargang, F. E., Radford, A., Selitrennikoff, C., Galagan, J. E., Dunlap, J. C., Loros, J. J., Catcheside, D., Inoue, H., Aramayo, R., Polymenis, M., Selker, E. U., Sachs, M. S., Marzluf, G. A., Paulsen, I., Davis, R., Ebbole, D. J., Zelter, A., Kalkman, E. R., O'Rourke, R., Bowring, F., Yeadon, J., Ishii, C., Suzuki, K., Sakai, W. and Pratt, R. (2004) Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiology and Molecular Biology Reviews* **68**: 1-108.

Bourne, Y. and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current opinion in structural biology* **11**: 593-600.

Brand, A. and Gow, N. A. R. (2009) Mechanisms of hypha orientation of fungi. *Current Opinion in Microbiology* **12**: 350-357.

Brosch, G., Loidl, P. and Graessle, S. (2008) Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiology Reviews* **32**: 409-439.

Brunner, K., Omann, M. Pucher, M. E., Delic, M., Lehner, S. M., Domnanich, P., Kratochwill, K., Druzhinina, I., Denk, D. and Zeilinger, S. (2008) *Trichoderma* G protein-coupled receptors: functional characterisation of a cAMP receptor-like protein from *Trichoderm atroviride*. *Current Genetics* **54**: 283-299.

Castellanos, F., Schmoll, M., Martinez, P., Tisch, D., Kubicek, C. P., Herrera-Estrella, A. and Esquivel-Naranjo, E. U. (2010) Crucial factors of the light perception machinery and their impact on growth and cellulase gene transcription in *Trichoderma reesei*. *Fungal Genetics and Biology* **47**: 468-476.

Chambergo, F. S., Bonaccorsi, E. D., Ferreira, A. J. S., Ramos, A. S. P., Junior, J. R. F., Abrahao-Neto, J., Farah, J. P. S. and El-Dorry, H. (2002) Elucidation of the Metabolic Fate of Glucose in the Filamentous Fungus *Trichoderma reesei* using Expressed Sequence Tag (EST)

Analysis and cDNA Microarrays. *The Journal of Biological Chemistry* **277**: 13983-13988.

Chung, H.-R., Dunkel, I., Heise, F., Linke, C., Krobitsch, S., Ehrenhofer-Murray, A. E., Sperling, S. R. and Vingron, M. (2010) The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE* **5**: 15754-15761.

Cichewicz, R. H. (2010) Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. *Natural product reports* **27**: 11-22.

Clark, D. J. (2010) Nucleosome positioning, nucleosome spacing and the nucleosome code. *Journal of Biomolecular Structure & Dynamics* **27**: 781-791.

Cockerill, P. N. (2011) Structure and function of active chromatin and DNase I hypersensitive sites. *The FEBS Journal* **278**: 2182-2210.

Coradetti, S. T., Craig, J. P., Xiong, Y., Schock, T., Tian, C. and Glass, L. N. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *PNAS* **109**: 7397-7402.

Cziferszky, A., Mach, R. L. and Kubicek C. P. (2002) Phosphorylation Positively Regulates DNA Binding of the Carbon Catabolite Repressor CRE1 of *Hypocrea jecorina* (*Trichoderma reesei*). *The Journal of Biological Chemistry* **277**: 14688-14694.

Dang, Y., Yang, Q., Xue, Z. and Liu, Y. (2011) RNA interference in fungi: pathways, functions and applications. *Eukaryotic Cell* **10**: 1148-1155.

Dashtban, M., Schraft, H. and Qin, W. (2009) Fungal Bioconversion of Lignocellulosic Residues; Opportunities and perspectives. *International Journal of Biological Sciences* **5**: 578-595.

Deacon, J. W. (2006) Fungal Biology 4th Edition. Blackwell publishing.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N. D., Yong-Hwan, L., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galangan, J. E. and Birren, B. W. (2005) The genome sequence of the rice blast fungus *Magnaporthe oryzae*. *Nature* **434**: 980-986.

Delmas, S., Pullan, S., Gaddipati, S., Kokolski, M., Malla, S, Blythe, M. J., Ibbett, R., Campbell, M., Liddell, S., Aboobaker, A., Tucker, G. A. and Archer, D. B. (2012) Uncovering the genome-wide transcriptional responses of the filamentous fungus *Aspergillus niger* to lignocellulose using RNA sequencing. *PLoS Genetics* **8**: e1002875.

De Vries, L. and Gist Farquhar, M. (1999) RGS proteins: more than just GAPs for heterotrimeric G proteins. *Cell Biology* **9**: 138-144.

Dodd, D. and Cann, I. K. O. (2009) Enzymatic deconstruction of xylan for biofuel production. *GCB Bioenergy* **1**: 2-17.

Dohlman, H. G. and Thorners, J. (1997) RGS Proteins and signalling by heterotrimeric G proteins. *The journal of biological chemistry* **272**: 3871-3874.

Do Nascimento Silva, R., Steindorff, A. S., Ulhoa, C. J. and Felix, C. R. (2009) Involvement of G-alpha protein GNA3 in production of cell wall-degrading enzymes by *Trichoderma reesei* (*Hypocrea jecorina*) during mycoparasitism against *Pythium ultimum*. *Biotechnology Letters* **31**: 531-536.

Dorschner, M. O., Hawrylycz, M., Humbert, R., Wallace, J. C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P. J., Kohli, A., Li, Q., McArthur, M. and Stamatoyannopoulos, J. A. (2004) High-throughput localisation of functional elements by quantitative chromatin profiling. *Nature Methods* **1**: 219-225.

Druzhinina, I., Komon-Zelazowska, M., Atanasova, L., Seidl, V. and Kubicek, C. P. (2010) Evolution and Ecophysiology of the industrial producer *Hypocrea jecorina* (anamorph *Trichoderma reesei*) and a new sympatric agamospecies related to it. *PLoS ONE* **5**: 1-14.

Druzhinina, I. S., Schmoll, M., Seiboth, B. and Kubicek, C. P. (2006) Global carbon utilisation profiles of wild-type, mutant, and transformant strains of *Hypocrea jecorina*. *Applied and Environmental Microbiology* **72**: 2126-2133.

El-Gogary, S., Leite, A., Crivellaro, O., Eveleigh, D. E. and El-Dorry, H. (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. *PNAS* **86**: 6138-6141.

Elkins, J. G., Raman, B. and Keller, M. (2010) Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Current Opinion in Biotechnology* **21**: 657-662.

Faghihi, M. A., and Wahlestedt, C. (2009) Regulatory roles of natural antisense transcripts. *Nature Reviews* **10**: 637-643.

Fekete, E., Karaffa, L., Kubicek, C. P., Szentirmai, A. and Seiboth B. (2007) Induction of extracellular β-galactosidase (BGA1) formation by D-galactose in *Hypocrea jecorina* is mediated by galactitol. *Microbiology* **153**: 507-512.

Fekete, E., Seiboth, B., Kubicek, C. P., Szentirmai, A. and Karaffa, L. (2008) Lack of aldose 1-epimerase in *Hypocrea jecorina* (anamorph *Trichoderma reesei*): A key to cellulase gene expression on lactose. *PNAS* **105**: 7141-7146.

Felsenfeld, G. and Groudine, M. (2003) Controlling the double helix. *Nature* **421**: 448-453.

Ferreira de Oliveira, J. M. P. and De Graaff, L. H. (2011) Proteomics of industrial fungi: trends and insights for biotechnology. *Applied Microbiology and Biotechnology* **89**: 225-237.

Flaus, A. and Owen-Hughes, T. (2011) Mechanisms of ATP-dependent chromatin remodelling: the means to the end. *The FEBS Journal* **278**: 3579-3595.

Fillingham, J., Kainth, P., Lambert, J.-P., Van Bakel, H., Tsui, K., Pena-Castillo, L., Nislow, C., Figeys, D., Hughes, T. R., Greenblatt, J. and Andrews, B. J. (2009) Two-color cell array screen reveals interdependent roles for histone chaperones and a chromatin boundary regulator in histone gene repression. *Molecular Cell* **35**: 340-351.

Fischle, W. (2011) Talk is cheap – cross-talk in establishment, maintenance, and readout of chromatin modifications. *Genes & Development* **22**: 3375-3382.

Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M. (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *The Journal of Biological Chemistry* **278**: 31988-31997.

Fukuda, H., Kondo, A. and Tamalampudi, S. (2009) Bioenergy: Sustainable fuels from biomass by yeast and fungal whole-cell biocatalysts. *Biochemical Engineering Journal* **44**: 2-12.

Fujii, T., Murakami, K. and Sawayama, S. (2010) Cellulase hyperproducing mutants derived from the fungus *Trichoderma reesei* QM9414 produced large amounts of cellulases at the enzymatic and transcriptional levels. *Bioscience Biotechnology and Biochemistry* **74**: 419-422.

Furukawa, T., Shida, Y., Kitagami, N., Mori, K., Kato, M., Kobayashi, T., Okada, H., Ogasawara, W. and Morikawa, Y. (2009) Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. *Fungal Genetics and Biology* **46**: 564-574.

García, I., Gonzalez, R., Gómez, D. and Scazzocchio, C. (2004) Chromatin Rearrangements in the *prnD-prnB* Bidirectional Promoter: Dependance on Transcription Factors. *Eukaryotic Cell* **3**: 144-156.

Georg, J. and Hess, W. R. (2011) *cis*-Antisense RNA, another level of gene regulation in bacteria. *Microbiology and Molecular Biology Reviews* **75**: 286-300.

Gonzalez, R. and Scazzocchio, C. (1997) A rapid method for chromatin structure analysis in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Research* **25**: 3955-3956.

Geysens, S., Whyteside, G. and Archer, D. B. (2009) Genomics of protein folding in the endoplasmic reticulum, secretion stress and glycosylation in the aspergilla. *Fungal Genetics and Biology* **46**: S121 – S140.

Graessle, S., Loidl, P. and Brosch, G. (2001) Histone acetylation: plants and fungi as model systems for the investigation of histone deacetylases. *Cellular and Molecular Life Sciences* **58**: 704-720.

Gremel, G., Dorrer, M. and Schmoll, M. (2008) Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*). *BMC Microbiology* **8**: 1471-1489.

Grewal, S. I. S. (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Current Opinion in Genetics & Development* **20**: 134-141.

Guangtao, Z., Seiboth, B., Wen, C., Yaohua, Z., Xian, L. and Wang, T. (2010) A novel carbon-source dependent genetic transformation system for the

versatile cell factory *Hypocrea jecorina* (anamorph *Trichoderma reesei*). *FEMS Microbiology Letters* **303**: 26-32.

Guangtao, Z., Hartl, L., Schuster, A., Polak, S., Schmoll, M., Wang, T., Seidl, V. and Seiboth, B. (2009) Gene targeting in a nonhomologous en joining deficient *Hypocrea jecorina*. *Journal of Biotechnology* **139**: 146-151.

Hainer, S. J. and Martens, J. A. (2011) Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. *Molecular and Cellular Biology* **31**: 3557-3568.

Häkkinen, M., Arvas, M., Oja, M., Aro, N., Penttilä, M., Saloheimo, M. and Pakula, T. M. (2012) Transcriptional analysis in täe presence of different lignocellulose substrates reveals co-regulated groups of genes encoding carbohydrate active enzymes in *Trichoderma reesei*. *Microbial Cell factories* unpublished paper.

Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J.-C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* **49**: 3305-3316.

Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology* **7**: 637-644.

Henrique-Silva, F., El-Gogary, S., Carle-Urioste, J.C., Metheucci Jr., E., Crivellaro, O. and E-Dorry, H. (1996) Two regulatory regions controlling basal and cellulose-induced expression of the gene encoding *cellobiohydrolase 1* of *Trichoderma reesei* are adjacent to its TATA box. *Biochemical and Biophysical research communications* **228**: 229-237.

Hervé, C., Rogowski, A., Blake, a. W., Marcus, S. E., Gilbert, H. J. and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Science* **107:** 15293-15298.

Higuchi, Y., Arioka, M. and Kitamoto, K. (2009) Endocytic recycling at the tip region in the filamentous fungus *Aspergillus oryzae*. *Communicative & Integrative Biology*, **2**: 327-328.

Hildén, L and Johansson, G. (2004) Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnology Letters* **26**: 1683-1693.

Hoyos-Carvajal, L., Orduz, S. and Bissett, J. (2009) Genetic and metabolic diversity of *Trichoderma* from Colombia and adjacent neotropic regions. *Fungal Genetics and Biology* **44**: 615-631.

Ike, M., Park, J.-Y., Tabuse, M. and Tokuyasu, K. (2010) Cellulase production on glucose-based media by the UV-irradiated mutants of *Trichoderma reesei*. *Applied Microbiology and Biotechnology* **87**: 2059-2066.

Ilmén, M., Saloheimo, A., Onnela, M.-L. and Penttilä, M. E. (1997) Regulation of Cellulase Gene Expression in the Filamentous Fungus *Trichoderma reesei*. *Applied and Environmental Microbiology* **63**: 1298-1306.

Jeoh, T., Michener, W., Himmel, M. E., Decker, S. R. and Adney, W. S. (2008) Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnology for Biofuels* **1**: 1186-1198.

Jouhten, P., Pitkänen, E., Pakula, T., Saloheimo, M., Penttilä, M. and Maaheimo, H. (2009) ^{13}C-metabolic flux ratio and novel carbon path analyses confirmed that *Trichoderma reesei* uses primarily the respirative pathway also on the preferred carbon source glucose. *BMC Systems Biology* **3**: 509-525.

Juhász, T., Szengyel, Z., Réczey, K., Siika-Aho, M. and Viikari, L. (2005) Characterisation of cellulases and hemicellulases produced by *Trichoderma reesei* on various Carbon Sources. *Process Biochemistry* **40**: 3519-3525.

Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C. and Sandgren, M. (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina* at 1.6 Å resolution. *Journal of Molecular Biology* **383**: 144-154.

King, A. J., Cragg, S. M., Li, Y., Dymond, J., Guille, M. J., Bowles, D. J., Bruce, N. C., Graham, I. A. and McQueen-Mason, S. J. (2010) Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut micrboes. *PNAS* **107**: 5345-5350.

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A. and Blanch, H. W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering* **109**: 1083-1087.

Knob, A., Terrasan, C. R. F. and Carmona, E. C. (2010) β-xylosidases from filamentous fungi: an overview. *World Journal of Microbial Biotechnology* **26**: 389-407.

Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M. and Seiboth, B. (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina. Biotechnology for Biofuels* **2**: 6834-6848.

Kulkarni, R. D, Thon, M. R., Pan, H. and Dean, R. A. (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea. Genome Biology* **6**: R24-R39.

Lapidot, M. and Pilpel, Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO reports* **7**: 1216-1222.

Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J. K., Culley, D. E., Collett, J. R., Martin, J., Druzhinina, I. S., Mathis, H. Monot, F., Seiboth, B., Cherry, B., Rey, M., Berka, R., Kubicek, C. P., Baker, S. E. and Margeot, A. (2009) Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *PNAS* **106**: 16151-16156.

Lee, H.-C., Li, L., Gu, W., Xue, Z., Crosthwaite, S. K., Pertsemlidis, A., Lewis, Z. A., Freitag, M., Selker, E. U., Mello, C. C. and Liu, Y. (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Molecular Cell* **38**: 803-814.

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R. and Nislow, C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**: 1235-1243.

Leeder, A. C. and Turner, G. (2008) Characterisation of *Aspergillus nidulans* polarisome component BemA. *Fungal Genetics and Biology* **45**: 897-911.

Lefkowitz, R. J. (1998) G Protein-coupled receptors. *The Journal of Biological Chemistry* **273**: 18677-18680.

Li, G.-D, Fang, J.-X., Chen, H.-Z., Luo, J., Zheng, Z.-H., Shen, Y.-M. and Wu, Q. (2007) Negative regulation of transcription coactivator p300 by orphan receptor TR3. *Nucleic Acids Research* **35**: 7348-7359.

Li, L., Wright, S. J., Krystofova, S., Park, G. and Borkovich, K. A. (2007) Heterotrimeric G protein signalling in filamentous fungi. *Annual Review of Microbiology* **61**: 423-452.

Linder, M. B., Szilvay, G. R., Nakari-Setälä, T. and Penttilä, M. (2005) Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiology Reviews* **29**: 877-896.

Ling, M., Qin, Y., Li, N. and Liang, Z. (2009) Binding of two transcriptional factors, XYR1 and ACE1 in the promoter region of cellulase *cbh1* gene. *Biotechnological Letters* **31**: 227-231.

Lingner, U., Münch, S., Deising, H.B. and Sauer, N. (2011) Hexose transporters of a hemibiotrophic plant pathogen: Functional variations and regulatory differences at different stages of infection. *The Journal of Biological Chemistry* **286:** 20913-20922.

Liu, Y.-S., Baker, J. O., Zeng, Y., Himmel, M. E., Haas, T. and Ding, S.-Y. (2011) Cellobiohydrolase hydrolyses crystalline cellulose on hydrophobic faces. *The Journal of biological Chemistry* **286**: 11195 – 11201.

Liu, T., Wang, T., Li, X. and Liu, X. (2008) Improved heterologous gene expression in *Trichoderma reesei* by *cellobiohydrolase I* gene (*cbh1*) promoter optimisation. *ABBS* **40**: 158-165.

Lo, C.-M. and Ju, L.-K. (2009) Sophorolipids-induced cellulase production in cocultures of *Hypocrea jecorina* Rut C30 and *Candida bombicola*. *Enzyme and Microbial Technology* **44**: 107-111.

Loney, E. R., Inglis, P. W., Sharp, S., Pryde, F. E., Kent, N. A., Mellor, J. and Louis, E. J. (2009) Repressive and non-repressive chromatin at native telomeres in *Saccharomyces cerevisiae*. *Epigenetics and Chromatin* **2**: 1756-1769.

Lorch, Y., Griesenbeck, J., Boeger, H., Maier-Davis, B. and Kornberg, R. D. (2011) Selective removal of promoter nucleosomes by the RSC chromatin-remodeling complex. *Nature Structural & Molecular Biology* **18**: 881-886.

Lu, Y., Su, C., Mao, X., Palaraniga, P., Liu, H. and Chen, J. (2008) Efg1-mediated recruitment of NuA4 to Promoters is required for hypha-specific Swi/Snf binding and activation in *Candida albicans*. *Molecular Biology of the Cell* **19**: 4260-4272.

Mach-Aigner, A. R., Grosstessner-Hain, K., Pocas-Fonseca, M. J., Mechtler, K. and Mach, R. L. (2010) From an electrophoretic mobility shift assay to isolated transcription factors: a fast genomic-proteomic approach. *BMC Genomics* **11**: 644-746.

Mach-Aigner, A. R., Pucher, M. E. and Mach, R. L. (2010) D-xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). *Applied and Environmental Microbiology* **76**: 1770-1776.

Mach-Aigner, A. R., Pucher, M. E., Steiger, M. G., Bauer, G. E., Preis, S. J. and Mach, R. L. (2008) Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. *Applied and Environmental Microbiology* **74**: 6554-6562.

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

Marfella, C. G. A. and Imbalzano, A. N. (2007) The Chd family of chromatin remodelers. *Mutational Research* **1**: 30-40.

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

Margolles-Clark, E., Ilmén, M. and Penttilä, M. (1997) Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. *Journal of Biotechnology* **57**: 167-179.

Marín-Navarro, J. and Polaina, J. (2011) Glucoamylases: structural and biotechnological aspects. *Applied Microbiology and Biotechnology* **89**: 1267-1273.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. and Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* **18**: 1509-1517.

Martinez, D., Berka, R. M, Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G. J., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C., Han, C. S., Ho, I., Larrondo, L. F., Lopez de Leon, A., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J.,

Barbote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M. and Brettin, T. S. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology* **26**: 553-560.

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

Mathieu, M., Nikolaev, I., Scazzocchio, C. and Felenbok, B. (2005) Patterns of nucleosomal organisation in the *alc* regulon of *Aspergillus nidulans*: roles of the AlcR transcriptional activator and the CreA global repressor. *Molecular Microbiology* **56**: 535-548.

Matzke, M. A. and Matzke, A. J. (2004) Planting the seeds of a new paradigm. *PLoS Biology* **2**:e133.

McCudden, C. R., Hains, M. D., Kimple, R. J., Siderovski, D. P. and Willard F. S. (2005) G-protein signalling: back to the future. *Cellular and Molecular Life Sciences* **62**: 551-577.

Medve, J., Lee, D. and Tjerneld, F. (1998) Ion-exchange chromatographic purification and quantitative analysis of *Trichoderma reesei* cellulases cellobiohydrolase I, II and endoglucanase II by fast protein liquid chromatography. *Journal of Chromatography* **808**: 153-165.

Mikus, M., Hatvani, L., Neuhof, T., Komón-Zelazowska, M., Dieckmann, R., Schwecke, T., Druzhinina, I. S., Von Döhren, H. and Kubicek, C. P. (2009) Differential regulation and posttranslational processing of the class II hydrophobin genes from the biocontrol fungus *Hypocrea atroviridis*. *Applied and Environmental Microbiology* **75**: 3222-3229.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods* **5**: 621-628.

Mulakala, C. and Reilly, P. J. (2005) *Hypocrea jecorina* (*Trichoderma reesei*) CEL7A as a Molecular Machine: A Docking Study. *PROTEINS: Structure, Function and Bioinformatics* **60**: 598-605.

Muro-Pastor, I, Gonzalez, R., Strauss, J., Narendja, F. and Scazzocchio, C. (1999) The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter. *The EMBO Journal* **18**: 1584-1597.

Nakari-Setälä, T., Paloheimo, M., Kallio, J., Vehmaanperä, J., Penttilä, M and Saloheimo, M. (2009) Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. *Applied and Environmental Microbiology* **75**: 4853-4860.

Ni, T., Tu. K., Wang, Z., Song, S., Wu, H., Xie, B., Scott, K. C., Grewal, S. I., Gao, Y. and Zhu, J. (2010) The prevalence and regulation of antisense transcripts in *Schizosaccharomyces pombe*. *PloS ONE* **5**: e15271.

Nishida, H., Motoyama, T., Yamamoto, S., Aburatani, H. and Osada, H. (2009) Genome-wide maps of mono- and di-nucleosomes of *Aspergillus fumigatus*. *Bioinformatics* **25**: 2295-2297.

Ohtaki, S., Maeda, H., Takahashi, T., Yamagata, Y., Hasegawa, F., Gomi, K., Nakajima, T. and Abe, K. (2006) Novel hydrophobic surface binding protein, HsbA, produced by *Aspergillus oryzae*. *Applied and Environmental Microbiology* **72**: 2407-2413.

Palmer, J. M., Perrin, R. M., Dagenais, T. R. T. and Keller, N. P. (2008) H3K9 methylation regulates growth and development in *Aspergillus fumigatus*. *Eukaryotic Cell* **7**: 2052-2060.

Pao, S. S., Paulsen, I. T. and Saier Jr. M. H. (1998) Major facilitator Superfamily. *Microbiology and Molecular Biology Reviews* **62**: 1-34.

Pel, H. J., De Winde, J. H., Archer, D. B., Dyer, P. S., Hofmann, G., Schaap, P. J., Turner, G., De Vries, R. P., Albang, R., Albermann, K., Andersen, M. R., Bendtsen, J. D., Benen, J. A. E., Van den Berg, M., Breestraat, S., Caddick, M. X., Contreras, R., Cornell, M., Coutinho, P. M., Danchin, E. G. J., Debets, A. J. M., Dekker, P., Van Dijck, P. W. M., Van Dijk, A., Dijkhuizen, L., Driessen, A. J. M., D'Enfert, C., Geysens, S., Goosen, C., Groot, G. S. P., De Groot, P. W. J., Guillemette, T., Henrissat, B., Herweijer, M., Van den Hombergh, J. P. T. W., Van den Hondel, C. A. M. J. J., Van der Heijden, R. T. J. M., Van der Kaaij, R. M., Klis, F. M., Kools, H. J., Kubicek, C. P., Van Kuyk, P. A., Lauber, J., Lu, X., Van der Maarel, M. J. E. C., Meulenberg, R., Menke, H., Mortimer, M. A., Nielsen, J., Oliver, S. G., Olsthoorn, M., Pal, K., Van Peij, N. N. M. E., Ram, A. F. J., Rinas, U., Roubos, J. A., Sagt, C. M. J., Schmoll, M., Sun, J., Ussery, D., Varga, J., Vervecken, W., Van der Vondervoort, P. J. J., Wedler, H., Wösten, H. A. B., Zeng, A.-P., Van Ooyen, A. J. J., Visser, J. and Stam, H. (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nature Biotechnology* **25**: 221-231.

Peñalva, M. Á. (2010) Endocytosis in filamentous fungi: Cinderella gets her reward. *Current Opinion in Microbiology* **13**: 684-692.

Piatti, P., Zeilner, A. and Lusser, A. (2011) ATP-dependent chromatin remodelling factors and their roles in affecting nucleosome fibre composition. *International Journal of Molecular Sciences* **12**: 6544-6565.

Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Chaabane, F. B., Linke, R., Seiboth, B. and Kubicek, C. P. (2011) Differential regulation of the cellulase transcription factors XYR1, ACE2 and ACE1 in *Trichoderm reesei* strains producing high and low levels of cellulase. *Eukaryotic Cell* **10**: 262-271.

Portnoy, T., Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sandor, E., Hartl, L., Karaffa, L., Druzhinina, I. S., Seiboth, B., Le Crom, S. and Kubicek, C. P. (2011) The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. *BMC Genomics* **12**: 269-281.

Prescott, E. M. and Proudfoot, N. J. (2002) Transcriptional collision between convergent genes in budding yeast. *PNAS* **99**: 8796-8801.

Pucher, M. E., Steiger, M. G., Mach, R. L. and Mach-Aigner, A. R. (2011) A modified expression of the major hydrolase activator in *Hypocrea jecorina* (*Trichoderma reesei*) changes enzymatic catalysis of biopolymer degradation. *Catalysis Today* **167**: 122-128.

Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J.-C. N., Johansen, K. S., Krogh, K. B. R.M., Jørgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J. and Walton, P. H. (2011) Insights into the oxidative degradation of cellulose by copper metalloenzyme that exploits biomass components. *PNAS* **108**: 15079-15084.

Rahman, Z., Shida, Y., Furukawa, T., Suzuki, Y., Okada, H., Ogasawara, W. and Morikawa, Y. (2009) Application of *Trichoderma reesei* cellulase and xylanase promoters through homologous recombination for enhanced production of extracellular β-glucosidase I. *Bioscience, Biotechnology and Biochemistry* **73**: 1083-1089.

Rahman, Z., Shida, Y., Furukawa, T., Suzuki, Y., Okada, H., Ogasawara, W. and Morikawa, Y. (2009) Evaluation and characterisation of *Trichoderma*

reesei cellulase and xylanase promoters. *Applied Microbiology and Biotechnology* **82**: 899-908.

Rey, M., Ohno, S., Pintor-Toro, J.A., Llobell, A. and Benitez, T. (1998) Unexpected homology between inducible cell wall protein QID74 of filamentous fungi and BR3 salivary protein of the insect *Chironomus*. *Proceedings of the National Academy of Sciences* **95**: 6212-6216.

Reyes-Dominguez, Y., Narendja, F., Berger, H., Gallmetzer, A., Fernandez-Martin, R., García, I., Scazzocchio, C. and Strauss, J. (2008) Nucleosome Positioning and Histone H3 Acetylation are independent processes in the *Aspergillus nidulans prnD-prnB* bidirectional promoter. *Eukaryotic Cell* **7**: 656-663.

Saloheimo, A., Aro, N., Ilmén, M. and Penttilä, M. (2000) Isolation of the *ace1* gene encoding a Cys$_2$-His$_2$ transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei*. *The Journal of Biological Chemistry* **275**: 5817-5825.

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

Saloheimo, M. Valkonen, M. and Penttilä, M. (2003) Activation mechanisms of the HAC1-mediated unfolded protein response in filamentous fungi. *Molecular Microbiology* **47**: 1149-1161.

Scazzocchio, C. and Ramón, A. (2008) The Aspergilli: Genomics, Medical Aspects, Biotechnology and Research Methods. Chapter 19: Chromatin in the Genus *Aspergillus*. Editors: Gustavo H. Goldman and Stephen A. Osmani. CRC Press Taylor and Francis Group.

Schmoll, M. (2008) The information highways of a biotechnological workhorse – signal transduction in *Hypocrea jecorina*. *BMC Genomics* **9**: 430-455.

Schmoll, M., Franchi, L. and Kubicek, C. P. (2005) Envoy, a PAS/LOV domain protein of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light. *Eukaryotic Cell* **4**: 1998-2007.

Schmoll, M., Schuster, A., Do Nascimento Silva, R. and Kubicek, C. P. (2009) The G-alpha protein GNA3 of *Hypocrea jecorina* (Anamorph

Trichoderma reesei) regulates cellulase gene expression in the presence of light. *Eukaryotic Cell* **8**: 410-420.

Schuster, A., Kubicek, C. P., Friedl, M. A., Druzhinina, I. and Schmoll, M. (2007) Impact of light on *Hypocrea jecorina* and the multiple cellular roles of ENVOY in this process. *BMC Genomics* **8**: 449-466.

Schuster, A., Kubicek, C. P. and Schmoll, M. (2011) Dehydrogenase GRD1 represents a novel component of the cellulase regulon in *Trichoderma reesei* (*Hypocrea jecorina*). *Applied and Environmental Microbiology* **77**: 4553-4563.

Schuster, A. and Schmoll, M. (2010) Biology and Biotechnology of *Trichoderma*. *Applied Microbiology Biotechnology* **87**: 787-799.

Seiboth, B., Hakula, S., Mach, R. L., Suominen, P. L. and Kubicek, C. P. (1997) Role of four major cellulases in triggering of cellulase gene expression by cellulose in *Trichoderma reesei*. *Journal of Bacteriology* **179**: 5318-5320.

Seiboth, B. and Metz, B. (2011) Fungal arabinan and L-arabinose metabolism. *Applied Microbial and Biotechnology* **89**: 1665-1673.

Seiboth, B., Pakdaman, B. S., Hartl, L. and Kubicek, C. P. (2007) Lactose Metabolism in filamentous fungi: how to deal with an unknown substrate. *Fungal Biology Reviews* **21**: 42-48.

Seidl, V., Gamauf, C., Druzhinina, I. S., Seiboth, B., Hartl, L. and Kubicek, C. P. (2008) The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. *BMC Genomics* **9**: 327-342.

Seidl, V., Seibel, C., Kubicek, C. P. and Schmoll, M. (2009) Sexual development in the industrial workhorse *Trichoderma reesei*. *PNAS* **106**: 13909-13914.

Seidl-Seiboth, V., Gruber, S., Sezerman, U., Schwecke, T., Albayrak, A., Neuhof, T., Von Doehren, H., Baker, S. E. and Kubicek, C. (2011) Novel hydrophobins from *Trichoderma* define a new hydrophobin subclass: protein properties, evolution, regulation and processing. *Journal of Molecular Evolution* **72**: 339-351.

Sekinger, E. A., Moqtaderi, Z. and Struhl, K. (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for

preferential accessibility of promoter regions in yeast. *Molecular Cell* **18**: 735-748.

Stals, I., Samyn, B., Sergeant, K., White, T., Hoorelbeke, K., Coorevits, A., Devreese, B., Claeyssens, M. and Piens, K. (2010) Identification of a gene coding for a deglycosylating enzyme in *Hypocrea jecorina*. *FEMS Microbiology Letters* **303**: 9-17.

Steiger, M. G., Vitikainen, M., Uskonen, P., Brunner, K., Adam, G., Pakula, T., Penttilä, M., Saloheimo, M., Mach, R. L. and Mach-Aigner, A. R. (2011) Transformation System for *Hypocrea jecorina* (*Trichoderma reesei*) that favours homologous integration and employs reusable bidirectionally selectable markers. *Applied and Environmental Microbiology* **77**: 114-121.

Stephanopoulos, G. (2007) Challenges in engineering microbes for Biofuels production. *Science* **315**: 801-804.

Steyaert, J. M., Weld, R. J., Mendoza-Mendoza, A. and Stewart, A. (2010) Reproduction without sex: conidiation in the filamentous fungus *Trichoderma*. *Microbiology* **156**: 2887-2900.

Strauss, J. and Reyes-Dominguez, Y. (2011) Regulation of secondary metabolism by chromatin structure and epigenetic codes. *Fungal Genetics and Biology* **48**: 62-69.

Stricker, A. R., Grosstessner-Hain, K., Wurleitner, E. and Mach, R. L. (2006) XYR1 (Xylanase Regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryotic Cell* **5**: 2128-2137.

Stricker, A. R., Mach, R. L. and H. de Graaff, L. (2008) Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Applied Microbiology and Biotechnology* **78**: 211-220.

Stricker, A. R., Steiger, M. G. and Mach, R. L. (2007) XYR1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS Letters* **581**: 3915-3920.

Stricker, A. R., Trefflinger, P., Aro, N., Penttilä, M. and Mach, R. L. (2008) Role of ACE2 (activator of cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. *Fungal Genetics and Biology* **45**: 436-445.

Sun, J., Tian, C., Diamond, S. and Glass, L. N. (2012) Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. *Eukaryotic Cell* **11**: 482-493.

Sunde, M., Kwan, A. H. Y., Templeton, M. D., Beever, R. E. and Mackay, J. P. (2008) Structural analysis of hydrophobins. *Micron* **39**: 773-784.

Suto, M. and Tomita, F. (2001) Review: Induction and Catabolite Repression Mechanisms of Cellulase in Fungi. *Journal of Bioscience and Bioengineering* **92**: 305-311.

Takashima, S., Iikura, H., Nakamura, A., Masaki, H. and Uozumi, T. (1996) Analysis of CRE1 binding sites in the *Trichoderma reesei cbh1* upstream region. *FEMS Microbiology Letters* **145**: 361-366.

Tisch, D., Kubicek, C. P. and Schmoll, M. (2011) New insights into the mechanism of light modulated signalling by heterotrimeric G-proteins: ENVOY acts on *gna1* and *gna3* and adjusts cAMP levels in *Trichoderma reesei* (*Hypocrea jecorina*). *Fungal Genetics and Biology* **48**: 631-640.

Todd, R. B. and Andrianopoulos, A. (1997) Evolution of a fungal regulatory gene family: The Zn(II)2Cys6 Binuclear Cluster DNA Binding Motif. *Fungal Genetics and Biology* **21**: 388-405.

Tolkunov, D., Zawadzki, K. A., Singer, C., Elfving, N., Morozov, A. V. and Broach, J. R. (2001) Chromatin remodelers clear nucleosomes from intrinsically unfavourable sites to establish nucleosome-depleted regions at promoters. *Molecular Biology of the Cell* **22**: 2106-2118.

Torigoi, E., Henrique-Silva, F., Escobar-Vera, J., Carle-Urioste, J. C., Crivellaro, O., El-Dorry, H. and El-Gogary, S. (1996) Mutants of *Trichoderma reesei* are defective in cellulose induction but not basal expression of cellulase-encoding genes. *Gene* **176**: 199-203.

Travers, A., Hiriart, E., Churcher, M., Caserta, M. and Di Mauro, E. (2010) The DNA sequence-dependence of nucleosome positioning *in vivo* and *in vitro*. *Journal of Biomolecular Structure and Dynamics* **27**: 713-723.

Tribus, M., Bauer, I., Galehr, J., Rieser, G., Trojer, P., Brosch, G., Loidl, P., Haas, H. and Graessle, S. (2010) A novel motif in fungal class 1 histone deacetylases is essential for growth and development of *Aspergillus*. *Molecular Biology of the Cell* **21**: 345-353.

Trifonov, E. N. (2010) Nucleosome positioning by sequence, state of the art and apparent finale. *Journal of Biomolecular Structure and Dynamics* **27**: 741-746.

Trojer, P., Dangl, M., Bauer, I., Graessle, S., Loidl, P. and Brosch, G. (2004) Histone methyltransferases in *Aspergillus nidulans*: Evidence for a novel enzyme with a unique substrate specificity. *Biochemistry* **43**: 10834-10843.

Verbeke, J., Coutinho, P., Mathis, H., Quenot, A., Record, E., Asther, M. and Heiss-Blanquet, S. (2009) Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*. *Biotechnology Letters* **31**: 1399-1405.

Vitikainen, M., Arvas, M., Pakula, T., Oja, M., Penttilä, M. and Saloheimo M. (2010) Array comparative genomic hybridisation analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. *BMC Genomics* **11**: 441-457.

Wang, L., Feng, Z., Wang, X., Wang, X. and Zhang, X. (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**: 136-138.

Weber, C., Farwick, A., Benisch, F., Brat, D., Dietz, H., Subtil, T. and Boles, E. (2010) Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Applied Microbiology and Biotechnology* **87**: 1303-1315.

Wendland, J. (2001) Comparison of morphogenetic networks of filamentous fungi and yeast. *Fungal Genetics and Biology* **34**: 63-82.

Werner, M. and Ruthenburg, A. J. (2011) The United States of histone ubiquitylation and methylation. *Molecular Cell* **43**: 5-7.

Werner, A. and Swan, D. (2010) What are natural antisense transcripts good for? *Biochemical Society Transactions* **38**: 1144-1149.

Wilson, R. A. and Talbot, N. J. (2009) Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews* **7**: 185-195.

Wolfenden, R., Lu, X. and Young, G. (1998) Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society,* **120**: 6814-6815.

Xu, F., Ding, H. and Tejirian, A. (2009) Detrimental effect of cellulose oxidation on cellulose hydrolysis by cellulase. *Enzyme and Microbial Technology* **45**: 203-209.

Xu, Z., Wei, W., Gagneur, J., Clauder-Muenster, S., Huber, W. and Steinmetz, L. M. (2011) Antisense expression increases gene expression variability and locus interdependency. *Molecular Systems Biology* **7**: 468-4677.

Xu, F. and Olsen, W. K. (2010) DNA architecture, deformability and nucleosome positioning. *Journal of Biomolecular Structure and Dynamics* **27**: 725-740.

Yassour, M., Pfiffner, J., Levin, J. Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D.-A., Friedman, N. and Regev, A. (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biology* **11**: 87-101.

Yu, Jae-Hyuk (2006) Heterotrimeric G-protein signalling and RGSs in *Aspergillus nidulans*. *The Journal of Microbiology* **44**: 145-154.

Zeilinger, S., Schmoll, M., Pail, M., Mach, R. L. and Kubicek, C. P. (2003) Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Molecular Genetics and Genomics* **270**: 46-55.

Zhang, J., Siika-aho, M., Tenkanen, M. and Viikari, L. (2012) The role of acetyl xylan esterase in the solubilisation of xylan and enzymatic hydrolysis of wheat straw and giant reed. *Biotechnology for Biofuels* **4:** 60-69.

Zhong, L, Matthews, J. F., Hansen, P. I., Crowley, M. F., Cleary, J. M., Walker, R. C., Nimlos, M. R., Brooks III, C. L., Adney, W. S., Himmel, M. E. and Brady, J. W. (2009) Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose Iβ: the enzyme-substrate complex. *Carbohydrate Research* **344**: 1984-1992.

Zhong, Y. H., Wang, T. H., Wang, X. L., Zhang, G. T. and Yu, H. N. (2009) Identification and characterisation of a novel gene, TrCCD1, and its possible function in hyphal growth and conidiospore development of *Trichoderma reesei*. *Fungal Genetics and Biology* **46**: 255-263.

Zhou, Q., Zu, J., Kou, Y., Lv, X., Zhang, X., Zhao, G., Zhang, W., Chen, G. and Liu, W. (2012) Differential involvement of β-glucosidases from Hypocrea jecorina in rapid induction of cellulase genes by cellulose and cellobiose. *Eukaryotic Cell* **11:**1371-1381.

Zhu, J. Y., Pan, X. and Zalesny Jr., R. S. (2010) Pre-treatment of woody biomass for biofuel production: energy efficiency, technologies and recalcitrance. *Applied Microbiology and Biotechnology* **87**: 847-857.

Znameroski, E. A., Croadetti, S. T., Riche, C. M., Tsai, J. C., Iavarone, A. T., Cate, J. H. D. and Glass, L. N. (2012) Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellodextrins. *PNAS* **109**: 6012-6017.