Chapter 1 Introduction

T. reesei is currently exploited for the large-scale production of proteins applicable to industrial processes. The number of proteins of interest is increasing and expanding into areas such as biopharmaceuticals. As a consequence, the *T. reesei* expression system requires further optimisation and diversification to cater for the specific requirements of each commercially-relevant protein. In particular, there is a need to increase the available choice of promoters to include constitutively active promoters and promoters which are not repressed by glucose. The following sections provide the background on *T. reesei* as a production host and outlines the discovery of the *hex1* gene and the potential of this promoter to drive heterologous protein expression in *T. reesei.*

1.1 The biology of filamentous fungi

All fungi possess a heterotrophic nutritional lifestyle and are dependent on the absorption of nutrients from the immediate surrounding environment for survival. The nutritional demands of fungi may be met by the decomposition of dead organic material or parasitism on other living organisms. In either situation, in the natural environment, fungi must secrete hydrolytic enzymes into the extracellular environment to convert complex compounds to simple nutrient molecules that are accessible for absorption. This innate ability of fungi to secrete enzymes makes them valuable industrial protein producers.

The growth of the cylindrical hyphae of filamentous fungi is by unidirectional extension from the hyphal tip. The cell wall at the tip is plastic due to the constant deposition and remodeling of the polysaccharides and becomes progressively rigid in

the subapical areas and lateral walls due to the covalent crosslinking of chitin and □(1□ 3) glucans. In the subapical region, compartmentalisation of the hyphae occurs by the formation of a septum, which coincides with the cessation of cell growth and nuclear division within that cell. The cells of the hyphae are typically multinucleate and the septal pore enables cytoplasmic continuity and organelle movement between neighbouring cells. In addition to the growth by the extension of the hyphal tips, branches develop from lateral walls to expand the area colonised by the mycelium and possibly enables it to access more nutrients from the environment. The extent of branching of a mycelium is affected by growth media and conditions, however the molecular mechanisms which determine this morphology are yet to be elucidated (Wongwicharn *et al*., 1999; Alcantara-Sanchez *et al*., 2004; Gatherar *et al*., 2004; Riquelme *et al*., 2004; reviewed by Yarden, 2004).

1.1.1 Fungi as protein producers

Fungi possess the elements of the cellular machinery necessary for efficient protein expression and secretion, and they can be grown on cheap undefined media making large-scale protein production economically feasible. Therefore, fungi continue to be investigated for their development as protein expression systems. Fungal systems are also advantageous over prokaryotic systems because fungi possess eukaryotic glycosylation and post-translational modification characteristics, which are especially important if mammalian proteins are to be expressed.

1.1.2 Enzymes in industrial processes

The use of enzymes as biocatalysts in numerous industrial processes is a $2 billion global industry (Nevalainen *et al*., 2003), which necessitates the development of

efficient protein production systems. The dominant areas of industry where biocatalysts are used are those involving detergents (34%), textiles (11%), starch (12%), baking (5%), animal feed (7%), beverages and brewing (7%) and dairy processes (14%) (Saxena *et al.*, 2004). The main enzymes exploited are amylases, proteases, lipases, xylanases, tannase and phytase, and are marketed under product names such as Vinozyme®, Scourzyme®, Denimax®, Plusweet® and Purfect®. The various enzymes available and their applications are shown in Table 1-1.

Enzymes have traditionally been used in food production, brewing, leather and linen manufacturing. The enzymes have been sourced from nature as crude extracts of gastric juices of slaughtered young animals (for rennet in cheese making) or from microorganisms such as *Saccharomyces sake* and *Aspergillus oryzae* (for the production of sake from the fermentation of rice). The development of fermentation technology has enabled the large-scale production of enzymes by mass culturing of the relevant microorganism. Recently, with developments in recombinant gene technology, the range of enzymes available has increased and the manufacturing processes can be further optimised. Enzymes no longer need to be produced by the native host as a range of heterologous expression systems are now available. *Trichoderma reesei* is one of the microorganisms continuing to be developed for this purpose.

1.1.3 The genus *Trichoderma*

The genus *Trichoderma,* has particular economic and industrial importance. *Trichoderma* species are represented on all continents and exist in a range of climatic zones. This genus is predominantly saprophytic but some species are capable of mycoparasitism (Klein *et al.*, 1998).

Table 1-1. Commercially available enzymes and their industrial applications. This list was compiled from the product catalogues from three of the world-leading enzyme producing companies Novozyme (www.novozyme.com), Genencor (www.genencor.com) and DSM (www.dsm.com).

Industry	Application	Enzyme
Detergents	Removal of protein, starch, fat and oil stains	Proteases, lipases, amylases, cellulases
Textiles	Desizing, increased wettability and even dyeing of fabric, bio-polishing, softening of wool and silk	Amylase, pectinase, cellulase, catalase, protease
Leather	Hair removal and cleaning of pelt, bating, increase softness	Lipase, protease, elastase
Pulp and paper	Removal of lignin and reduction of pitch for increased whiteness	Xylanase, lipase
	Smooth starch coating of paper	Amylase
Agricultural animal feeds	Improved feed utilisation through increased digestibility and accessibility to minerals	Phytase, glucanase, amylase, protease, xylanase, pectinase, cellulase
Ethanol fuel and brewing	Release of protein from grain mash, liquification, saccharification, shortened maturation time for beer	Protease, amylase, amyloglucosidase, α-acetolactate decarboxylase
Sweetener production	Saccharification, liquification, increased sugar yields from cane, maltodextrin production from starch	Glucoamylase, α-amylase, dextranase, pullulanase, glucose isomerase
Baking	Anti-staling, increased volume, even crumb structure, attractive crusts	Amylases
	Increased dough strength and elasticity, dough conditioning and bleaching	Glucose oxidase, lipase, lipoxygenase, xylanase, protease
Dairy	Hydrolysis of lactose, milk formula preparation and piquancy in cheese	β-galactosidase, lipase, protease
Food	Fruit maceration, juice extraction and clarification, vegetable oil extraction	Pectinesterase, polygalacturonase, pectin lyase
	Hydrolysis of meat by-products, production of flavour enhancers	Protease

Some species, such as *T. harzianum*, are suitable biocontrol agents against fungal and bacterial disease on agricultural crops. During the attack on other fungi, chitinases, glucanases and proteases are secreted to destroy the cell wall of the host (Chet *et al*., 1998). Biocontrol of pathogens is also effective due to the secretion of antibiotic compounds such as alkyl pyrones, isonitriles, polyketides, petaibols, diketopiperazines, sesquiterpenes and various steroids (Howell, 1998).

Trichoderma reesei degrades plant material in soil by secreting a barrage of cellulolytic and hemicellulolytic enzymes which act synergistically in the hydrolysis of cellulose and hemicellulose to oligosaccharides and monomeric sugars. *T. reesei* was first recognised as a high secretor of cellulolytic enzymes during World War II where it was identified as the culprit organism responsible for causing rapid degradation of US military canvas tents and cotton clothing in Bougainville (Linko, 1993). Since the isolation of *T. reesei*, hyperproducers of cellulolytic enzymes have been developed by classical mutagenesis and screening (Montenecourt *et al*., 1979; Bailey *et al*., 1981). Improved strains are capable of secreting up to 40 g/L of protein into the growth medium (Durand *et al*., 1988; Pakula *et al*., 2000) making *T. reesei* one of the highest known protein-secreting organisms.

Due to the strong protein secretion potential of *T. reesei*, many of its enzymes are currently produced on a large scale for application to industrial processes. Commercially available *Trichoderma* enzyme products such as Crystalzyme, Econase, Stonenzyme, Cellubrix, Ecopulp and Pulpzyme (Penttilä *et al*., 2004) include some of the native hydrolytic enzymes.

1.2 The cellulolytic enzyme system of *Trichoderma reesei*

Since the discovery of *T. reesei,* there has been intensive investigation into the strong cellulolytic enzyme system (summarised in Table 1-2). Initially, this involved the identification of various carbon sources that either enhanced or repressed cellulase gene expression (reviewed by Penttilä *et al*., 1993). It is now clear that the genes encoding cellulases and hemicellulases are co-ordinately regulated and highly expressed in medium containing cellulose, cellobiose, hemicellulose, lactose, xylobiose and sophorose – the strongest known inducer of cellulase in *T. reesei.* The cellobiohydrolase 1 gene (*cbh1*) encodes the main secreted protein (CBHI enzyme), for which expression is upregulated several thousand fold on induction and can yield up to 24 g/L (Ilmén *et al*., 1996a). Alternatively, when cultures are grown on readily metabolised carbon sources such as glucose, cellulase transcripts are undetectable (Ilmén *et al*., 1997), as it is unnecessary for the fungus to expend energy in producing and secreting large quantities of enzymes. Glucose is an active repressor of cellulase expression by the mechanism of carbon repression. In order for the cellulolytic and hemicellulolytic enzymes to function in synergy in the hydrolysis of complex molecules, their regulation is multilayered and highly complex. Some of the factors involved in the regulation of transcription are described in Section 1.5.8.

1.2.1 Heterologous protein production in *Trichoderma reesei*

In addition to the production of enzymes for industrial processes, *Trichoderma* is being developed as a heterologous protein expression host for any proteins of interest,

Table 1-2. The extracellular cellulolytic enzymes of *Trichoderma reesei* known to date. The enzyme names are given by the original protein name (eg. CBHI) as well as the name designated by the nomenclature developed by Henrissat *et al*. (1998). The name of the preferred substrate eg. cellulase, is abbreviated and the family to which the enzyme belongs is denoted as a number eg. CEL7. If an organism produces multiple enzymes of the same family, a letter denotes the particular family member eg. CEL7A. (Table adapted from Miettinen-Oinonen, 2004).

Enzyme	**Gene**	**GenBank accession no.**	**Reference**
CEL7A (CBHI)	*cbh1/cel7a*	P00725	Shoemaker *et al*. (1983), Teeri *et al*. (1983)
CEL6A (CBHII)	*cbh2/cel6a*	M16190	Teeri *et al*. (1987)
CEL7B (EGI)	*egl1/cel7b*	M15665	Penttilä *et al*. (1986)
CEL5A (EGII)	*egl2/cel5a*	M19373	Saloheimo *et al*. (1988)
CEL12A (EGIII)	*egl3/cel12a*	AB003694	Ward *et al*. (1993), Okada *et al*. (1998)
CEL61A (EGIV)	*egl4/cel61a*	Y11113	Saloheimo *et al*. (1997)
CEL45A (EGV)	*egl5/cel45a*	Z33381	Saloheimo *et al*. (1994)
EGVI	not determined		Bower *et al*. (1998)
CEL74A (EG)	*cel74a*	AY281371	Foreman *et al*. (2003)
CEL61B (EG)	*cel61b*	AY281372	Foreman *et al*. (2003)
CEL5B (EG)	*cel5b*	AY281373	Foreman *et al*. (2003)
CEL3A (BGLI)	*bgl1/cel3a*	U09580	Barnett *et al*. (1991), Mach *et al*. (1993)
CEL1A (BGLII)	*bgl2/cel1a*	AB003110	Takashima *et al*. (1999), Saloheimo *et al*. (2002)
CEL3B (BGL)	*cel3b*	AY281374	Foreman *et al*. (2003)
CEL3C (BGL)	*cel3c*	AY281375	Foreman *et al*. (2003)
CEL1B (BGL)	*cel1b*	AY281377	Foreman *et al*. (2003)
CEL3D (BGL)	*cel3d*	AY281378	Foreman *et al*. (2003)
CEL3E (BGL)	*cel3e*	AY281379	Foreman *et al*. (2003)

including the cheap production of pharmaceuticals which are conventionally produced by mammalian cell culture. Recently, the potential of *T. reesei* for the production of antibodies has been investigated (Joosten *et al*., 2003). These antibodies are not only targeted for the pharmaceutical industry but are to be incorporated into a whole range of products such as shampoos and skin care creams for targeting the active ingredients to the specific intended surface, such as hair but not the scalp (www.genencor.com).

The large scale cultures of industrially important strains of fungi and the fungal products themselves are considered to be safe with respect to humans. In Europe *T. reesei* is considered a Group 1 low risk microorganism (EC Directive 90/219/EEC) (Nevalainen *et al*., 1998) and many *Trichoderma*-sourced products are approved from the relevant local authorities .

In order to optimise *T. reesei* as a heterologous protein producer, a greater understanding of the basic fungal biology and the cellular elements involved from transcription to the secretion of a functional protein are necessary.

1.3 A review of the secretory pathway

In order to manufacture homologous or heterologous gene products efficiently in fungal hosts, the proteins must be successfully transported through the secretory pathway. The extracellular proteins are passed through a series of intracellular compartments (Figure 1-1) for various steps of protein modification and folding which are outlined below.

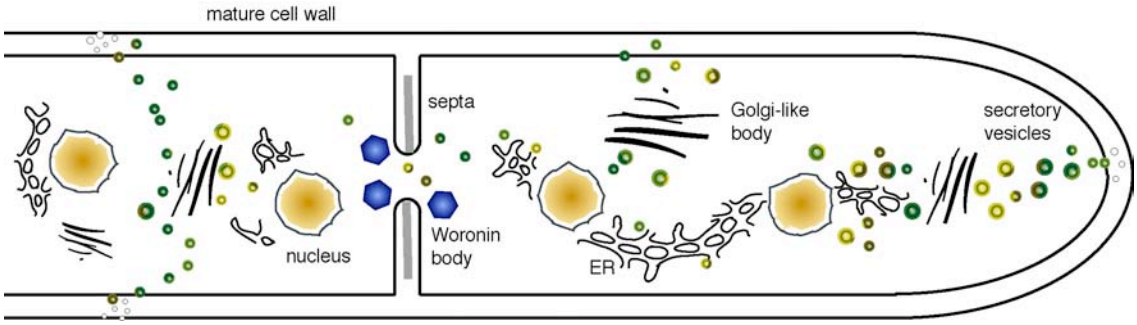

Figure 1-1. A hyphal apex showing an overview of the secretory pathway. See text for details.

Much of the understanding of the secretory pathway in filamentous fungi has been generated from studies in yeast and mammalian neuronal cells. However, the mammalian and yeast models are not entirely applicable due to two features signatory of filamentous fungi: the growth of the mycelium as long branching hyphae, and the ability to secrete comparatively copious quantities of protein into the growth medium (reviewed in Raikhel *et al*., 2000; van Vliet *et al*., 2003).

1.3.1 The passage of protein through a cell

The proteins expressed within a cell that are destined for secretion into the surrounding medium are identified by a secretion signal peptide on the N-terminus. The secretion signal peptide (ss) contains several charged amino acids followed by a hydrophobic core, as demonstrated by the ss of the main cellobiohydrolase of *T. reesei,* MYRKLAVISAFLATARA. The ss of a particular protein can be interchangeable between unlike proteins or between different organisms. For example, the *cbh1* ss has successfully targeted heterologous proteins for secretion from *T. reesei* (Paloheimo *et al*., 1993; Penttilä, 1998; Te'o *et al*., 2000), and the fusion of the ss from the secreted viral K28 protein toxin to the N-terminus of green fluorescent protein (GFP) enabled the secretion of GFP from the yeasts *Candida glabrata, Pichia pastoris, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Eiden-Plach *et al*., 2004).

As a mRNA is translated on a free ribosome, the emerging ss is recognised by the cytosolic signal recognition particle (SRP), a ribonucleoprotein complex, which pauses further translation until the ribosome is transferred to the ER membrane. In yeast, the SRP docks the ribosome and emerging protein chain with the SRP receptor on the ER membrane and the translocation complex composed of the proteins Sec66p, Sec67p and

Sec63p (Deshaises *et al.*, 1991; Feldheim *et al.*, 1993), before being released. The Sec63p protein of the translocation complex spans the ER membrane and is in direct association with the chaperone, immunoglobulin binding protein (BiP), in the ER lumen (Brodsky *et al.*, 1993). On the transfer of the nascent polypeptide chain and attached ribosome to the Sec61p ER membrane channel, co-translational translocation into the ER lumen occurs (Pilon *et al.*, 1998). ATP is hydrolysed in the process and BiP is released into the ER lumen (Brodsky and Schekman, 1993). The ss on the immature protein is cleaved by a signal peptidase upon entering the ER lumen.

Within the ER lumen, chaperones and foldases assist the nascent protein to achieve the proper tertiary or quaternary conformation (Chevet *et al.*, 2001). Two of these chaperones, BiP and protein disulfide-isomerase (PDI) have been characterised in *T. reesei* (Saloheimo *et al.*, 1999) and *A. niger* (Ngiam *et al.*, 2000). BiP binds to hydrophobic regions of the partially folded protein to shield them and prevent the tendency to incorrectly fold upon itself or form aggregates (Gething, 1999). PDI breaks and reforms correct disulfide bonds between the sulfhydryl groups of cysteine residues (Freedman *et al.*, 1994). Proteins of incorrect conformation are retained in the ER by chaperones for correction or subjected to the ER-associated protein degradation pathway (ERAD). ERAD involves retrotransportation to the cytosol, ubiquination and degradation by the proteasome (reviewed by Spiro, 2004).

The overexpression of protein can create stress within the ER, which in turn activates the unfolded protein response (UPR). The UPR initiates a cascade of events that alter the expression pattern of numerous genes involved in the secretory pathway including BiP and PDI (Pakula *et al.*, 2003; Saloheimo *et al.*, 2003; Mulder *et al.*, 2004; Ogawa *et al.*, 2004; Penttilä and Limón, 2004). The effects of the UPR on protein production

continue to be investigated (Pakula *et al.*, 2003; Valkonen *et al.*, 2003; Al-Sheikh *et al.*, 2004).

From the ER proteins are selected and packaged into coatomer-protein II (COPII) coated vesicles which bud off and are transported to the Golgi via passage along microtubules (Phillipson *et al.*, 2001). In mammalian cells, the transport may be a two step process where the cargo is first transferred to coatomer-protein I (COPI) coated vesicles for transport to the Golgi (Stephens *et al.*, 2000). If there is a backlog in the flow of the secretory pathway, the vesicles fuse to form larger ER-Golgi intermediate compartments (ERGIC) where some further sorting occurs and ER-associated proteins can be recycled by retrograde transport.

On reaching the Golgi, the protein may undergo further folding, glycosylation and phosphorylation. The chaperones and enzymes involved in these processes are not uniformly dispersed but show distinct gradients of concentration, which are representative of the progressive stages of protein modification that occurs in the direction from the *cis* to the *trans* Golgi network (TGN). Proteins in the TGN have a number of destinations and the sorting of these proteins occurs in response to the specific signals they contain (Keller *et al.*, 1997).

The correct glycosylation of a protein within the ER and Golgi is important for its successful secretion and for protein functionality. Consequently, differences between mammalian and fungal glycosylation patterns present a challenge for the production of mammalian proteins in fungi. However, the basic core N-glycan structure in *T. reesei* is similar to the mammalian-like $Man_3GlcNAc_2$ core (reviewed by Nevalainen *et al.*,

2004a). This similarity offers the opportunity to build upon this common foundation and replicate mammalian glycosylation in fungi (Maras *et al*., 1997; Maras *et al*., 1999).

The Golgi and TGN are also the sites for Kex2-like endoprotease cleavage of protein proregions. In the absence of other signals, the default pathway for soluble proteins in the TGN lumen is for packaging and exocytosis from the cell (Raikhel and Chrispeels, 2000). In neurons, this pathway to secretion can be influenced by the proteins which are co-expressed and the levels at which they are expressed (Farhadi *et al*., 2000).

Whilst there is a wealth of information on membrane-associated proteins, membrane complexes and the carrier peptides involved in protein secretion in yeast and mammalian cells, the knowledge of the fundamental recognition mechanism upon specific fungal proteins is still lacking.

1.3.2 Secretion of protein from the mycelium into the extracellular matrix

The final stage in the secretion of a protein occurs via two alternative pathways. The constitutive secretory pathway is a result of a steady rate of packaging of proteins into vesicles from the TGN and the exocytosis of their contents. The regulated secretory pathway involves the packaging of extracellular proteins into vesicles that accumulate intracellularly until exocytosis is induced by an external stimulus.

Many of the details of this final translocation step in the secretory pathway are not clearly understood in filamentous fungi. Microtubules facilitate the translocation of secretory vesicles from the TGN to the cell membrane in mammalian cells (Schmoranzer *et al*., 2003) and actin filaments are involved in the transport of

glucoamylase to the cell wall in *A. niger* (Gordon *et al.*, 2000a). The Rho family of small GTPases have been implicated in vesicle trafficking and cell polarisation in yeast (Takai *et al.*, 2001) and the *rho3* gene of *T. reesei* is important for growth and total protein secretion (Vasara *et al.*, 2001). The involvement of the cytoskeleton indicates that the passage of secretory vesicles to hyphal tips is not via a passive mechanism.

1.3.2.1 The site of protein secretion from the mycelium

For a long time, the general consensus was that all secretion from the mycelium occurred only from the apical cells or subapical regions of hyphae (Wösten *et al.*, 1991). This concept was supported by the bulk flow of protoplasm toward the tips and the dynamic nature of the cell wall in this region. In addition, the distribution of a GFP:glucoamylase fusion protein in *A. niger* showed that secretion occurred dominantly from the hyphal apex, even though some fluorescence accumulated in the cell wall of the hyphae (Gordon *et al.*, 2000a). However, more recent data suggest that the site of secretion is variable depending on the protein. Immunohistochemical labelling for CBHI localised the protein throughout the cytoplasm and within the cell wall of mature hyphae, indicating that active secretion was also occurring away from the hyphal tips (Nykänen, 2002).

Evidence is also emerging to suggest that there is a spatial relationship between the region of transcription and the site of protein secretion from the cell. For example, the distribution of CBHI protein throughout the mycelium of *T. reesei* mostly mirrored the distribution of *cbh1* transcripts (Nykänen, 2002) and the distribution of barley endopeptidase B (EPB) protein and *epb* transcripts were co-localised to the apical and subapical compartments of hyphae at the edge of the colony (Nykänen *et al.*, 1997).

This co-localisation of transcripts and protein suggests that transcription and secretion occurs locally and that broader translocation of either transcript or protein through the mycelium for secretion may be unlikely to occur, except for a some transport of protein to the hyphal apex (Nykänen *et al.*, 1997).

Localised transcription and secretion implies differential promoter activity throughout the mycelium. For example, glucoamylase (GLA) is secreted by only some hyphae at the colony periphery in *A. niger* (Wösten, *et al.* 1991). The expression of GFP under the *A. niger glaA* promoter revealed that hyphal branches displayed significantly lower levels of fluorescence than the leading hyphae but expression of GFP under the *gpdA* promoter demonstrated a more even distribution of fluorescence across the mycelium. The authors suggested that the differential distribution of fluorescence was an indication of spatial differences in transcriptional activity of the *glaA* promoter across the mycelium, which resulted in the observed localised secretion of glucoamylase (Vink *et al.*, 2004).

The information at hand indicates that the mycelium can not be considered as a uniform mass, but more as a dynamic multicellular organism which displays spatial regulation of transcription and secretion throughout.

1.4 The effects of proteolysis on protein production

The degradation of heterologous proteins expressed in fungi is a major cause for the low yields observed. Degradation can occur intracellularly within lysosomes, within the ER or on secretion by membrane associated proteases, or extracellularly by secreted proteases. Studies into the expression of human interleukin 6 (hIL-6) in *A. awamori*

highlight various different proteolytic events that can occur in heterologous protein expression. When hIL-6 was expressed under the endogenous 1,4-β-endoxylanase A (*exlA*) promoter, fused to the *exlA* prepro sequence, the very low level of protein in the culture supernatant did not reflect the relatively high mRNA levels. Purified hIL-6 was shown to be stable in culture supernatants, indicating that extracellular proteases were not the cause of the low yield of secreted protein. In addition, intracellular degradation by the UPR was unlikely, as BiP expression was not upregulated (Gouka *et al.*, 1996). When hIL-6 was expressed as a fusion to glucoamylase joined by a Kex2 cleavage site, the ratio of glucoamylase to hIL-6 protein indicated that 60 to 90% of the hIL-6 must have been degraded intracellularly or by mycelium-associated proteases (Gouka *et al.*, 1997). The degradation of hIL-6 occurred intracellularly during or immediately on translation when expressed without the glucoamylase fusion partner or fused to the glucoamylase at the C-terminus (Gouka *et al.*, 1997). Therefore, proteolytic activity can have a major effect on protein production in fungi especially when concerning heterologous products.

The protease profile for *A. niger* has been partially characterised. These proteases include aspartyl-, serine-, and carboxy proteases (reviewed in van den Hombergh *et al.*, 1997) but the profile can vary according to the strain and growth medium. Similar proteases are also present in *T. reesei* (Bradford, 2000). Classical mutagenesis and screening has generated some protease-deficient strains of *A. niger* (Mattern *et al.*, 1992) and *T. reesei* (Mäntylä *et al.*, 1994) which have been beneficial for heterologous protein expression (Broekhuijsen *et al.*, 1993; de Faria *et al.*, 2002). However, the isolation of protease genes or genes involved in their regulation allows for the specific genetic manipulation and inactivation of these loci (reviewed in Berka *et al.*, 1990;

Nevalainen *et al.*, 2004a). For example, deletion of the gene encoding the aspartic proteinase Trichodermapepsin from *T. reesei* resulted in a 94% decrease in acid protease activity (Mäntylä *et al.*, 1998).

1.4.1 Prosequences in protein production

Some proteins, such as extracellular enzymes (Monod *et al.*, 2002), are translated with a propeptide, adjacent to the N-terminal ss or on the C-terminus, which is not present on the mature protein due to proteolytic cleavage in the secretory pathway. The propeptides of proteins serve different functions and have been described to assist in protein folding and protein stability or act as inhibitors of enzyme activity (Bzymek *et al.*, 2004; Zschenker *et al.*, 2004). The proregion may also contribute as a localisation signal which determines protein passage through the cell and may be vital for protein secretion (Keller and Simons, 1997).

Propeptides have been exploited for the production of heterologous proteins in fungi. For example, proinsulin-like molecules were efficiently secreted from *S. cerevisiae* when expressed as a translational fusion to the endogenous □-factor ss and propeptide (Kjeldsen, 2000). In order to increase the effectiveness of fusion of heterologous proteins to endogenous propeptides, more needs to be understood on the mechanisms of propeptide processing.

The identification of the endopeptidases responsible for the specific processing of the propeptides is important. A number of Kex2-like subtilisin-related serine proteases (SRSP) called the proprotein convertases (PCs) have been discovered within the mammalian secretion pathway. The SRSPs PC4, PACE4, PC5/PC6, PC7/8 and furin act

within the constitutive secretory pathway (Gensberg *et al*., 1998). These PCs are localised in the Golgi, TGN, secretory granules, the cell surface or are extracellular (Gensberg *et al*., 1998; Wouters *et al*., 1998). However, the endoproteinases themselves are subject to regulation by posttranslational modification, which determines their subcellular localisation and their state of activity by altering the endoproteinases' conformation (Wouters *et al*., 1998). The recognition sequences can be shared between different SRSPs making identification of specific SRSP activity difficult (Wouters *et al*., 1998). They can also be composed of neighbouring basic residues or residues up and downstream of the cleavage site (Wantanabe *et al*., 1992). The mammalian Kex2 homologue furin, is anchored into the membrane of the TGN and cleaves the precursor protein sequence Arg-X-Lys/Arg-Arg, in the constitutive secretory pathway (Nagahama *et al*., 1991; Wantanabe *et al*., 1992).

In fungi, the *A. niger* Kex2 homologue KexB cleaves after the typical Lys-Arg dibasic moiety and after the tetrapeptide Arg-Val-Arg-Arg (Jalving *et al*., 2000). In *T. reesei*, cleavage of the proprotein on the carboxy side of the amino acids Arg-Arg and Lys-Arg by a Kex2-like dibasic endopeptidase is vital for secretion of the proteins XYNI and XYNII (Goller *et al*., 1998). *T. reesei* may also possess other endoproteolytic proprotein processing enzymes in addition to the Kex2-like dibasic endopeptidase (Goller *et al*., 1998) but these remain uncharacterised to date.

1.5 Fungal promoters - the driving force for protein production

A high level of gene transcription is generally the first step towards high levels of protein expression. This transcription is driven by elements within the gene promoter

that dictate whether the activity is broadly classified as inducible or constitutive. Inducible promoters commence transcription in response to a specific inducing element that may be representative of an environmental stimulus, such as temperature or photoperiod, a particular developmental or growth phase, or in response to the presence or absence of certain metabolites. Constitutive promoters are considered to be constantly active at a consistent intensity and are generally required in genes that have a continuous "housekeeping" role in the maintenance of cellular homeostasis.

There are three promoters used for the expression of protein in filamentous fungi that dominate in the literature. These are the inducible *T. reesei* cellobiohydrolase I (*cbh1*) gene promoter, the inducible *A. awamori* glucoamylase A (*glaA*) gene promoter and the constitutive *A. nidulans* glyceraldehyde-3-phosphate dehydrogenase (*pgk1*) gene promoter. These promoters are discussed below with the constitutive promoters for the pyruvate kinase (*pki1*) and the alcohol dehydrogenase (*alcA*) genes, to illustrate the current choices of fungal promoters applicable to enhanced protein production.

1.5.1 The *cbh1* gene promoter of *Trichoderma reesei*

The *T. reesei cbh1* promoter is one of the strongest inducible promoters known in fungi and consequently, the promoter is used extensively in the production of industrial enzymes. Some of the proteins which have been expressed under the *cbh1* promoter and are described in the literature, are shown in Table 1-3. In conjunction with the overexpression of endogenous genes, the *cbh1* promoter has been used for the expression of heterologous genes with varying degrees of success. Expression of proteins fused to the CBHI core has been used to improve extracellular yields of foreign proteins such as human antibody fragments by 150 fold and of calf chymosin by 5 fold

Table 1-3. Examples of proteins which have been expressed under the *cbh1* promoter in *T. reesei*.

Protein	Gene source	Yield	Method	Reference
Hydrophobin I Hydrophobin II	*T. reesei*	0.5 g/L $_f$ 240 mg/L $_f$	Coding sequence directly under *cbh1* promoter	Bailey *et al*. (2002)
Xylanase	*Actinomadura flexuosa* (bacterium)	ND	Fusion to mannanase I core and hinge region	Paloheimo *et al*. (1998)
Thermophilic xylanase B	*Dictyoglomus thermophilum* (bacterium)	100 mg/L$_s$	Fusion to *cbh1* ss	Te'o *et al*. (2000)
Thermophilic xylanase II	*Humicola grisea* var. *thermoidea* (filamentous fungus)	0.5 g/L$_s$	Fusion to *cbh1* ss	de Faria *et al*. (2002)
Calf chymosin	bovine	20 mg/L $_s$ 100 mg/L $_s$	Fusion to *cbh1* ss Fusion to *cbh1* core-linker	Harkki *et al*. (1989) Nykänen *et al*. (2002)
Interleukin-6	human	5 mg/L$_s$	Fusion to *cbh1* core-linker	Demolder *et al*. (1994)
Fab antibody fragments	mouse	1 mg/L $_s$ 40 mg/L $_s$ 150 mg/L $_f$	40 mg/L and 150 mg/L yields from fusion to *cbh1* core-linker	Nyyssönen *et al*. (1993)

$_s$ yield from shake flask culture, $_f$ yield from fermenter cultivation, ND not determined

(Nykänen, 2002), but the fusion strategy was not beneficial in the expression of a bacterial thermophilic xylanase enzyme, *xynB* (Te'o *et al*., 2000).

In general, the *cbh1* promoter is strongly induced by sophorose and cellulose. However, when grown on crystalline cellulose, *cbh1* induction is dependent upon secondary products formed by the activities of CBHII and EGII (Seiboth *et al*., 1997). Alternatively, the *cbh1* promoter may also be induced by the soluble carbon source, lactose, independently of the activity of other cellulases. Evidence suggests that the *cbh1* promoter has more than one mechanism of induction (Aro *et al*., 2003) and one of these mechanisms involves the transcription enhancer ACEII, as described in Section 1.5.8.3 (Aro *et al*., 2001).

The *cbh1* promoter is strongly repressed by glucose via the action of the transcriptional repressor, CREI (Section 1.5.8.2). Glucose repression has been relieved by mutation of a single CREI binding site in the *cbh1* promoter (Ilmén *et al*., 1996a). However, an additional eight CREI binding sites were discovered in the *cbh1* promoter using electrophoretic mobility shift assays and DNase I footprinting assays (Takashima *et al*., 1996) which suggests that the multiple CREI binding sites have an additional role in the *cbh1* promoter.

1.5.2 The glucoamylase A gene promoter of *Aspergillus niger* var. *awamori*

Glucoamylase (GLA) is an extracellular enzyme produced by *A. niger* to hydrolyse starch into smaller oligosaccharides and glucose. The promoter of the glucoamylase A gene (*glaA*) of *A. niger* is strongly induced by starch and repressed by xylose (Nunberg

et al., 1984). Regulation of GLA expression occurs at the level of transcription by catabolite repression (Nunberg *et al.*, 1984; Fowler *et al.*, 1990).

Two forms of GLA protein can be produced from the single gene by the differential splicing of a 169 bp intron from the 3' region of the primary transcript (Boel *et al.*, 1984). The *tsp* of the *glaA* gene is within –52 to –73 from the start codon and only a further 140 bp is required for transcription initiation (Nunberg *et al.*, 1984; Fowler *et al.*, 1990). However, the binding regions for trans-acting regulatory proteins that direct high-level expression, have been mapped to the region –318 and –800 along the *glaA* gene promoter (Fowler *et al.*, 1990; Verdoes *et al.*, 1994). Within this region is a CCAAT-box that enhances transcription of the gene. This enhancement was demonstrated when multiple CCAAT-boxes were introduced into the promoter (Liu *et al.*, 2003).

The *glaA* gene promoter is the dominant promoter used for industrial enzyme production in *A. niger*. When the promoter is used in high protein-secreting strains, they are capable of producing greater than 30 g/L of GLA (Ward, 1989). Glucoamylase driven expression systems are used for the industrial production of enzymes such as arabinofuranosidase, amino-peptidase, glucoamylase, pectinases and proteases for the food industry and phytases and xylanases for animal feed production (www.dsm.com). Improved protein production under the *glaA* promoter has been achieved by using a translational fusion of the glucoamylase catalytic domain to the N-terminus of the heterologous gene via a short linker. This has become a common approach to protein production under *glaA* in *Aspergillus*.

Recently, high levels of a functional IgG antibody were expressed in *A. niger* under the *glaA* promoter (Ward *et al.*, 2004). The expression cassette designed encoded a fusion protein composed of the GLA ss, proregion and the catalytic and linker regions fused to the heavy or light chain by a triglycine linker and a KexB cleavage site. Using this expression cassette, the fusion protein was correctly processed through the secretory pathway such that 100% of the secreted heavy and light chains were cleaved from the GLA fusion. The success of this study may pave the way for further positive outcomes in the secretion of mammalian proteins in fungi.

1.5.3 The glyceraldehyde-3-phosphate dehydrogenase gene promoter of *Aspergillus nidulans*

The glyceraldehyde-3-phosphate dehydrogenase enzyme (GPD) is a key enzyme in the glycolysis pathway responsible for the oxidation of D-glyceraldeyde-3-phosphate to the high energy compounds 1,3-bisphosphoglycerate and NADH. The fundamental role of GPD is in basic cellular metabolism, and the fact that its expression is not affected by growth on different carbon sources (Punt *et al.*, 1990) suggests that the *gpdA* gene is constitutively expressed. GPD is an intracellular protein, unlike GLA or CBHI. The *gpdA* promoter is predominantly applied when consistent levels of protein expression are required.

The constitutive nature of the *gpdA* promoter has enabled applications such as the development of a green fluorescent protein (GFP) reporter expression system in *P. chrysosporium* (Ma *et al.*, 2001), the use of *gpdA* transcripts as loading controls on northern blots (Saloheimo *et al.*, 2003) and for the expression of the *hph* gene in *Ophiostoma piceae*, which confers resistance to hygromycin B (Lee *et al.*, 2002).

The *gpdA* promoter has been used to constitutively express the transcription factor, ALCR, to overcome CREA repression on the *alcR* gene promoter (Mathieu *et al.*, 1994). This application of the promoter enabled the identification of further effects of CREA on genes involved in ethanol metabolism (for further discussion on these transcription factors refer to Section 1.5.8).

The *A. niger gpdA* gene promoter contains a promoter-enhancing element termed the *gpd* box. Deletion of this 50 bp element resulted in a 50% decrease in promoter activity (Punt *et al.*, 1990) and insertion of this *gpd* box sequence upstream of an *amdS* promoter (see Section 1.5.6) fused to *lacZ* resulted in a 6 to 30 times increase in *amdS* promoter activity in *A. nidulans* (Punt *et al.*, 1992).

For the fluorescent labelling of fungal hyphae, fluorescent reporter genes are often expressed under the *gpdA* promoter. GFP was expressed in *O. piceae* under the *A. nidulans gpdA* promoter to enable the detection of fungal colonisation in wood (Lee *et al.*, 2002). The fluorescent reporter protein DsRed2 was placed under the control of the *A. nidulans gpdA* promoter in *Fusarium oxysporum* f. sp. *lycopersici* to visualise the colonisation pattern of the pathogenic fungus in tomato plant tissues (Nahalkova *et al.*, 2003). The mutant form of the DsRed fluorescent reporter protein, DsRed-Express, was expressed under the *A. nidulans gpdA* promoter in the fungi *Penicillium paxilli*, *T. harzianum* and *T. virens* (Mikkelsen *et al.*, 2003) to trial the use of the new fluorescent reporter mutant in filamentous fungi.

1.5.3.1 The *gpdA* gene promoter for the expression of industrial proteins

The *gpdA* promoter has been successfully applied for the industrial production of some proteins. When grown on the cheap, undefined industrial waste medium of molasses, *A. niger* successfully expressed the *T. reesei xyn2* and *egl1* genes when placed under the *A. niger gpdA* promoter with their own endogenous signal sequence (Rose *et al.*, 2002). The specific activity of XYN2 in the supernatant was almost one and a half times greater than the reported complete xylanolytic activity produced by *T. reesei* RutC-30 strain (see Section 1.5.8.2). The level of endoglucanases expressed was also high but most of the protein was inactive, which the authors attributed to different glycosylation in the heterologous host.

The overexpression of the homologous *A. niger* extracellular feruloyl esterase (FAEA) was successfully carried out by inserting *faea* cDNA, including the signal sequence, under the *A. nidulans gpdA* promoter and 5' untranslated region (de Vries *et al.*, 1997; Record *et al.*, 2003). When grown on glucose, the FAEA yield was 1 g/L making it potentially useful for the pulp and paper industry.

Both of the above cases demonstrate the use of a promoter for a gene that encodes an intracellular protein, for the expression and secretion of a heterologous extracellular protein. It is interesting to see that the yields for XYN2 and FAEA are relatively high for a promoter seemingly geared for "housekeeping" of intracellular protein levels. This observation suggests that potentially all promoters may be capable of large-scale protein production under the right conditions.

1.5.4 The *pki1* gene promoter of *Trichoderma reesei*

Pyruvate kinase is an enzyme in glycolysis, responsible for the dephosphorylation of phosphoenolpyruvate to generate pyruvate and ATP. The promoter of the gene encoding pyruvate kinase in *T. reesei, pki1,* is considered to be a constitutive promoter (Schindler *et al*., 1993). This promoter is used when a consistent level of expression is required. For example, the *pki1* promoter is commonly used on expression cassettes to drive the expression of the selection marker gene *hph* in *T. reesei* (Mach *et al*., 1994; Te'o *et al*., 2000; Cziferszky *et al*., 2002).

In a study investigating the induction of endochitinase (*ech42*) and N-acetyl-β-D-glucosamine (*nag1*) expression in *T. harzianum* during mycoparasitic interactions. An expression cassette with GFP under the control of the *pki* promoter was used successfully to demonstrate constitutive expression compared to the upregulated expression of GFP under the *ech42* and *nag1* gene promoters (Zeilinger *et al*., 1999).

T. reesei xyn1 and *xyn2* along with their endogenous signal sequences, have been expressed under the *pki* promoter on glucose medium in order to express xylanases in the absence of cellulases. Transformants with multiple copies of the expression cassette integrated into their genomes produced 76 and 145 U/mg of XYN I and XYN II protein respectively, compared to 26 U/mg by the parental strain indicating that the *pki* promoter can be a useful promoter for protein production on glucose (Kurzatkowski *et al*., 1996).

1.5.5 The *alcA* gene promoter of *Aspergillus nidulans*

The *alcA* gene encodes alcohol dehydrogenase (ADHI) in *A. nidulans*, which is an intracellular protein that has a small role in the maintenance of cellular homeostasis under normal conditions. However, the *alcA* promoter is also strongly induced by ethanol and the ketones, cyclopentanone (Waring *et al.*, 1989) and 2-butanone (Felenbok *et al.*, 2001). The *alcA* promoter is considered one of the strongest promoters in *A. nidulans* of which the genetics is well known. However, *A. nidulans* is not used for industrial protein production.

The *alcA* gene is induced by binding of the transcription enhancer, AlcR, to three regions along the promoter (reviewed in Felenbok *et al.*, 2001). The consensus core binding motif is 5' WGCGG and exists as repeats either in tandem or as a palindrome. The intensity of *alcA* induction is variable depending upon which combinations of the three AlcR binding sites are occupied. AlcR binding to all three sites is necessary for high level transcription (Kulmburg *et al.*, 1992). The *alcA* promoter is also subject to carbon catabolite repression via the action of CreA (Kulmburg *et al.*, 1993). For a description of the repression mechanism by CreA, refer to Section 1.5.8.2.

The strength of the *alcA* promoter and its tight regulation make it useful for a range of applications. The *alcA* promoter has been used to drive an expression system in *A. fumigatus* to investigate the function of genes involved in human pathogenesis (Romero *et al.*, 2003). It has also been applied to express a fusion of the protein phosphatase, BimG to GFP in *A. nidulans* to visualise the distribution of the BimG protein in living hyphae (Fox *et al.*, 2002). This promoter has further been used on a series of expression vectors created for fluorescent labelling of cellular organelles using the fluorescent

reporters DsRed(T4), mRFP1 and the blue fluorescent protein (BFP) (Toews *et al.*, 2004). The inducible nature of the *alcA* promoter has made it useful for the expression of transgenes in plants (Caddick *et al.*, 1998) and to drive the expression of human proteins in *A. nidulans*. Human lactoferrin fused to the GLA signal peptide was expressed and secreted at levels of up to 5 μg/mL which demonstrates the potential use of the *alcA* promoter for heterologous protein production (Ward *et al.*, 1992).

1.5.6 The *amdS* gene promoter of *Aspergillus nidulans*

The promoter of the gene encoding acetamidase (*amdS*) in *Aspergillus nidulans* is well characterised and has served as a model filamentous fungal promoter for understanding promoter regulation. The pioneering work by Hynes (Hynes *et al.*, 1970; Hynes, 1972; Hynes, 1978a; Hynes, 1978b; Hynes, 1979; Hynes, 1980; Hynes, 1982; Hynes *et al.*, 1983) describes the use of mutation and phenotypic analysis to deduce a profile of the inducing and repressing conditions affecting the *amdS* gene promoter and to identify the transcription factors involved.

Aspergillus is able to use acetamide as both a carbon source and nitrogen source by the degradation of acetamide to CH_3OOH and ammonia by acetamidase (Hynes, 1970). Other nitrogen-containing compounds such as benzoate, benzamide, β-alanine and other α-amino acids, and sources of acetyl coenzymeA are also inducers of acetamidase expression (Hynes, 1978a). Compounds such as ammonia, L-glutamate and L-glutamine are repressors of *amdS* expression. It was noted early that the regulatory elements functioned independently and additively (Hynes, 1978a) by affecting the rate of transcription. The sequencing of the *amdS* gene and flanking sequences has since

enabled the fine analysis of the transcription binding sites along the promoter (Figure 1-2).

In a repressed or inactive state there is an exact positioning of nucleosomes over regions of the promoter and coding sequence. On derepression, the nucleosomes are repositioned and all of the cis-acting motifs of the promoter emerge within the nucleosome-free region (Figure 1-2).

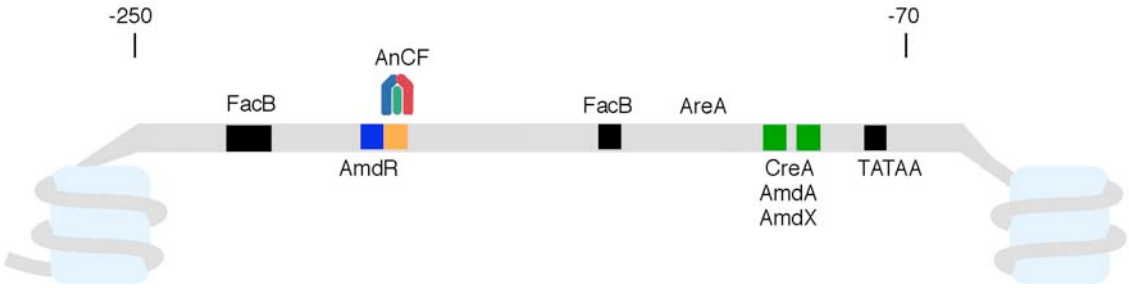

Figure 1-2. The positioning of regulatory motifs along the promoter of the *amdS* gene. Binding of the AnCF complex (homologue to the Hap complex, Section 1.5.8.1) to the CCAAT box (orange) mediates chromatin rearrangement such that the sequence between –70 and –250 from the translation start site becomes a nucleosome-free region. The DNA-binding proteins are discussed in the text.

The transcriptional activator AmdR is constitutively expressed at low levels which sets the basal level of induction of *amdS* (Andrianopoulos *et al*., 1988). AmdR is activated by □-amino acids and induces *amdS* in response. AmdR also binds to the promoters of the genes *gabA*, *gabB*, *lamA* and *lamB* (Katz *et al*., 1989; van Heeswijck *et al*., 1991). The binding site for AmdR on the *amdS* promoter overlaps with a CCAAT-box (Littlejohn *et al*., 1992) and AmdR binding only occurs if the CCAAT-box is bound by the AnCF complex (Steidl *et al*., 1999).

The AnCF binding complex is composed of the proteins HapB, -C and –E which are homologues to the *Saccharomyces cerevisiae* HAP2, HAP3 and HAP5 proteins

(Papagiannopoulos *et al*., 1996; Steidl *et al*., 1999). The binding of AnCF to the CCAAT box mediates chromatin remodelling along the *amdS* locus such that a nucleosome-free region materialises within the promoter and encompasses all of the known binding sites for various trans-acting transcription regulatory elements (Narendja *et al*., 1999). Unlike the situation for AmdR, binding of the transcription factors AmdA, AreA, FacB and CreA is not dependent on AnCF binding to the CCAAT box, however *amdS* expression was markedly reduced without the formation of the nucleosome-free region (Narendja *et al*., 1999).

Approximately 58 bp upstream from the transcription start point (*tsp*) is a GAGGGGG motif which is the core DNA-binding site for three independent transcription factors which all possess similar C2H2 zinc finger DNA-binding motifs: the CreA binding site, which is a mediator of carbon catabolite repression (see Section 1.5.8.2), AmdX which is an activator of *amdS* expression (Murphy *et al*., 1997) and AmdA. AmdA induces *amdS* expression in response to acetate and also binds to the promoters of the genes *aciA* and *alcB* (Andrianopoulos *et al*., 1997).

FacB is the activator of genes involved in acetate utilisation and therefore induces *amdS* expression in response to acetate. The transcription factor FacB binds to the promoter regions of the genes for carnitine acetytransferase (*facC*) (Stemple *et al*., 1998), acetyl-CoA synthase (*facA*), isocitrate lyase (*acuD*) and malate synthase (*acuE*) (Todd *et al*., 1998). However, the sequences of the binding sites between these genes is dissimilar and include two possible consensus sequences, 5' TCC/GN_{8-10}C/GGA and 5' GCAGNTNNCCN_{1-2}GGC, which are considered to be more conserved in secondary structure than in sequence. Therefore, the binding of FacB demonstrates the difficulty in identifying putative transcription regulatory elements simply based on sequence.

Overall, the knowledge gained from the in-depth studies into the *A. nidulans amdS* promoter highlight the levels of complexity involved in promoter regulation.

1.5.7 The concept of a constitutive promoter

Actin is a key protein in the cellular cytoskeleton and the promoter is regarded as a constitutive promoter. Importantly, in *T. reesei* it has been shown that actin expression does not depend on the carbon source employed for growth (Matheucci *et al*., 1995). Thus, the level of actin expression is often referred to as a loading control on northern blots (Ilmén *et al*., 1996a; Ilmén *et al*., 1997; Zeilinger *et al*., 2003). However, expression under this constitutive promoter is not always constant and a ten-fold difference in actin transcript levels was reported between three day old cultures grown on a range of media (Margolles-Clark *et al*., 1997).

The *gpd* gene promoter has also been considered constitutive (see Section 1.5.3). However, in *T. harzianum* following the simulation of mycoparasitism by the addition of *R. solani* cell walls, or induction of conidiation by a pulse of blue light, the level of *gpd* mRNA decreased strongly in the mycelium or in the conidiophore respectively (Puyesky *et al*., 1997). The above mentioned studies demonstrates that even genes fundamental to cellular integrity, such as *act1* and *gpd*, are subject to variable levels of expression in some circumstances.

It is expected that during different phases of the fungal life cycle, for example, conidiation, germination or exponential growth, the transcriptional profile of a cell would be variable to accommodate the phenotypic changes. This supposition has now been demonstrated clearly by the use of proteomics and transcriptomics (Hegde *et al*.,

2003; Tyers *et al*., 2003) which reflect the multifactorial changes in an organism. A fundamental rule for each of these disciplines is that the culture conditions are as precise as possible for satisfactory replication of results. With these technologies, the dynamic nature of cellular physiology has received renewed appreciation.

1.5.8 Regulatory elements of *Trichoderma reesei* gene promoters

The mechanisms of gene regulation must be understood in order to develop fungi as efficient protein expression systems. Some of the transcription factors which have been identified and characterised in filamentous fungi are shown in Table 1-4. A thorough list of transcription factors characterised in yeast is available at http://biochemie.web.med.uni-muenchen.de/Yeast_Biology/YTF_alpha_2.htm. Since the production of the main cellulases in *T. reesei* is regulated at the transcriptional level in response to the carbon source available, a fair amount of knowledge on fungal promoter regulation is founded on an understanding of the regulation of the cellulase genes and the process of carbon repression.

It is clear that the regulation of the genes encoding the cellulolytic enzymes is not a simple switch mechanism, but requires a coordinated, overlapping network of transcription factors to allow dynamic expression patterns to enable the enzymes to function synergistically in the degradation of complex polymers. The transcription factors which have been most investigated are described below.

Table 1-4. Overview of some transcription factors characterised in filamentous fungi.

DNA-binding protein	DNA binding site 5' to 3'	Organism	Target genes	Element function	Reference
Protein unidentified	CCAGN_{13}CTGG	*Trichoderma atroviride*	*nag1*	Essential for induction of the *nag1* gene by N-acetyl-D-glucosamine. N_{13} motif contains a CCAAT box.	Peterbauer *et al*. (2002a)
Aba AbaA	CATTCY GGAATG	*Fusarium oxysporum* *Aspergillus nidulans* *T. atroviride*	*pgx2*, *endopg1*, *endopg2* *brlA, wetA, rodA* 1,3-β-glucosidase	Establishes feedback regulatory loop that is required for conidiophore development.	Andrianopoulos *et al*. (1994), de las Heras *et al*. (2002), Donzelli *et al*. (2001)
ACEI	AGGCA(AA)	*T. reesei*	*cbh1, xyn1, xyn2*	Represses cellulase and xylanase genes in cellulose and sophorose.	Aro *et al*. (2003) Saloheimo *et al*. (2000)
ACEII	GGCTAATAA	*T. reesei*	*cbh1, cbh2, egl1, xyn2*	Activates cellulase and xylanase genes on cellulose-induction. Sophorose induction bypasses ACEII activation.	Aro *et al*. (2001)
AFLR	TCGN_5CGA	*A. parasiticus* *A. nidulans*	aflatoxin biosynthesis genes	Transcription activator.	Chang *et al*. (1999), Yu *et al*. (2004)
AGAA box binding protein	AGAA	*T. reesei*	*xyn2*	Binds only under noninducing conditions to function as a repressor.	Würleitner *et al*. (2003)
AGGGG binding protein	AGGGG	*T. atroviride*	*nag1, ech42*	Essential for induction of the *nag1* gene by N-acetyl-D-glucosamine.	Peterbauer *et al*. (2002a), Brunner *et al*. (2003)
AlcR	WGCGG tandem or inverted repeats	*A. nidulans*	*alcA, aldA*	Transcription activator of genes involved in ethanol catabolism. AlcR is self regulating by positive feedback.	Kulmburg *et al*. (1992), Flipphi *et al*. (2001)
AmdA	GCGGGG GAGGGG	*A. nidulans*	*aciA, alcB, amdS*	Transcription activator on acetate induction.	Andrianopoulos *et al*. (1997), Saleeba *et al*. (1992)
AmdR	TCGGCGAA	*A. nidulans*	*amdS, gatA, gatB, lamA, lamB*	Induced by ω-amino acids. AmdR binding on the *amdS* promoter is dependent on bound AnCF.	Katz *et al*. (1989), van Heeswijck *et al*. (1991)
AmdX	GCGGGG GCGGGG	*A. nidulans*	*amdS*	Binding site on *amdS* promoter overlaps with AmdA and CreA binding sites. AmdX is an activator of *amdS* expression.	Murphy *et al*. (1997)

AreA	GATA core	*A. nidulans*	*fmdS*, *amdS, niiA, niaD*	Induced by glutamine and ammonium and creates a hierarchy of gene expression. Competes with CreA site on *amdS* promoter. Mediates nitrogen metabolite repression. Causes chromatin remodelling. Regulated *areA* mRNA degradation is a form of self-regulation.	Fraser *et al*. (2001), Muro-Pastor *et al*. (1999), Morozov *et al*. (2001), Davis *et al*. (1993)
AreB	undefined	*A. nidulans*	not determined	*areB* encodes three proteins that are transcription regulators.	Conlon *et al*. (2001)
AnCP/AnCF	see Hap complex				
BrlA	VDAGGGGG/A	*T. atroviride* *A. nidulans*	1,3-β-glucosidase gene	Binds to several conidiation responsive elements.	Donzelli *et al*. (2001), Adams *et al*. (1988)
cAMP responsive element (CRE)	GTGACGTCAC	*N. crassa*	*grg-1*	Required with NRS to repress expression in glucose.	Wang *et al*. (1994)
CpcA Cpc1	TTGACTCT ATGACTCA GTGACTCAT	*A. nidulans* *N. crassa*	Genes for amino acid biosynthetic enzymes	Upregulation during amino acid starvation. Under autoregulation.	Hoffman *et al*. (2001), Wanke *et al*. (1997), Ebbole *et al*. (1991)
CreA CREI	G/CYGGGG SYGGRG SYGGRG	F. oxysporum *A. niger* *A. nidulans* *T. reesei* *T. atroviride*	*pgx1*, *pgx2*, *endopg1*, *endopg2* *lacA, alcR* *cbhI, xyn1* 1,3-β-glucosidase gene	Carbon catabolite repression binding element. Represses cellulase and xylanase genes in glucose.	Kulmburg *et al*. (1993), de las Heras *et al*. (2002) de Vries *et al*. (2002) Mathieu *et al*. (1994), Dowzer *et al*. (1991) Aro *et al*. (2003), Takashima *et al*. (1996), Mach *et al*. (1996), Donzelli *et al*. (2001)
CYS3	ATGRYRYCAT	*N. crassa*	Genes for ATP sulfurylase, APS kinase and cysteine synthase, *cys-14*	Master regulatory protein for sulfur regulatory genes.	Kanaan *et al*. (1991), Marzluf *et al*. (1997), Li *et al*. (1996)
FacB	TCC/GN_{8-10}C/GGA GGAGNTNNCCN_{1-2}GGC	*A. nidulans*	*facC, facA, amdS, acuD, acuE*	Acetate induction of *amdS*.	Todd *et al*. (1998), Stemple *et al*. (1998)
FlbD	undefined	*A. nidulans*	Not determined	Binds to BrlA and activates genes involved in conidiation.	Wieser *et al*. (1995)
GTAATA box	GTAATA	*T. reesei*	*cbh2*	Adjacent to the CCAAT box that together form the *cbh2* activating element (CAE).	Zeilinger *et al*. (2003)

HACA	undetermined	*A. niger*	*bipA, cypB, pdiA, hacA, prpA, tigA*	Upregulates expression of genes involved in UPR.	Mulder *et al*. (2004)
HAP complex hapB/C/E subunits	ATTGG CCAAT GGGTAAATTGG ATTGG	*T. reesei* *A. nidulans* *T. reesei* *A. niger* *T. atroviride*	*cbh2* *amdS, taaG2, gatA* *xyn2* *lacA* *nag1*	Binding induces chromatin remodelling. Hap complex binding is essential for efficient expression.	Zeilinger *et al*. (2001), Steidl *et al*. (1999), Kato *et al*. (1997), van Heeswijck *et al*. (1991), Wurleitner *et al*. (2003), de Vries *et al*. (2002), Peterbauer *et al*. (2002a)
Heat shock protein	NGAAN	*F. oxysporum*	*pgx1*, *pgx2*, *endopg1*, *endopg2*	Induction in response to stress conditions.	de Vries *et al*. (2002)
NirA	CTCCGHGG	*A. nidulans*	*niaD, niiA*	Potentially binds to several sites along a promoter. Each site has unequal importance in gene induction by nitrate.	Punt *et al*. (1995)
NIT2	TATC, TATCTA, GATA	*N. crassa* *A. nidulans*	100 genes in nitrogen metabolism eg. *nit-3, alc, lao* *niaD, niiA*	Lifts nitrogen catabolite repression when cellular glutamine is limited. Activates expression with direct protein interaction with NIT4 on nitrate induction and nitrogen source derepression.	Tao *et al*. (1999), Mo *et al*. (2003) Fu *et al*. (1990)
NIT4	GATA	*N. crassa*	*nit-3*	Positive regulation of genes involved in nitrogen metabolism.	Chiang *et al*. (1995)
NmrA	binds DNA via AreA	*A. nidulans*	as for AreA	Transcription repressor in nitrogen metabolism. Possible redox sensor. Binds to AreA.	Lamb *et al*. (2003). Andrianopoulos *et al*. (1998)
NRE	GATA	*Penicillium chrysogenum*	*PcbAB, niiA-niaD*	Belongs to GATA family of transcription factors.	Haas *et al*. (1995), Zadra *et al*. (2000)
NRS (*Neurospora* repressor site)	TTGCTAGCAA	*Neurospora crassa*	*grg-1*	Works with CRE to repress expression in glucose.	Wang *et al*. (1994)
NUC-1	CACGTG	*N. crassa*	*pho2+, pho4+*	Positive regulation of genes for phosphorous acquisition enzymes. Negatively controlled by the proteins PREG and PGOV.	Peleg *et al*. (1994), Kang *et al*. (1993)
PacC	GCCARG	*P. chrysogenum* *F. oxysporum* *A. nidulans*	*pcbAB-pcbC* *endopg1* *gabA, ipnA*	Regulation of gene expression in response to pH. Active form occurs in alkaline conditions.	Suarez *et al*. (1996), Caracuel *et al*. (2003), Espeso *et al*. (1996), Espeso *et al*. (2000), Tilburn *et al*. (1995)

qa-1F	GGATAAACAATTATC	*N. crassa*	quinic acid gene cluster (*qa*)	Activator of genes that enable utilisation of quinic acid as a carbon source.	Asch *et al.* (1991)
Seb1	AGGGG	*T. atroviride*	*nag1*, *ech42*	Important in osmotic stress response but not chitinase gene expression.	Peterbauer *et al.* (2002b)
SRE	GATA	*N. crassa*	*sid1*	Regulator of genes involved in iron metabolic pathway. Negative regulator of siderophore genes.	Zhou *et al.* (1999), Harrison *et al.* (2002)
SteAp	undefined	*A. nidulans*	not determined	Involved in sexual reproduction.	Vallim *et al.* (2000)
Ste12p	TGAAACA	*F. oxysporum*	*endopg1*	Homologous to yeast Ste12. Involved in induction of filamentous growth and pathogenesis.	Di Petro and Roncero (1998)
StuAp	A/TCGCGT/ANA/C	*F. oxysporum* *A. nidulans*	*pgx2*, *endopg1*, *endopg2* *cpeA*	Regulator of genes involved in development, cell cycle, sexual reproduction and ascosporogenesis.	Dutton *et al.* (1997), Scherer *et al.* (2002)
TOXE	ATCTCNCGNA	*Cochliobolus carbonum*	*tox2* locus	Activates transcription from all TOX2 genes. Important in plant pathogenicity.	Pedley *et al.* (2001)
XLNR	GGCTAA	*A. niger*	*xlnB, xlnC, xlnD, eglA, eglB, aguA, axeA, axhA, faeA, lacA*	Transcription activator of xylanases and endoglucanases.	van Peij *et al.* (1998)

1.5.8.1 The HAP complex

The CCAAT box is one of the most common elements of eukaryotic fungal promoters found in either the forward or reverse orientations (Mantovani, 1998). The complex which binds to this motif is composed of three to four proteins which have been highly conserved throughout evolution (Li *et al*., 1992). The CCAAT-binding complex was first identified in *Saccharomyces cerevisiae* where it is a heterotrimeric DNA-binding complex (HAP) composed of the proteins Hap2, Hap3 and Hap5 (Pinkham *et al*., 1985; Forsburg *et al*., 1989; Olesen *et al*., 1990). A fourth subunit, Hap4 interacts with the HAP complex and functions as a transcriptional activator (McNabb *et al*., 1995). Binding of the HAP complex to the CCAAT motif was found to increase expression of the gene (Papagiannopoulos *et al*., 1996). The CCAAT motif is very common within eukaryotic promoters and the HAP 2/3/5 complex is highly conserved through evolution. The mammalian equivalent of the complex is composed of the subunit NF-YA, NF-YB and NF-TC (Nakshatri *et al*., 1996). In filamentous fungi, a HAP-like complex has been identified in *A. nidulans* (Kato *et al*., 1997; Steidl *et al*., 1999), *A. oryzae* (Tanaka *et al*., 2001) and *T. reesei* (Zeilinger *et al*., 2001) and is composed of the subunits HapB, HapC and HapE. Strains deficient in either of these proteins grow slowly and have poor conidiation (Steidl *et al*., 1999).

A fourth subunit HapX, which may not be a functional homologue to the *S. cerevisiae* Hap4, was only recently identified in *A. nidulans*. HapX not only interacts with the HAP complex as a transcription activator, but also possesses a leucine zipper domain and three cysteine-rich regions at the N-terminus, which may indicate that HapX can also interact directly and independently with DNA (Tanaka *et al*., 2002). Sequence

homologues have also been identified in *A. oryzae* and *N. crassa* (Tanaka *et al.*, 2002). The mechanism of HAP complex assembly is shown in Figure 1-3.

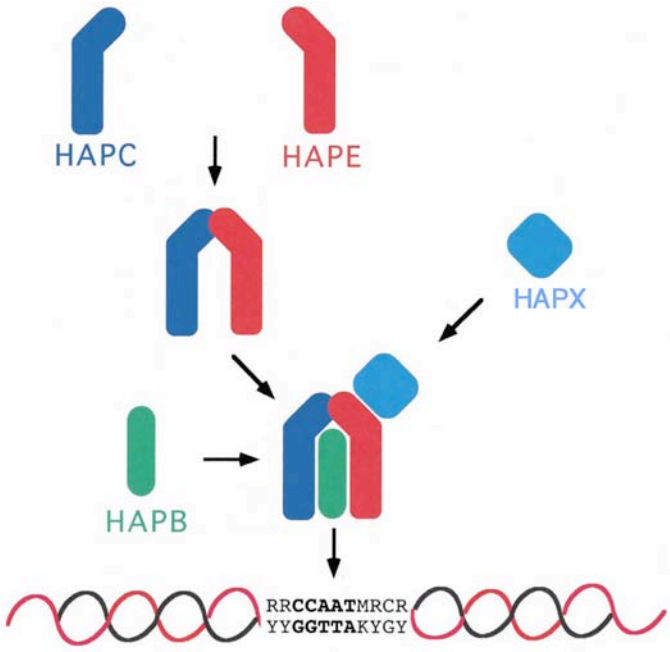

Figure 1-3. Schematic representation of the assembly of the HAP complex in *A. nidulans*. Subunits HapC and HapE combine first, followed by subunit HapB. The heterotrimer can then bind to the CCAAT motif of promoters with or without the fourth subunit HapX. Figure modified from Brakhage *et al.* (1999).

Functional CCAAT boxes have been identified on *T. reesei xyn2* promoter (Würleitner *et al.*, 2003) and *cbh2* promoter (Zeilinger *et al.*, 2001), *A. nidulans amdS* promoter (Narendja *et al.*, 1999), taka-amylase A (*taa*) gene promoter and *gatA* promoter (Kato *et al.*, 1997). The presence of HAP-binding CCAAT motifs on fungal genes has been reviewed by Brakhage *et al.* (1999).

Binding of the HAP complex onto the promoter of the *amdS* gene results in the formation of a DNase I-hypersensitive site that is demonstrative of nucleosome remodelling along the *amdS* locus (Narendja *et al.*, 1999). The nucleosomes become positioned such that the binding motifs for other transcription factors become exposed

in a nucleosome free region of the promoter (see Section 1.5.6). A similar mechanism has been described for HAP complex binding and nucleosome positioning along the *T. reesei cbh2* promoter. The *cbh2* activating element (CAE) of the *cbh2* promoter is composed of a CCAAT box on the template strand, adjacent to a GTAATA box on the coding strand. Binding of trans-acting elements to each component of the CAE during inducing conditions is vital for the shifting of nucleosomes to make the TATA box accessible (Zeilinger *et al*., 2003).

1.5.8.2 The carbon repression mediator CREI

The *cre1* gene of *T. reesei* and *T. hazianum* is homologous to the *creA* gene of *A. nidulans* (Strauss *et al*., 1995; Ilmén *et al*., 1996b) and both regulate the gene expression of several genes influenced by carbon repression (Ilmén *et al*., 1996b), The conserved DNA-binding motif is 5'-SYGGRG (Strauss *et al*., 1995) and can exist in either orientation (Ilmén *et al*., 1996a; Takashima *et al*., 1996).

The serine at position 241 in CREI must be phosphorylated for binding of the carbon catabolite repressor to its DNA target motifs (Cziferszky *et al*., 2002) and may form a complex with other proteins (Strauss *et al*., 1995). Binding of CREI protein was important for the positioning of nucleosomes on the *cbh2* promoter, however induction of the promoter could still occur with mispositioned nucleosomes (Zeilinger *et al*., 2003).

The function of CREI was confirmed using the *T. reesei* Rut-C30 strain which is a hyperproducer of cellulase, even in the presence of glucose (Ilmén *et al*., 1996b). It was discovered that Rut-C30 contains a truncated version of the *cre1* gene which codes for

only one of the two zinc finger domains (described for *Aspergillus* CreA by Dowzer and Kelly (1991)), rendering the protein inefficient at DNA binding and resulting in a derepressed phenotype. Transformation of this strain with an intact copy of the *cre1* gene resulted in a return to carbon repression in the presence of glucose. Likewise, the deletion of *cre1* from a parent strain and replacement with the truncated form of *cre1,* resulted in higher levels of cellulase transcripts, 20-fold higher cellobiohydrolase and endoglucanase activity and 10-fold higher xylanase activity when grown on lactose (Nakari-Setälä *et al.*, 2004).

CreA-mediated carbon repression is involved in the utilisation of ethanol in *A. nidulans*. Functional CreA binding sites have been located on the upstream region of the transcription-activator gene *alcR*. The ALCR protein is a DNA-binding transcription activator of the alcohol dehydrogenase encoding gene (*alcA*) and the aldehyde dehydrogenase encoding gene (*aldA)*. Functional CreA binding sites have also been identified on the upstream region of *alcA* (Kulmburg *et al.*, 1993). Therefore CreA represses the utilisation of ethanol at two different levels of the metabolic pathway ie. by preventing expression of the trans-acting transcription factor (AlcR) and by directly inhibiting transcription of the target gene (*alcA*).

1.5.8.3 The transcription regulators ACEI and ACEII

The transcription regulator ACEI contains three zinc finger motifs and binds *in vitro* to at least eight sites on the *cbh1* promoter with the conserved core sequence AGGCA (Saloheimo *et al.*, 2000). When ACEI was initially identified it was aptly named as activator of cellulase expression because ACEI induced transcription of *T. reesei cbh1* in a yeast expression system (Saloheimo *et al.*, 2000). However, deletion of *ace1* from

T. reesei resulted in earlier and higher expression levels of the cellulase genes *cbh1*, *cbh2*, *egl1* and *egl2* and of the xylanase genes *xyn1* and *xyn2* when induced by cellulose. These results indicated that ACEI may act as a repressor of cellulase expression under some conditions (Aro *et al.*, 2003). The growth of a *ace1* deletion mutant was impaired on the neutral carbon source sorbitol, suggesting that ACEI also affects genes outside those of the cellulolytic system (Aro *et al.*, 2003).

Within the 5' upstream region of *ace1*, there are 11 putative CREI binding sites. However, there was no difference in the levels of *ace1* gene transcript levels between the RutC-30 strain and RutC-30 transformed with a functional *cre1* gene when cultured on different carbon sources (Aro *et al.*, 2003). The authors concluded that *ace1* is not affected by CREI or glucose repression.

ACEII was also discovered as an inducer of expression from the *cbh1* promoter in a yeast expression system (Aro *et al.*, 2001). ACEII possesses a zinc binuclear cluster DNA-binding domain of the type ($Zn(II)2Cys_6$) and has been shown to bind to the 5'-GGCTAATAA motif of the *cbh1* promoter *in vitro*. Deletion of the *ace2* gene resulted in a significantly reduced level of *cbh1*, *cbh2* and *egl1* expression and protein activities on cellulose medium compared to the parental strain[1]. However, this reduced level became less pronounced by day six of the culture period, suggesting that induction had only been slowed in the deletion mutant. Deletion of *ace2* also reduced *xyn2* expression

[1] The reduction in expression of the genes *cbh1*, *cbh2* and *egl1* in the *ace2* deletion mutant was not apparent when the cultures were induced by sophorose, indicating that the mode of induction by the glucose polymer cellulose is not the same as the mode of induction by the disaccharide sophorose.

levels but had no effect on *xyn1* expression. Therefore ACEII functions as a transcriptional activator (Aro *et al.*, 2003).

The identification of several transcription factors and their characterisation has led to a greater understanding of the regulation of the cellulolytic enzyme system of *T. reesei*. The data demonstrates the complexity of the regulation of the cellulase gene system and suggests that other factors, yet to be identified, must also be involved in the simultaneous regulation of several promoters to create the dynamic patterns of protein expression in response to the carbon source on which the fungus is growing. The highly interconnected regulation mechanisms of the cellulase system exemplify the complexity of gene expression that is likely to be paralleled in other systems.

1.6 The need for novel promoters in industrially-relevant fungi

Under conditions conducive to *cbh1* induction, the expression of a number of secreted enzymes is upregulated and CBHI constitutes approximately 60% of the total protein. Alternatively, in conditions unfavourable for cellulase induction, only a few proteins are secreted (Mäntylä *et al.*, 1998) which makes them of a higher purity and in particular, without cellulase contamination. This is important for the production of individual hydrolysing enzymes such as xylanase, which is used for removing xylan from wood pulp, where the presence of cellulase is detrimental to pulp processing for paper manufacture. In addition, the higher purity of a secreted protein is advantageous for the production of most heterologous proteins, especially for the production of bio-pharmaceuticals. Therefore, *Trichoderma reesei* promoters effectively expressed on glucose are desirable for various applications.

The degree of contamination of the secreted protein of interest with cellulases has been decreased to date by using cellulase knockout strains of *T. reesei* to express the gene of interest (Harkki *et al.*, 1991; Wang *et al.*, 2004). However, it would be simpler to have a constitutively active promoter rather than creating multiple knockout strains for each situation.

Industrial enzymes are often produced in large fed batch cultures in which the mycelium cycles between active growth phases and stationary phases. Gordon and colleagues (2001) have attempted to establish a novel expression system in *Fusarium venenatum,* using a combination of two different promoters to ensure that the expression of the recombinant protein is at a constant high rate in fed-batch culture. The *A. niger glaA* encoding sequence was placed under the *F. oxysporum* trypsin-like protease promoter which has an optimum production rate at pH 6.0 and is regulated independently of the specific growth rate. This recombinant was then transformed with several copies of the *A. niger glaA* gene under its own promoter, which has optimal activity in the exponential phase of growth at pH 4.5. Fed batch cultures were started at pH 4.5 then raised to pH 5.8 once the culture had reached stationary phase. The amount of glucoamylase in the culture supernatants increased more gradually in the double transformant compared to the rate of production under the trypsin-like protease promoter alone, but resulted in similar total quantities of glucoamylase. Although the protein yields from this trial were not improved, the concept of adapting the application of gene promoters to suit a particular set of conditions is feasible. However, this approach may be restricted by the small number of fungal promoters which have been characterised so far for high-level gene expression.

The over-expression of homologous proteins in *T. reesei* has been very successful and has resulted in the commercial application of many enzyme products (see Table 1-1). Consequently, *T. reesei* is currently being developed as an expression system for heterologous proteins and a better understanding of many of the abovementioned stages of protein production, such as secretion, glycosylation and proteolysis, is elementary to the success of *T. reesei* for this purpose. In addition, the proteins of interest and their industrial applications are now very diverse. The characteristics of the protein itself or its destined application may dictate the conditions under which the protein must be produced. For example, secretion at certain pH, in particular medium, or secretion in the absence of other enzymes. The likelihood that a single engineered strain, even with modifications at numerous stages the gene expression and product secretion, would be sufficient for the expression of many heterologous products with their specific requirements is remote. Therefore, the specifics of the expression system must be tailored to cater for the diversity of industrially applicable proteins. One of the areas of the expression system, which is in need of diversification, is the available choice of promoters.

1.6.1 Methods employed in the search for novel promoters

In the search for useful novel promoters, the following methods have been or could be used. The techniques are based on identifying high levels of either the gene transcript or protein product.

Freitag and colleagues (2001) took a random event approach to identifying strong promoters in *N. crassa*. The fungus was co-transformed with a selection marker and promoter-less fragments of EGFP, in the chance that the EGFP fragments would

integrate in frame under an endogenous promoter and thereby be expressed. This method was unsuccessful as none of the 10 000 transformants screened expressed the fluorescent reporter.

The complete genomes of some filamentous fungi such as *N. crassa, A. nidulans, A. niger, Phanerochaete chrysosporium* and *M. grisea*, have been sequenced (www.fgsc.net and www.jgi.doe.gov/programs/whiterot.htm). The raw *T. reesei* genome-sequencing project is complete but access to data is limited at this stage. With genome data available, promoters to homologous genes in other species can be identified. However, the identification of the homologous gene is not demonstrative of identifying a useful promoter. For example, the *T. reesei* phosphoglycerate kinase (*pgk1*) gene was isolated based on hybridisation to the *Saccharomyces cerevisiae PGK* gene. The *PGK* gene promoter has been used for protein production in yeast (Bao *et al*., 1999; van Mullem *et al*., 2003) but the *T. reesei* equivalent was inadequate for high level protein expression (Nakari *et al*., 1993). It would be more beneficial to isolate novel promoters based on the specific conditions in which they are active.

The identification of highly expressed genes under a set of specific conditions is possible by analysis of expressed sequence tags (ESTs) (Panabières *et al*., 2002). The creation of cDNA libraries from ESTs and the screening of the recombinants is generally representative of those genes highly expressed. A novel method to identify promoters active under specific growth conditions was developed by Nakari and colleagues (1993). The total RNA isolated from *T. reesei* grown in glucose fermentation culture was used to generate a cDNA lambda library and a cDNA probe library. The lambda library was hybridised with the probe library and the plaques which produced the strongest signals were postulated to corresponded to the highly expressed genes. To

confirm the findings, the cloned cDNA of a positive plaque was sequenced and a specific probe was designed. The new probe was then used to screen the original cDNA library to demonstrate that the strong signal of the plaque corresponded to a highly represented transcript in the starting RNA pool. Two promoters with strong activity on glucose medium were identified by this method: the translation elongation factor 1a (*tef1*) promoter and a novel promoter of a gene of unknown function, labelled the cDNA1 promoter. The activity of these two promoters was compared to a formerly identified *pgk1* promoter and a modified *cbh1* promoter which lacked a 1.1 kb fragment that contained the glucose repression regulatory motifs. A larger amount of endoglucanase 1 core protein (up to 190 mg/L) was expressed from the cDNA1 promoter than from the modified *cbh1* promoter when grown on glucose. This method was effective in identifying novel promoters which are capable of driving strong protein expression on glucose medium.

Genome data combined with microarray technology now offers the opportunity for identifying highly expressed genes on a large-scale (reviewed by Choi, 2004). Specifically, using this technology, genes which are highly expressed under a particular set of growth conditions can be identified. A cDNA microarray can be generated from cloned ESTs or total RNA and this array is probed with fluorescently tagged cDNAs, extracted under the particular conditions of interest. This technology has been used widely in fungal genomics (Nakajima *et al.*, 2000; Chambergo *et al.*, 2002; Shi *et al.*, 2003; Allen *et al.*, 2004; Sims *et al.*, 2004). In addition to identifying strongly expressed genes, microarrays also allow the identification of genes which are co-expressed with the gene of interest. The *T. reesei* genes *envoy* and *minute,* which affect cellulase expression, were discovered by this approach (Schmoll *et al.*, 2004). Currently,

microarray technology is still expensive and is dependent upon the generation of genome cDNA libraries or complete genome sequencing for the assembly of the DNA-array. Nevertheless, the holistic approach to cellular genomics offered by microarrays has insured that the technology is an increasingly popular method of choice.

A novel way of identifying strong promoters is by identifying the highly expressed genes using a proteomic approach. The separation of cellular proteins by two-dimensional gel electrophoresis enables the quantification of specific proteins within the entire proteome based on spot intensity. The quantity of protein may be indicative of the level of expression of that protein and the activity of the gene's promoter. In order to study the proteins associated with the cell wall of *T. reesei,* Lim and colleagues (2001) prepared cell wall extracts from mycelia grown under *cbh1* inducing and *cbh1* non-inducing conditions. The cell wall extracts were then separated by two dimensional gel electrophoresis and a range of protein spots were sequenced by mass spectrometry. The proteins were identified as vacuolar protease A, enolase, glyceraldehyde-3-phosphate dehydrogenase, transaldolase, protein disulfide isomerase, mitochondrial outer membrane porin, diphosphate kinase, translation elongation factor beta and HEX1. The HEX1 protein is the major component of the Woronin body found in filamentous fungi (see Section 1.7.2). Almost half of the spots isolated and sequenced by mass spectrometry failed to match any sequence on a BLAST database search which was a reflection of the lack of data available on fungal genome and proteome projects at the time of the research.

An interesting result from the above research was the discovery of several intense spots which were identified as the HEX1 protein. The authors suggested that the spots of lower molecular mass than 21 kDa were degraded products of the 21 kDa HEX1 protein

however, no explanation was offered for the nature of the spots greater than 21 kDa. It was not until the crystal structure of the Woronin body was solved in 2003 (Yuan *et al.*, 2003) that the presence of multiple spots representing HEX1 was placed in context. These findings are discussed in depth later with reference to the findings of the work conducted in this thesis (see Section 3.2.7.1). Nonetheless, a large spot identified as a1 in Figure 1-4 indicated that HEX1 was an abundant protein within the cell envelope extract. In fact, from mycelia grown in glucose medium, six of nine spots were identified as HEX1, which amounted to approximately 50% of the total protein extracted from the cell envelope (Figure 1-4). HEX1 was also an abundant protein

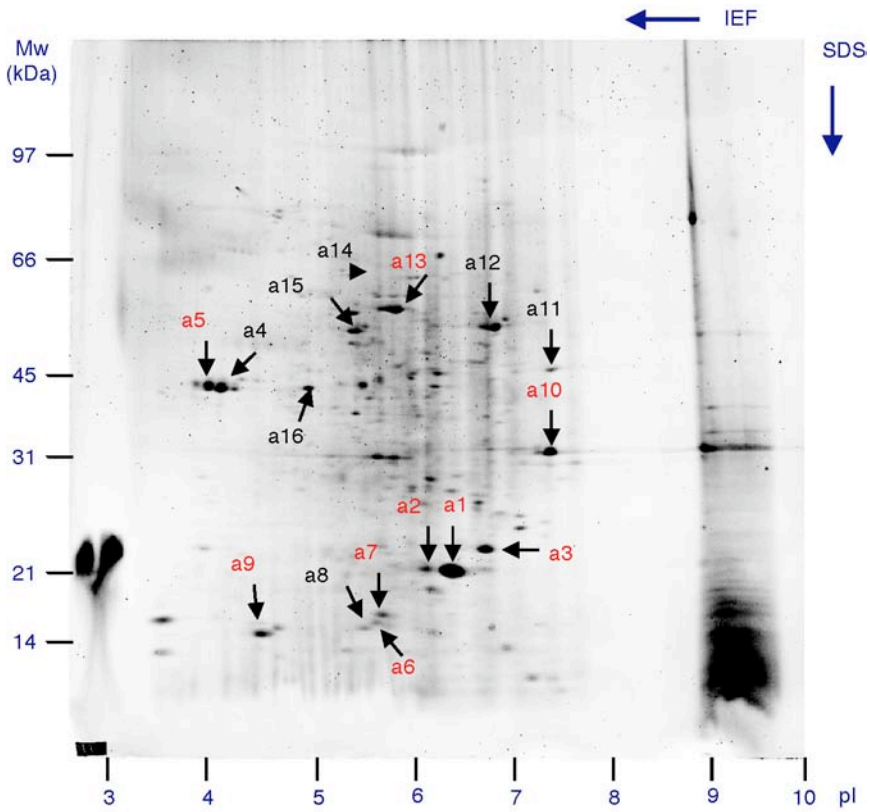

Figure 1-4. Two dimensional gel separation of proteins from the cell envelope of *T. reesei* grown on glucose medium. All spots that were subjected to protein sequencing are labelled and those shown in red were identified as HEX1. Figure modified from Lim *et al.* (2001).

associated with the cell envelope of mycelia grown on cellulase-inducing medium containing cellulose and lactose. These results suggested that a strong constitutive promoter drives the expression of HEX1 and the possibility of the discovery of a novel constitutive promoter warranted further investigation. This thesis is an exploration of the *hex1* gene, its pattern of expression and role in *T. reesei* physiology. The following sections describe the current knowledge of the HEX1 protein in fungal biology.

1.7 The HEX1 protein and Woronin bodies

The HEX1 protein has been previously isolated from Woronin bodies of *N. crassa* (Jedd *et al.*, 2000; Tenney *et al.*, 2000). Woronin bodies are organelles unique to filamentous fungi and prior to the discovery of HEX1, the composition and functioning of the Woronin body was undetermined. Early observations of Woronin bodies by light microscopy showed them as globular or oval in shape. More recently, studies using electron microscopy have made it clear that Woronin bodies are in fact hexagonal rods and the shape visualised by microscopy is dependent on the axis of symmetry observed. Woronin bodies are often associated with septa in filamentous fungi (Figure 1-5).

1.7.1 The function of Woronin bodies

Highly refractive particles located near the septum were first recorded in fungi by the Russian mycologist Michael Stepanovitch (1838-1903) by light microscopy and they were named Woronin bodies by Buller in 1933. Woronin bodies have only been identified in filamentous ascomycete and deuteromycete fungi. Even though early observations were restricted by the limitations of light microscopy it was thought that Woronin bodies may act as plugs for septum in response to rupture of the hyphae

(Markham *et al.*, 1987). It was not until developments in molecular biology and electron microscopy advanced that more became known about Woronin bodies.

Two groups independently and concurrently published research confirming the composition and functioning of the *N. crassa* Woronin Body in 2000 (Jedd and Chua, 2000; Tenney *et al.*, 2000). Each group used a slightly different method in the purification of Woronin bodies but both generally involved centrifuging homogenised cells through a sucrose gradient with a series of washes and lysis steps. The resulting

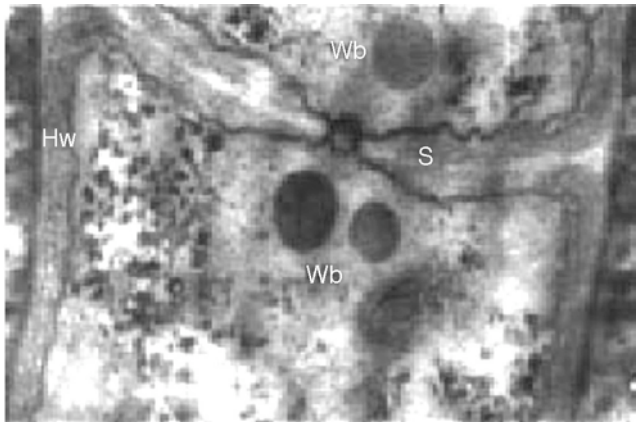

Figure 1-5. Woronin bodies surrounding the septal pore within a hyphae of *Glomerella cingulata* f. sp. *phaseoli*. Hw, hyphal wall, S, septa, Wb, Woronin body. TEM, 37.35x (Roca *et al.*, 2000).

dense core pellet was resolved by SDS-PAGE to reveal a 19 to 20 kDa protein amongst others. The band was subjected to protein sequencing and the results were used to search against the *N. crassa* sequence database. The *hex1* gene, which encoded for a 19 kDa protein was identified. This protein was named HEX1 and *hex1* knockout mutants were generated to prove the association between the protein and Woronin bodies.

The *hex1* deletion mutants, which lacked the presence of Woronin bodies on light microscopy observation, could not prevent cytoplasmic bleeding after damage to the

hyphae. A translational fusion of the fluorescent reporter DsRed2 to HEXA, expressed in *Aspergillus oryzae,* resulted in red fluorescent Woronin bodies at the septal pore adjacent to apical cells which were lysed by hypotonic shock (Maruyama *et al.*, 2004). In a *hex1* deletion mutant, the protoplasm leaked from a damaged hyphal tip extends several cell volumes along the hyphae (Yuan *et al.*, 2003). The ability to plug septal pores is therefore a vital survival mechanism for filamentous fungi as the protoplasm of the mycelium is a continuum.

Woronin bodies are usually randomly dispersed in young hyphal cells but are directly associated with the cell wall or septa in older cells (Brenner *et al.*, 1968; Momany *et al.*, 2002). There are some reports that the Woronin body is surrounded with a meshwork, which is continuous with the cell wall, or with a membrane which is continuous with the cell membrane (McKeen, 1971; Trinci *et al.*, 1974).

The plugging of the septal pore by a Woronin body has been observed by microscopy in numerous species. It is unclear how the Woronin body is transported to the septal pore in reaction to cytoplasmic leakage from the adjacent cell. The view generally held is that translocation to the septum by the Woronin body occurs passively with the movement of protoplasm on the release of turgor pressure within the damaged hyphal cell. This is possible since the release of cell contents and the movement of organelles through the hyphae is rapid. A 1-2 μm vacuole has been recorded to travel at approximately 200 μm per minute with bulk protoplasmic flow and cell leakage continued for up to 15 min in a damaged *hex1* deletion mutant (Tenney *et al.*, 2000). However, this theory has not been proven and other lines of evidence suggest the contrary. Considering there are multiple Woronin bodies associated with each septum it

is unlikely that only one should plug the septum if movement was by passive cytoplasmic flow. Momany and colleagues (2002) reported ribosome-free areas surrounding Woronin bodies as if they were connected in a meshwork with the septum and reported possible fibrous connections between the Woronin body and the septum. This evidence may be an indication that intracellular movement of Woronin bodies is more precisely directed by transportation along the cytoskeleton.

Complete plugging of the septum adjacent to the site of damage by a Woronin body can occur in under 3.6 seconds in *Penicillium chrysogenum* (Markham and Collinge, 1987). The Woronin body which plugs the septum enters from the undamaged cell and the septum is almost always blocked by a single Woronin body (Roca *et al*., 2000). Septa several cells back from the site of damage are also blocked by a Woronin body, but this response is slower than the immediate blockage of the neighbouring septa (Collinge *et al*., 1985).

The number of Woronin bodies surrounding septa does not appear to be correlated to the environmental conditions, even though cell damage may be more likely in some instances. The hyphae of *Penicillium janczewskii zaleski* have an average cell wall thickness of 384 nm when grown on solid medium containing inulin. When they are grown on sucrose, the cell wall is significantly thicker at 800 nm. The thinner hyphae have been observed to be more fragile and easily crushed during preparation for electron microscopy. However, the average number of Woronin bodies remained two to four per side of the septum, regardless of the fragility or robustness of the hyphal walls (Pessoni *et al*., 2002).

Once the septum is plugged by a Woronin body, further reinforcement of the septum commences. Additional material similar to that of the septal wall is deposited around the Woronin body such that it becomes fused with the septum wall on the undamaged cell side. This process was completed after three hours in *Penicillium chrysogenum* (Markham and Collinge, 1987). In *N. crassa,* the material stained positive for chitin (Jedd and Chua, 2000). There is no further detail in the literature concerning the binding of the Woronin body to the septum wall or the deposition of chitin. It is also unknown if the HEX1 lattice is exocytosed from the Woronin body membrane for plugging of the septum or if the membrane remains intact and becomes embedded within the suggested chitin deposition.

1.7.2 The intracellular localisation of HEX1

The purification of Woronin bodies from *N. crassa* led to the identification of the HEX1 protein. Alignment of homologous gene sequences from other fungi revealed that all HEX1 sequences have a Type 1 peroxisomal targeting signal (PTS1) on the carboxy terminus (Jedd and Chua, 2000). Once translated in the cytosol, the HEX1 protein is concentrated within a microbody where individual molecules self assemble into a dense hexagonal crystal lattice through polymerisation (Yuan *et al*., 2003). This has been observed in *A. nidulans* (Momany *et al*., 2002).

When Tenney and colleagues (2000) analysed *N. crassa* cell extracts concentrated for Woronin bodies, they identified several protein bands of 17, 19, 40 and 80 kDa in size which all labelled with the anti-HEX1 polyclonal antibody. This indicated that HEX1 oligomerises but it was not until the crystal structure of HEX1 was solved that the formation of a Woronin body core could be described.

1.7.3 The assembly of Woronin bodies

Purified HEX1 protein readily crystallises *in vitro* to form hexagonal rod shaped crystals (Jedd and Chua, 2000). Recently, the structure of the HEX1 protein from *N. crassa* has been solved and the nature of the formation of the Woronin body has been described (Yuan *et al*., 2003) (Figure 1-6). The HEX1 protein is divided into distinct N- and C-terminal domains each structured as antiparallel β-barrels. The N-terminal domain contains six antiparallel strands and one helix composed of four residues. The C-terminal β-barrel contains five strands and two helices.

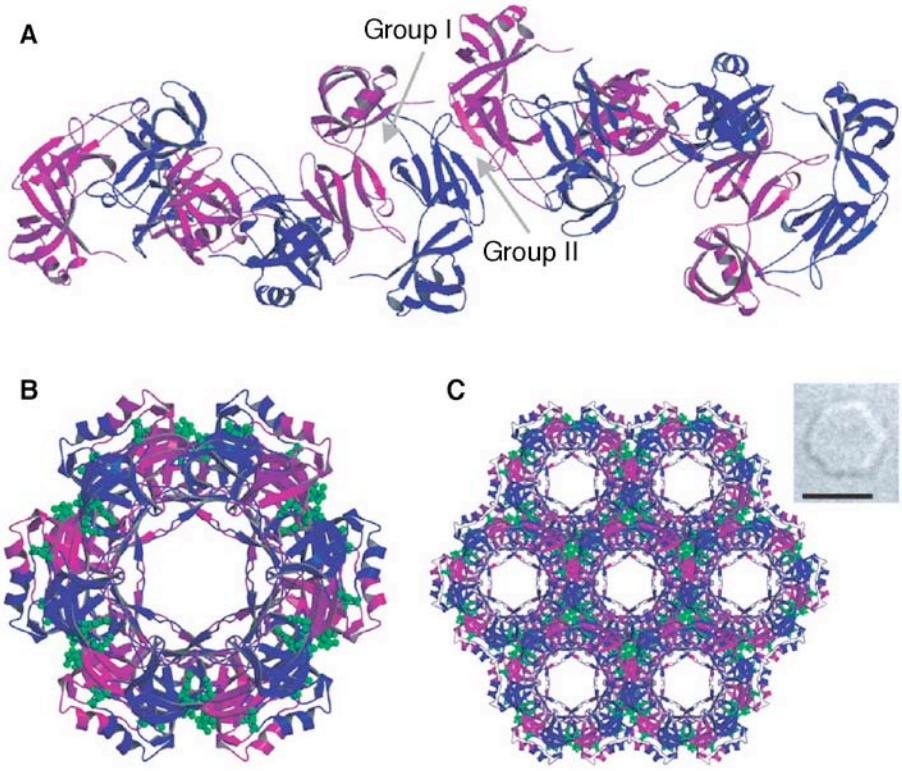

Figure 1-6. Organisation of HEX1 molecules within a Woronin body. A. Polymerisation of HEX1 molecules into a long helical spiral. Alternating HEX1 molecules are shown in purple and blue with alternating Group I and Group II interactions shown. B. Cross section view of HEX1 helical spiral filament. The residues which form Group III interactions are shown in green. C. Cross section view of seven HEX1 spiral filaments bound by Group III interactions to form the hexagonal crystal structure of the Woronin body. Insert is an EM image of a Woronin Body from *N. crassa*. Figure taken from Yuan *et al*. (2003).

A combination of hydrogen bonds and salt bridges between the residues Asp44 to Arg41 and His39 and hydrogen bonds between Ser92, Glu121 and Gln105 are collectively termed Group I interactions. Salt bridges between residues Arg68 and Glu81 and hydrogen bonds between Gln49 and Ser84, and Ser61 and Glu81 are collectively termed Group II interactions. The polymerisation of HEX1 occurs due to the alternating Group I and Group II interactions between adjacent HEX1 molecules. It is not stated how many HEX1 molecules in total form each full length spiral filament. Residues Gln127, Val125, Ile56 and Gln134 are accessible on the outside surface of the helical spiral and form hydrogen bonds between neighbouring HEX1 helical filaments as Group III interactions. The complete crystalline structure of a Woronin body core is composed of seven helical filaments tightly bound by Group III interactions in the conformation shown in Figure 1-6.

Mutations of residue His39 in *N. crassa* HEX1 did not affect the folding of the HEX1 molecule but prevented the polymerisation of HEX1 *in vitro* by disrupting the necessary □-interaction and hydrogen bond with the adjacent HEX1 molecule. When a *N. crassa* HEX1 deletion mutant was transformed with this His39G HEX1, Woronin bodies still formed but were spherical in shape without the hexagonal crystalline structure. These spherical Woronin bodies also failed to prevent cytoplasmic leakage indicating they were dysfunctional. Therefore the solid HEX1 crystalline lattice is vital for the Woronin body to plug septal pores against the turgor pressure generated in lysed hyphae (Yuan *et al.*, 2003).

The uptake of HEX1 molecules into the Woronin body and the subsequent polymerisation and crystallisation to form a Woronin body may have been observed indirectly in *Aspergillus nidulans* (Momany *et al.*, 2002). Cryosections of young *A.*

nidulans hyphae were gold immunolabelled using HEX1 polyclonal primary antibodies. Transmission electron micrographs showed a difference in the labelling pattern in Woronin bodies near the germ tube apex and those near a mature septum, suggesting different protein organisation between the two locations. Gold particles in apical Woronin bodies were more diffuse at the periphery of the electron dense core whilst septal Woronin body labelling was concentrated within the electron dense core.

1.8 Fluorescent reporter proteins

The use of fluorescent reporter proteins commenced with the identification of the green fluorescent protein from the jellyfish *Aequorea victoria* and the cloning of its gene (Prasher *et al*., 1992). The number of mutants available and the number of applications have been steadily increasing. Fluorescent reporters are now widely employed as biosensors and in studies into gene transcription and protein localisation. Fluorescent reporters have the advantage over other detection systems because they require no addition of co-factors and can be imaged *in situ*. The fluorescent reporters are also available with a range of spectral properties to allow multi-labelling within one living system.

Passive applications of fluorescent reporter proteins include the labelling of cells, the labelling of proteins expressed as a translational fusion tag, or the use as a reporter of gene expression. GFP has also been used as an indicator of the production phase of heterologous protein on an industrial scale (Albano *et al*., 1998; Baker *et al*., 2002) and for various studies in fungal biology (summarised in Jensen *et al*., 2004). In particular, it has been a useful marker for studies into protein localisation and secretion, and for studies in parasitic interactions (reviewed by Cormack, 1998; Lorang *et al*., 2001;

March *et al*., 2003; Jensen and Ashulz, 2004). Further examples of the use of fluorescent reporters in fungal biology are summarised in Table 1-5.

Table 1-5. Examples of the uses of fluorescent reporter proteins in fungal biology.

Host organism	Fluorescent reporter protein	Function	Reference
Aspergillus nidulans	sGFP	Protein localisation. Fusion to chitinase export signal and ER retention signal.	Fernandez-Abalos *et al*. (1998)
A. flavus	EGFP	Whole organism visualisation. Transcription reporter.	Du *et al*. (1999)
A. niger	sGFP	Protein fusion to GLA.	Gordon *et al*. (2000a)
A. nidulans	DsRed (T4) mRFP1 BFP	Nuclear and mitochondrial localisation. Creation of fusion vectors for GATEWAY.	Toews *et al*. (2004)
Phanerochaete chrysosporium	EGFP	Transcription reporter.	Ma *et al*. (2001)
Trichoderma harzianum	sGFP	Transcription reporter during mycoplasmic interaction.	Zeilinger *et al*. (1999), de las Mercedes Dana *et al*. (2001)
T. harzianum	EGFP	Whole organism visualisation.	Inglis *et al*. (1999)
Ustilago maydis	sGFP	Whole organism visualisation.	Spellig *et al*. (1996)
Aureobasidium pullulans	EGFP	Transcription reporter.	Wymelenberg *et al*. (1999)
Neurospora crassa	EGFP	Transcription reporter.	Freitag *et al*. (2001)
Ophiostoma piliferum	sGFP	Whole organism visualisation.	Lee *et al*. (2002)
Fusarium oxysporum F. sp. lycopersici	DsRed2	Whole organism visualisation.	Nahalkova *et al*. (2003)
Penicillum paxilli *T. harzianum* *T. virens*	DsRed-Express	Whole organism visualisation.	Mikkelsen *et al*. (2003)

GATEWAY, rapid cloning method for DNA fragments, without restriction digestion and ligation steps. sGFP, synthetic GFP is the original developed by (Chiu *et al*., 1996) with S65T and F64L substitutions and is now commercially available as enhanced green fluorescent protein (EGFP) (BD Biosciences Clontech). The DsRed(T4) mutant has 9 substitutions and reduced lag-time in maturation of the fluorophore compared to the wild-type. The DsRed2 mutant has an improved maturation time compared to the wild-type. Monomeric red fluorescent protein (mRFP1) has a reduced tendency to aggregate within a cell. DsRed-Express has a rapid fluorophore maturation time.

Other applications of fluorescent reporter proteins take advantage of certain physical and biochemical properties of the fluorophore and adapt these to yield specific

experimental data. GFP mutants have been used as bioindicators of intracellular and intra-organelle pH (Olsen *et al.*, 2002), oxygen levels (Albano *et al.*, 2001) and temperature (Naik *et al.*, 2001). Fluorescent reporters can also be targeted to specific organelles by the fusion with the necessary localisation signal (De Giorgi *et al.*, 1996). So far there are commercially available fluorescent reporters targeted to the ER, Golgi, cell membrane, nucleus, vacuoles, cytoplasm, peroxisomes, mitochondria and cytoskeleton (www.bdbiosciences.com/clontech/) or for secretion (El Meskini *et al.*, 2001). The only limitation to intracellular localisation of fluorescent reporter proteins seems to be the identification of specific targeting signal sequences.

Autofluorescence of the material being studied can be a problem when using fluorescent probes. Autofluorescence can be caused by excitation of phenols, tannic acids, cell wall compounds, flavins, NADH and NADPH and lipofuscins which have broad excitation and emission spectra (Knight *et al.*, 2001; Jensen and Ashulz, 2004). The level of autofluorescence can be affected by growth conditions of the sample, fixation method and mounting medium used (Graf *et al.*, 1998). The intensity of autofluorescence can be minimised by a number of methods. These include the correct use of optical filters, adopting techniques such as dual-wavelength correction, manipulation of data with image software (Knight and Billinton, 2001), or selecting a fluorescent reporter with an emission spectrum that does not overlap with the autofluorescence spectrum (Jensen and Ashulz, 2004).

1.8.1 Different mutants of the red fluorescent reporter

DsRed is a fluorescent protein isolated from the sea anemone (*Discosoma sp.*). There are a number of improved mutant alleles commercially available including DsRed1-E5,

commercialised as the fluorescent timer (BD Biosciences Clontech, USA). This mutant allele was produced by error prone PCR amplification from the wild-type gene which resulted in two substitutions, V105A and S197T. The first substitution is responsible for a two-fold increase in quantum yield of fluorescence and the S197T substitution is responsible for a spectral shift in the maturing fluorophore, such that the emitted fluorescence changes from green (excitation 483 nm, emission 500 nm) to red (excitation 558 nm, emission 583 nm) over time (Terskikh *et al*., 2002). This change in spectral shift also occurs in the wild-type DsRed when the fluorophore passes through an intermediary conformation, similar to EGFP, that emits green fluorescence. In DsRed1-E5, the transition from the green emitting fluorophore, to the mature red emitting conformation is delayed due to the S197T substitution (Gross *et al*., 2000). This change in spectral properties potentially can offer a greater insight into protein expression compared to the continuous fluorescence of a nonchanging fluorescent reporter. Green fluorescence only is indicative of recently induced protein expression whilst the presence of red fluorescence alone is indicative of completed protein expression. The concurrent presence of green and red fluorescence signifies continuous expression. Like other fluorescent reporters, the only requirement for DsRed1-E5 fluorescence is oxygen.

Other DsRed variants include DsRed2 and DsRed (T4). DsRed2 which has increased brightness and solubility compared to the wild-type, and was used successfully in *Fusarium oxysporum* f. sp. *Lycopersici* under the *A. nidulans gpdA* gene promoter to trace the growth of the parasitic fungal hyphae through tomato seedling roots (Nahalkova and Fatehi, 2003). The DsRed(T4) variant has a rapid maturation time and

has been applied in fungi to localise the transcription factor StuA to the nucleus in *A. nidulans* (Toews *et al*., 2004).

An important feature of fluorescent reporters, which can limit their usefulness is the length of time taken from transcription to the visualisation of a mature, folded fluorophore. The time taken can be variable depending on the system in which it is expressed. After the induction of DsRed1-E5 expression, the first appearance of green fluorescence occurred after six to nine hours in HEK293 mammalian cells and two hours in *C. elegans* (Terskikh *et al*., 2000). The time then to convert from the green fluorophore to the red fluorophore was three hours in HEK 293 mammalian cells and two hours in *C. elegans* (Terskikh *et al*., 2000). Red fluorescent proteins with faster maturation times have been created, namely DsRed variant E57 which takes three to four hours for fluorescence to appear after induction but the intensity of the fluorophore is reduced compared to variants that mature slower (Terskikh *et al*., 2002).

For this study into monitoring promoter activity and heterologous protein expression, the DsRed1-E5 fluorescent timer variant was the fluorescent reporter of choice. The additional information provided by the changing spectral properties of this variant were considered to be advantageous over the other types available. To our knowledge the work carried out by Curach *et al*. (2002) was the first to describe the use of any DsRed variant in filamentous fungi.

1.8.2 Methods for the detection of fluorescent reporters

The applications for fluorescent reporter proteins are ever increasing due to technological advances in microscopy and image analysis software. For instance,

fluorescently labelled plant, bacterial and mammalian cells are often analysed and sorted by fluorescent-activated cell sorting (FACS) (Galbraith *et al*., 1995; Cormack *et al*., 1996).

Fluorescence reporter proteins can also be used to study intermolecular interactions by fluorescence resonance energy transfer (FRET). FRET involves the use of two different fluorophores where the emission spectra of one (the donor) overlaps with the excitation spectra of the second (the acceptor). When the two fluorophores are in close proximity (<100 Å apart) FRET can occur and is analysed by the ratio of the emission intensity from the donor and acceptor. FRET has been used to study protease activity on specific peptide sequences by separating the two fluorophores BFP and GFP by a peptide linker. Protease recognition of the peptide and cleavage was confirmed by the release of the fluorophores and very low acceptor emissions (Heim *et al*., 1996). FRET has also been applied to analyse molecule dimerisation and ligand receptor interactions (Romoser *et al*., 1997).

Confocal laser scanning fluorescence microscopy (CLSM) increases the resolution capability of conventional fluorescent microscopy (Jensen and Ashulz, 2004). In fluorescence microscopy, the excitation lightpath is generated from a mercury lamp and excitation and barrier filters are used to narrow the excitation and emission wavelengths. The focal plane of excitation is broad and flare from out of focus positioned chromatophores reduces the resolution of the image captured. Confocal laser scanning microscopy is much more advantageous for fluorescent tissue observation. Excitation by a He/Ar laser is focused onto a particular x/y coordinate of the specimen and the image is scanned on a single plane which prevents exposure of the entire sample to photobleaching. Images can also be acquired by scanning in a vertical (z) direction to

create a three dimensional image. The images are recorded on imaging software which allows various degrees of analysis and manipulation.

In addition to CLSM, new microscopy techniques have been developed for improved application to the growing use of fluorescent proteins. Two-photon laser-scanning microscopy (TPLSM) uses pulsed infra-red laser light to excite the fluorophore which is normally exited by visible light. In this method, two photons will only hit a fluorescent molecule simultaneously at the exact point of focus so no fluorescence out of the focal plane is excited. This technique reduces photobleaching and provides improved sensitivity and resolution compared to CLSM (Potter *et al.*, 1996). In addition to a CLSM, additional specialty equipment such as a titanium/sapphire laser and an argon-ion pump laser are required. Other improved confocal microscopy methods include total internal reflection fluorescence microscopy (TIR-FM). In this technique, the focal plane of TPLSM is decreased from 500-800 nm to less than 100 nm, which allows much finer resolution and is reportedly capable of single molecule detection (Toomre *et al.*, 2001). However, a specialised setup is required to employ these techniques.

1.9 Aims of this project

The primary aim of this thesis was to investigate the activity and regulation of the *hex1* gene promoter with consideration for the potential of this promoter to regulate the expression of economically relevant proteins in the industrially exploited filamentous fungus, *T. reesei*. The regulation of the *hex1* promoter was monitored through *hex1* transcript analysis as well as through the expression of the fluorescent reporter protein DsRed1-E5. In the course of this work, the usefulness of the DsRed1-E5 reporter protein in fungi was also explored and assessed. As information on the regulation of the

hex1 gene promoter became apparent, the relevance of the findings were placed in the context of the role of *hex1* in the biology of the fungal mycelium. With this knowledge, the *hex1* gene was targeted for manipulation of the mycelium physiology for the enhancement of protein production in *T. reesei*.

Chapter 2 Materials and Methods

2.1 Frequently used reagents and solutions

All chemicals used were of analytical, biochemical or molecular grade and were purchased from Sigma (Australia) unless otherwise stated. Agarose and SDS were purchased from Amresco (USA). The reagents, solutions and media used throughout this work were prepared using H_2O filtered through the Millipore Milli-Q Academic filtration system and sterilised by autoclaving at 121°C for 20 min per L. Some of the commonly used solutions are listed in Table 2-1. The quantities provided as a percentage are in w/v unless otherwise stated.

Table 2-1. Solutions frequently used throughout experimental work for this thesis.

Alkaline phosphatase buffer	15.8 g Tris-HCl and 5.8 g NaCl in 1 L H_2O, pH 9.5
Blocking solution	1% (w/v) Blocking Reagent (Roche, Germany) in Maleic acid buffer
LB	1% Tryptone, 0.5% yeast extract and 1% NaCl, pH 7.0
Citrate buffer	50 mM citric acid solution was added to 50 mM sodium citrate solution until the pH reached 5.0
Maleic acid buffer	11.6 g Maleic acid and 8.8 g NaCl in 1 L H_2O, pH adjusted to 7.5 using NaOH pellets
$MgSO_4$-sodium phosphate buffer	1.2 M $MgSO_4$, 10 mM NaH_2PO_4 solution was added to 1.2 M $MgSO_4$, 10 mM Na_2HPO_4 solution, until the pH reached 5.8
PBS (20x)	160 g NaCl, 4 g KCl, 28.8 g Na_2HPO_4 and 4.8 g KH_2PO_4 in 1 L H_2O, pH 7.4
SSC (20x)	175.3 g NaCl and 88.2 g sodium citrate in 1 L H_2O, pH 7.0
TBE (10x)	90 mM Tris-HCl, 1.25 mM EDTA, 90 mM Boric acid, pH 8.4
TBS (1x)	50 mM Tris base and 150 mM NaCl, pH 8.0
TE (10x)	100 mM Tris-HCl and 10 mM EDTA, pH 8.0
Washing buffer	Maleic Acid buffer with 0.3% (v/v) Tween 20 (BDH)

2.1.1 Antibiotics and reagents for the selection of recombinants

Recombinant *E. coli* were identified based on blue white selection from the inclusion of X-Gal (Progen, Australia) at 20 μg/mL and IPTG (Progen, Australia) at 0.4 μg/mL, in the growth medium. Ampicillin was used at a concentration of 100 μg/mL to screen for recombinants.

Hygromycin B (Calbiochem, USA) was used at 60 U/mL for selection of fungal recombinants.

2.2 Fungal growth media

2.2.1 Minimal medium

The preparation of minimal medium was as described by Penttilä *et al.* (1987). To make 1 L, 15 g KH_2PO_4, 5 g $(NH_4)_2SO_4$, 10 mL 100 x trace elements (50 mg $FeSO_4 \bullet 7H_2O$, 16 mg $MnSO_4 \bullet H_2O$, 14 mg $ZnSO_4 \bullet 7H_2O$, 20 mg CoCl in 100 mL dH_2O) and 2 mL 10% (v/v) Tween 80 (BDH Chemicals Ltd., England) was added to 892 mL H_2O and the pH adjusted to 5.5 with 5 M KOH. This minimal salt solution was sterilised by autoclaving before adding 100 mL 20% D-(+)-glucose, 2.4 mL 1M $MgSO_4$ and 5.4 mL 1M $CaCl_2$, which had been pre-sterilised separately. For solid medium, agar (Spectrum Chemicals, USA) was added to the minimal salt solution to a final concentration of 2%.

2.2.2 Cellulose medium

Cellulose medium is based upon minimal medium with cellulose added as an alternative to glucose (to induce *cbh1* gene expression). The following procedure was carried out to

prepare multiple flasks of this medium to ensure equal distribution of the insoluble components. Fifty millilitres of minimal salt solution (Section 2.2.1) was distributed into 250 mL flasks followed by the addition of Avicel-cellulose (2%) (Fluka Biochemika, Switzerland) and soy bean flour (Type I) (1.5%). After sterilisation of the flasks, 120 µL of 1M $MgSO_4$ and 270 µL 1M $CaCl_2$, which had been pre-sterilised, were added. Glucose was omitted and replaced by sterile □-lactose (1%).

2.2.3 Cellobiose/Lactose/Soy hydrolysate medium

Cellobiose, lactose, soy hydrolysate (CLS) medium has been described previously (Lim *et al.*, 2001). This medium is based upon minimal medium supplemented with a soluble carbon source that induces *cbh1* expression. The components of the minimal salt solution were prepared as in Section 2.2.1 with the addition of D-(+)-cellobiose (1%). After sterilisation, □-lactose (1%) and soy bean flour extract (3%) were added instead of glucose. The soy bean flour extract was pre-prepared by sterilising a 3% soy bean flour solution by autoclaving and then removing the insoluble components by centrifugation at 9000 x *g* for 20 min.

2.2.4 Glycerol medium

Glycerol medium is neutral in respect to induction or repression of *cbh1* gene expression (Ilmén *et al.*, 1997; Margolles-Clark *et al.*, 1997). The components of glycerol medium were prepared as for minimal medium (Section 2.2.1) with glucose substituted by glycerol (2%).

2.2.5 Protoplast regeneration medium

Protoplast regeneration medium was prepared as for minimal medium but supplemented with a final concentration of 1 M sorbitol as an osmotic stabiliser. To make plate medium, bacteriological grade agar (Oxoid, UK) was added to a final concentration of 1.8%, and for the overlay agar the concentration was 2.5%.

2.3 Fungal strains and culture conditions

2.3.1 Fungal strains

The high cellulase-secreting strain of *Trichoderma reesei*, VTT-D-79125 was derived by UV mutagenesis (Bailey and Nevalainen, 1981).

Ophiostoma floccosum strain J2026 was isolated from wood chip piles by Joanne Thwaites in Kinleith New Zealand.

2.3.2 Culture conditions

Cultures were maintained on Potato Dextrose Agar (PDA) plates (Oxoid, UK). Inoculated plates were incubated at 28°C in the dark for 3-4 days, until the mycelium had covered the plate, then moved under fluorescent lighting at 22°C to encourage conidiation. After 7-10 days, conidia were collected by flooding the plate with 8 mL 0.9% NaCl, 0.01% Tween 80 solution and gently abraded with a glass rod spreader. The conidial suspension was filtered through absorbent cotton wool to remove any hyphae and the quantity of conidia within the suspension was counted using a haemocytometer. Conidial suspensions were used immediately or frozen at –20°C for medium term storage until required.

All liquid cultures were inoculated with freshly harvested conidia. Liquid cultures were grown in flat bottom conical flasks at 28°C with shaking at 240 rpm. Cultures destined for fluorescence analysis of the mycelia or supernatant, were grown in the dark to limit any effect of photobleaching of the fluorescent reporter protein.

2.3.2.1 Inhibition of protease activity

Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets (Roche, Germany), were added at the dosage rate of one tablet per 100 mL culture volume where stated.

2.3.3 Determination of *Trichoderma reesei* growth rate on different carbon sources

Conical flasks (250 mL volume) containing 50 mL of the relevant medium were set up in triplicate and inoculated with 3 x 10^8 conidia. Cultures were grown at 28°C and 240 rpm for 24 h, 54 h, 74 h, 94 h, and 154 h. At each time point, the mycelia from two flasks were collected on preweighed GF filter papers (Whatman, England) by vacuum, washed with 1 volume H_2O, then freeze dried. The mycelial dry weight was calculated by subtracting the weight of the filter paper. The mycelium of the third flask was collected for extraction of total RNA (refer to Section 2.9).

2.4 General molecular techniques

DNA manipulations were as described by Sambrook and Russell (2001). All plasmids were propagated in *Escherichia coli* strain DH5 cultured at 37°C, except recombinant DH5 cells carrying plasmids containing 2.4 kb of the *hex1* gene promoter, which were cultured at 28°C. Plasmids were extracted using the Qiagen QIAprep spin miniprep kit

(Germany) following the manufacturers instructions. The Qiagen QIAfilter plasmid maxi kit was used for large-scale plasmid preparations.

PCR products and restriction digest products were purified using the Qiagen QIAquick PCR purification kit when appropriate. DNA bands excised from agarose gels was purified using the Qiagen QIAquick gel extraction kit as directed by the manufacturer.

2.4.1 Blunt-ending of DNA fragments with 5' or 3' overhangs

The enzymes T4 DNA Kinase and T4 DNA Polymerase required for DNA manipulations were supplied by Roche (Germany). The blunt-ending reaction for DNA fragments with incompatible overhangs contained 400 ng of cut plasmid, 2.4 U T4 DNA kinase, 1.25 U T4 DNA polymerase, 1 mM ATP, 0.2 mM dNTP mix (BioLine, USA), 1x end repair buffer (33 mM Tris-acetate [pH 7.8], 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT) and H_2O to 50 μL. The reaction mix was incubated at 37°C for 1 h. The blunt-ended DNA was purified by a phenol/chloroform extraction (Sambrook and Russell, 2001) and resuspended in 10 μL H_2O.

2.4.2 Restriction endonuclease digestion

All restriction enzymes were obtained from Fermentas (USA) unless otherwise stated. Restriction digests of DNA were carried out in the commercially provided buffers and following the manufacturer's instructions.

2.4.3 Ligations

Ligation reactions were carried out using either the Fast-Link DNA ligation kit (Epicentre Technologies, USA) or T4 DNA ligase purchased from Roche (Germany), following the manufacturer's instructions.

2.4.4 Phosphatase treatment of DNA fragments

Linearised plasmids were treated with Shrimp Alkaline Phosphatase (Boehringer, Germany), to prevent self-ligation, when required. Reactions were carried out in accordance with the manufacturer's instructions in the buffer supplied. The reaction was terminated by incubation at 70°C for 10 min.

2.4.5 Polymerase chain reaction

A number of PCR systems were used throughout this work. The Herculase Hotstart DNA Polymerase (Stratagene, USA), MasterTaq Kit (Eppendorf, Germany), Triple Master PCR system (Eppendorf, Germany) and Platinum *Pfx* DNA polymerase (GibcoBRL Life Technologies, USA) were all used following the protocols supplied by the manufacturers. PCR was carried out on the Perkin Elmer Gene Amp 2400 PCR System.

A standard PCR reaction used the AmpliTaq Gold polymerase system (Applied Biosystems, USA) and contained 1 unit of polymerase, 10 pmol of each primer, 3 mM $MgCl_2$, 12.5 mM dNTPs, approximately 10 ng of template DNA and H_2O to 50 μL. Standard reaction conditions were 94°C, 10 min, 35x (94°C, 30 s; 45-70°C, 30 s; 72°C, 1 min/1.5 kb), 72°C, 5 min.

PCR products were separated on agarose (Amresco, USA) gels containing 0.5 μg/mL ethidium bromide and were run in 1x TBE which also contained ethidium bromide (0.05 μg/mL), unless otherwise stated. PCR products which required isolation, were excised from the agarose gel and eluted using the Qiagen QIAquick gel extraction kit (Section 2.4).

2.4.5.1 Primers for PCR

All primers were synthesised by Sigma-Genosys (Sydney, Australia). The primers used in PCR throughout this work are shown in Table 2-2.

2.4.6 DNA sequencing

DNA sequencing was carried out using the ABI Prism 377 (Applied Biosystems, USA) DNA sequencing facility at Macquarie University, Australia. DNA sequences were analysed and manipulated using programmes of the Genetics Computer Group (GCG) (Devereux *et al*., 1984) and BioManager web interface run by the Australian National Genomic Information Service (ANGIS). Other web based molecular biology analysis tools are described in the text when used.

2.5 Chromosome walking

2.5.1 Isolation of *hex1* gene consensus DNA fragment from *Trichoderma reesei* and *Ophiostoma floccosum*

Degenerate primers hex1fwd.pr ACATCTTCCAAAATGGGNTAYTAYGA and hex1rev.pr ACGGGGCCGCACATNGTYTGNAC were designed by Dr. M. Gibbs based on translation of conserved HEX1 sequences from *T. reesei* and *A. nidulans*

Table 2-2. Summary of the primers used in PCR. Features such as the primer sequence, inclusion of restriction enzyme recognition sites, the target of the PCR and the annealing temperature used, are also provided.

Primer	Sequence 5' to 3'	Restriction enzyme recognition sequences and other features	Annealing temperature (°C)
actin.fwdpr	TCAGTCCCATGGAGGGTACGTGCACCTGTC	for amplification of *act1* probe	56
actin.revpr	CTGTATCCCATGGTCCTTAGAAGCACTTG	for amplification of *act1* probe	56
cbhIfusion.seqpr	CAAGCAGCTGACTGAGATG	for amplification of *cbh1* probe	55
cbh1hphint.revpr	GCTACGTTGTCATCGTCTTGACAGCAATGC	for screening transformants	60
dsred.fwdpr	CTCTAGAGGAGAATTCGGTGTCGACCGCCACCATGGTG	*Eco*RI external to *Sal*I	60
dsred.revpr	GCTCAGTTGGAAGCTTAGGGTACCGGTCACTACAGGAAC	*Kpn*I internal to *Hin*dIII	60
dsredafl2.revpr	GCTCAGTTGGAAGCTTAGAGCTTAAGTCACTACAGGAAC	*Afl*II internal to *Hind*III	60
dsrednheI.fwdpr	CTCTAGAGGAGAATTCGGTGCTAGCCGCCACCATGGTG	*Eco*RI external to *Nhe*I	60
dsredprobe.fwdpr	CCACCGAGCGCCTGTACC	for amplification of *dsred* probe	50
dsredprobe.revpr	CTACAGGAACAGGTGGTG	for amplification of *dsred* probe	50
GWSF	GCAGGAAACAGCTATGAC	linker primer	50
GWSR	GTCCTTTGTCGATACTGG	linker primer	50
hex1prombamhi. fwdpr	GCTTGAGGATCCTTCTTGTCTTCC	*BamH*I	60
hexex1.revpr	AGAACGTGCTAGCCTGATATCCTCGTCGTCGTAGTAACCCA	*hex1* exon 2, *EcoR*V internal to *Nhe*I	60
hexhismcs.revpr	GCGGACAGCTAGCCTGATATCGTGCTCTGTGATCGTAGTG	*hex1* His region, *EcoR*V internal to *Nhe*I	60
hexinte.fwdpr	AGCCGGCATTGGCGAAGCAGGATAG	for screening transformants	60
hexintmcs.revpr	GCCGTGCGCTAGCCTGATATCAGAGCCTGCCGGCGACACGC	*hex1* intron 2, *EcoR*V internal to *Nhe*I	60
hexprobe2.fwdpr	CCTCAAGCACGGCGTCGCC	for amplification of *hex1* probe	60
hexprobe2.revpr	CCTTCATCTCAACAGCGAGC	for amplification of *hex1* probe	60

hexpromsphI.fwdpr	ATGGACAAGATGCAGCATGCGCTCGGACGG	*SpH*I	60
hexss.revpr	CACGCGATATCACGTGAGATCTGATCACTTAAGCTAGCCTGGCGCGACTGAGCACGAGCTGTG GCCAAGAAGGCCGAGATGACGGCCAACTTCCGATACATCTTGGCGGATTGTTCGGGGACTTG	*Nhe*I, *Afl*II, *Bcl*I, *Bgl*II, *Pml*I and *Eco*RV	60
hextermpmlI.fwdpr	GAGATGAAGCACGTGTAGTGATCTCGCCTGTA	*Pml*I	50
hextermsphI.revpr	GAGTCCTGAATTCTGCATGCAAGATATACACGCC	*SpH*I internal to *Eco*RI	50
hphbgl2.revpr	CGGGGATCCAGATCTCATGCATCTATTC	*Bgl*II	50
M13R	CAGGAAACAGCTATGAC	pUC18/19 sequencing primer	50
phexcbh1ssnew.pr	TTAGTATAGCTAGCTGGCGCGACTGAGCACGAGCTGTGGCCAAGAAGGCCGAGATGACGGCCA ACTTCCGATACATGTAACCCATCTTGGCGGATTGTTCGGGGACTTGC	*cbh1* ss, *hex1* exon 2, *Nhe*I	60
pkibclI.fwdpr	CTCGACTCTTGATCACTCTATCGAGATAACGG	*Bcl*I	50
TSTR.pr	TTCTACGGGTTATGAACGGG	for amplification of *cbh1* probe	55

(Lim *et al.*, 2001) at positions 2488 and 3015 on the *T. reesei hex1* gene sequence (GenBank Accession AY517639). The degenerate PCRs, carried out by Dr. J. Te'o, yielded a 550 bp product from *T. reesei* genomic DNA and a 747 bp product from *O. floccosum* genomic DNA. These DNA sequences were used to design primers for chromosome walking PCR.

2.5.2 Generation of genomic DNA linker libraries for chromosome walking PCR

The chromosome walking technique (Morris *et al.*, 1995) was applied to isolate the remaining 5' and 3' coding regions and flanking sequences of the *hex1* genes from *T. reesei* and *O. floccosum*. Linker libraries were created from *T. reesei* and *O. floccosum* genomic DNA. The linker libraries were used as template in chromosome walking PCR. The conditions for chromosome walking PCR were as a standard PCR reaction (Section 2.4.5) using the relevant gene specific (Table 2-3 and Table 2-4) and linker specific primers with a 4 min extension time. Chromosomal walking PCR products were either sequenced directly using primers GWSF and GWSR (Table 2-2), or ligated into the pCR 2.1 plasmid of the TA cloning kit (Invitrogen, USA). The recombinant plasmids were then isolated and sequenced using primer M13R (Table 2-2).

Sequence editing and contig assembly were carried out using programmes Seqed, Gelenter, Gelmerge, Gelassemble and Gelview of the Genetics Computer Group (GCG) sequence analysis software (Devereux *et al.*, 1984).

Table 2-3. Primers designed for chromosome walking of the *T. reesei hex1* gene.

Primer	Sequence 5' to 3'
hex1rev2.gwpr	GCGTATGCTTGCACGCGTCTG
hex1rev3.gwpr	GCAACAGAAGTGTGCTCGTCG
hex1rev4.gwpr	GTGTAGCGGCGAGAAGTGTCG
hex1rev5.gwpr	GGTTGACGGCTTGCGCTGGC
hex1rev6.gwpr	TAGGCGTGCTCGCTGTGTTGC
hex1rev7.gwpr	GTGGACTCCTCGGACATGATGC
hex1fwd1.gwpr	CTCCAACCCTGCCCCCAGC
hex1fwd2.gwpr	GCTACGTCACCGCCATGACCG
hex1fwd3.gwpr	CAGTACACGTTGCCTCTTGACG
hex1fwd4.gwpr	GAGATAACATGTAGGTAGCGCG
hex1fwd5.gwpr	GGCATGTAAGCAATGAGGACCG

Table 2-4. Primers designed for chromosome walking of the *O. floccosum hex1* gene.

Primer	Sequence 5' to 3'
Ohex1rev4.gwpr	GAGAAACCATAGCAACGTGCAG
Ohex1rev5.gwpr	TGATGCCCTTGGGACGGACAATCG
Ohex1rev6.gwpr	GAGATCGTGCCCAACACGTGGC
Ohex1rev7.gwpr	GACACCAGGACGGTAGCGTTCC
Ohex1fwd1.gwpr	AGTCGTCGTTCGTCTCGAGC
Ohex1fwd2.gwpr	CACCAAGAAGCTGCACGAGGAGTCG
Ohex1fwd3.gwpr	TCCACGATTGACTATGCTGCT
Ohex1fwd4.gwpr	GTGACACTCTGTAAATGTGAACCC
Ohex1fwd5.gwpr	CAGTTATGTTTGTCTCGTCTCG

2.5.3 Isolation of a consensus fragment of the *Trichoderma reesei qde-2* gene

A search of Genbank and genome sequencing projects for fungi, led to the identification of gene sequences similar to the *qde-2* gene described for *N. crassa* (Catalanotta *et al.*, 2000) from *B. graminis*, *A. nidulans*, *S. pombe* and two sequences from *M. grisea*. The protein sequences derived from the DNA sequences, were aligned using the ClustalX programme (Thompson *et al.*, 1997) to identify blocks of conserved amino acids. Based on this protein alignment, the corresponding genes were then aligned using Tranalign software (Rice *et al.*, 2000), which ensured that the codons encoding the blocks of conserved amino acids were maintained in alignment. Degenerate primers were designed based on the blocks of conserved sequences of the *qde-2* genes. The degenerate primers were: qde2fwd.degenpr 5' ATCTTCCGTGATGGAGTNTCNGARGGNCA and qde2rev.degenpr 5' TTGGTGACGCCGCGGTCNACNACNGTNCC. The components of the PCR reactions were as described in Section 2.4.5 for the AmpliTaq Gold Polymerase system and included approximately 10 ng of *T. reesei* genomic DNA. The conditions of the PCR were 1x (94°C, 15 min), [94°C, 30 s, 72°C, 30 s (with a stepdown of 1.5°C per cycle for 25 cycles); 72°C, 30 s], 25x (94°C, 30 s; 60°C, 30 s; 72°C, 30 s), 1x (72°C, 4 min).

2.6 Construction of gene expression plasmids

All gene expression plasmids were designed to facilitate homologous integration by containing extended sequences, homologous to the targeted gene locus, at the 5' and 3' termini of the expression cassettes. All fusion sites of ligated fragments were verified

by sequencing. For fungal transformation (Section 2.7), the complete DNA expression cassette was excised from the pUC19 plasmid by digestion with *Sph*I and purified.

2.6.1 Amplification and screening of DsRed1-E5-encoding DNA fragments

The DNA sequence encoding the fluorescent reporter protein DsRed1-E5, was amplified from the pTimer plasmid (BD Biosciences Clontech, USA). The DsRed1-E5 variant of the DsRed protein is abbreviated as DsRed from this point forward.

Before the construction of expression plasmids (see below), the amplified DNA fragments encoding DsRed, were first screened for functionality. *DsRed* DNA fragments were generated by PCR and digested by *Eco*RI and *Hind*III before ligation into the corresponding sites on pUC18. DH5 cells were transformed with the pUC18-*dsRed* plasmid and colonies were screened by observation under fluorescence microscopy. Bacterial recombinants demonstrating green and red fluorescence when excited by blue and green light respectively, indicated that the sequence of the *dsRed* fragment was functional. The *dsRed* fragment from a positive recombinant was excised from pUC18 by digestion with the relevant restriction enzymes and ligated into the corresponding expression vector (see below).

2.6.2 The *cbh1* promoter driven expression plasmid

The 700 bp DNA fragment encoding DsRed was amplified from the pTimer plasmid using the primers dsred.fwdpr and dsred.revpr and was screened for functionality (Section 2.6.1). The pUC18-*dsRed* plasmid from a positive bacterial recombinant was purified and the *dsRed* DNA fragment was excised by digestion with *Sal*I and *Kpn*I.

The *Sal*I/*Kpn*I fragment was ligated into the corresponding sites on the pHEN54RQ vector, which had been constructed previously (Bergquist *et al.*, 2004).

2.6.3 The *hex1* promoter driven expression plasmid

Approximately 2.4 kb of the *T. reesei hex1* gene promoter region was amplified as a fusion to the *cbh1* signal sequence (ss) using the primers hexpromsphI.fwdpr and hexss.revpr. This PCR product was digested with *Sph*I and *Sma*I and ligated into the respective sites on pUC19 to create plasmid phexP (Figure 2-1).

The *hex1* gene terminator region was amplified from *T. reesei* genomic DNA using the primers hextermpmlI.fwdpr and primer hextermsphI.revpr with the Platinum *Pfx* DNA polymerase PCR system following the manufacturer's recommendations. The terminator fragment was digested by *Pml*I and *Eco*RI and ligated into the respective sites on phexP to make phexPTm (Figure 2-1).

The DNA fragment encoding DsRed was amplified from pTimer plasmid with primers dsrednheI.fwdpr and dsredafl2.revpr and screened for functionality as described in Section 2.6.1. The functional *dsRed* fragment was excised from pUC18 by digestion with *Nhe*I and *Afl*II and ligated into the respective sites on plasmid phexPTm to make plasmid phexPRTm (Figure 2-1).

The *pki-hph* DNA fragment included the constitutively active *T. reesei pki1* gene promoter and the bacterial *hph* gene, which encodes hygromycin B phosphotransferase and confers resistance to hygromycin B (Mach *et al.*, 1994). The 1.7 kb *pki-hph* fragment was amplified from plasmid pHEN54RQ (Section 2.6.2) using primers

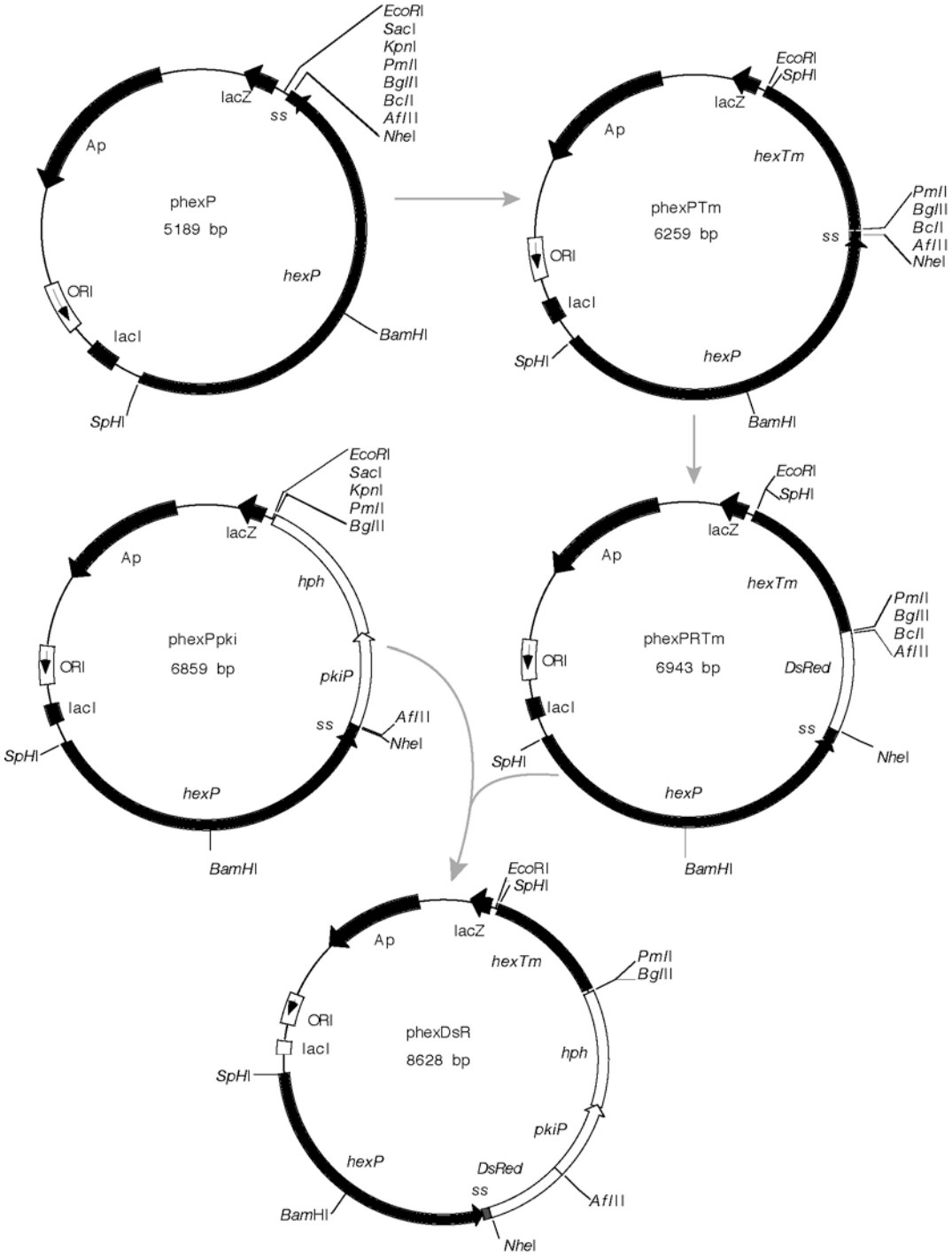

Figure 2-1. Plasmids constructed in the process of assembling the expression plasmid phexDsR. Elements derived from pUC19 include the *lacZ* promoter and N-terminal fragment of the *lacZ* gene (*lacI*, *lacZ*), bacterial plasmid replication origin (ORI) and the beta-lactamase encoding gene that confers ampicillin resistance (Ap). Elements of the phexDsR expression cassette include the *hex1* gene promoter (*hexP*), *cbh1* gene signal sequence (ss), *hex1* gene terminator region (*hexTm*), DsRed encoding sequence (*DsRed*), the *pki1* gene promoter (*pkiP*) and the hygromycin B phosphotransferase encoding gene (*hph*). See Section 2.6.3 for the description of the construction of phexDsR.

pkibclI.fwdpr and hphbgl2.revpr (Table 2-2). The amplified product was digested by the restriction enzymes *Bcl*I and *Bgl*II, which produced compatible ends. However, *Bcl*I does not digest DNA generated in *E. coli* strains with Dam^+ methylation, such as DH5α. For this reason, plasmid phexP was linearised with *Bgl*I and treated with Shrimp Alkaline Phosphatase (Section 2.4.4) prior to ligation with the *Bgl*I/*Bcl*I *pki-hph* fragment to produce plasmid pPpki (Figure 2-1). Correct orientation of the insert was confirmed by PCR. In order to confirm that the amplified *pki-hph* PCR fragment was functional, *T. reesei* was transformed with plasmid pPpki then screened on PDA plates containing 60 U/mL hygromycin B (Section 2.7). This *pki-hph* DNA fragment was removed by digestion with *Afl*II and *Bgl*II and ligated into the corresponding sites on plasmid phexPRTm to produce expression plasmid phexDsR (Figure 2-1).

2.6.4 The *hex1* fusion expression plasmids

Three expression vectors were created which incorporated different lengths of the 5' *hex1* ORF fused to the same 2.4 kb *hex1* promoter sequence, to replicate the layout of the endogenous *hex1* gene. These vectors were constructed to ascertain if any of the *hex1* 5' ORF sequences were influential on *hex1* gene expression. The *cbh1* ss was excluded from these vectors but DsRed was used again as a reporter of gene expression.

A primer (hex1prombamhi.fwdpr) was designed to anneal approximately 1.2 kb upstream from the start codon of the *hex1* gene and incorporate the recognition sequence of a unique *Bam*HI restriction site at this location. Primer hex1prombamhi.fwdpr was used in conjunction with one of three reverse primers hexex1.revpr, hexintmcs.revpr or hexhismcs.revpr (Table 2-2) in a standard reaction (Section 2.4.5). The amplified PCR products were isolated and digested by *Bam*HI and

*Nhe*I. Plasmid phexDsR was linearised by digestion with *Bam*HI and *Nhe*I and the 1.2 kb intervening fragment was removed. The *Bam*HI/*NheI hex1* fragment generated with primer Hexex1.revpr was ligated into the linearised phexDsR fragment to produce plasmid phexeDsR. The *Bam*HI/*NheI hex1* fragment amplified with primer hexintmcs.revpr was ligated into the linearised phexDsR fragment to produce plasmid phexiDsR and the *Bam*HI/*NheI hex1* gene fragment amplified with primer hexhismcs.revpr was ligated into the linearised phexDsR fragment to produce plasmid phexhDsR (Figure 2-2).

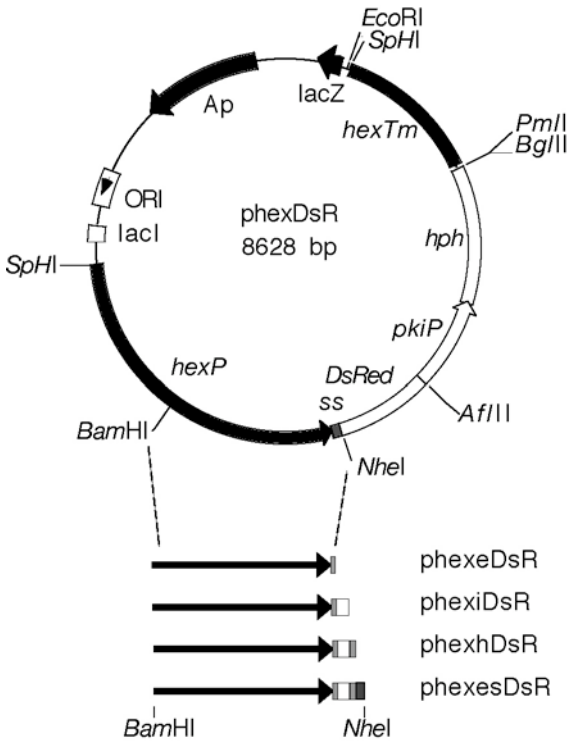

Figure 2-2. Construction of the *hex1* fusion expression plasmids phexeDsR, phexiDsR, phexhDsR and phexesDsR. Within the *BamHI*/*Nhe*I inserts, the *hex1* coding sequence, *hex1* intron sequence and the histidine-rich sequence are shown in light grey, white and dark grey respectively. Other abbreviations are the same as in Figure 2-1.

2.6.5 The *hex1* fusion expression plasmid containing the *cbh1* signal sequence

An expression plasmid that contained a fusion of the *hex1* exon 2, the *cbh1* ss and the DsRed coding sequence was constructed. A PCR using primers hex1prombamhi.fwdpr and phexcbh1ssnew.pr amplified a 1.3 kb DNA fragment from plasmid phexSRL (previously constructed by Dr. J. Te'o). The PCR components of a 50 μL reaction included 10 pmol of each primer, 1x High Fidelity buffer with Mg^{2+}, 200 μM dNTPs, 3 U Triple Master polymerase mix and approximately 10 ng DNA. The reaction conditions were 1x (96°C, 2 min), 30x (94°C, 20 sec; 60°C, 20 sec; 72°C, 2 min). The gel-eluted PCR product was digested with restriction enzymes *Bam*HI and *Nhe*I and ligated into the corresponding sites of plasmid phexeDsR, to produce plasmid phexesDsR (Figure 2-2).

2.6.6 The *hex1* open reading frame knockout plasmid

A plasmid was designed to integrate at the *hex1* locus and interrupt the endogenous *hex1* coding sequence with the *pki-hph* fragment. The plasmid phexSRL was digested with restriction enzymes *Bpl*I and *SnaB*I to remove a 1kb fragment, which included the DsRed coding sequence and 240 bp of an internal section of the *hex1* gene sequence (Figure 2-3). The 3' overhang produced by the *Bpl*I digest was blunt-ended (Section 2.4.1) and the plasmid self-ligated to produce plasmid phexnull (Figure 2-3).

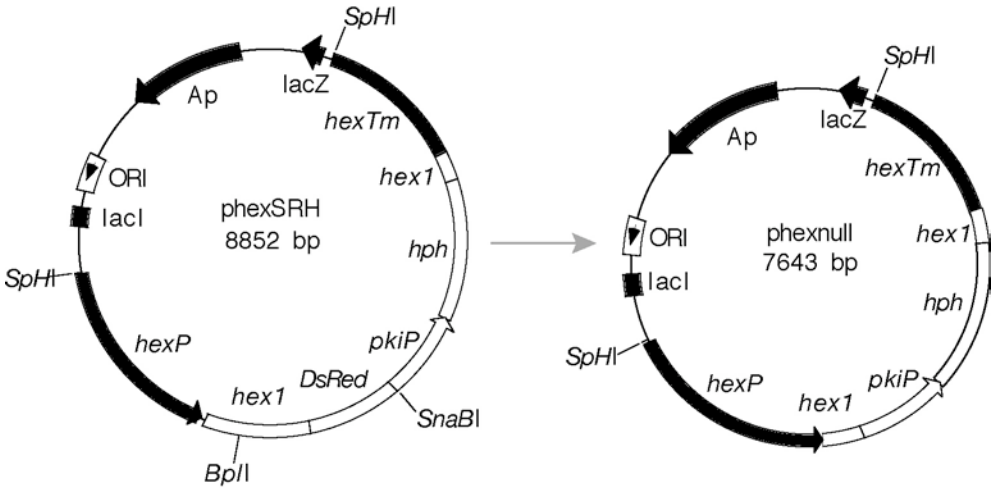

Figure 2-3. Construction of the *hex1* gene knockout plasmid phexnull.

2.7 Fungal transformation

2.7.1 Standard Biolistic Bombardment procedure

The uninucleate conidia of *T. reesei* were transformed by particle bombardment using the Bio-Rad model PDS-1000/He biolistic particle delivery system (Gene Gun) as described by Te'o *et al.* (2003). The hepta-adapter accessory was used which increases the area bombarded by seven times compared to the single barrel system (Hazell *et al.*, 2000).

2.7.1.1 Preparation of DNA coated microprojectiles

Approximately 40 mg of M-10 tungsten particles (BioRad, USA) were washed in 1 mL absolute ethanol (molecular grade) three times then washed once in 1 mL H_2O. The tungsten was resuspended in 1 mL H_2O and aliquoted into 100 μL portions. The volume of H_2O equivalent to 1.4 μg linear DNA was removed from the tungsten sample. The tungsten was kept in suspension by continuous vortexing during the following additions of DNA and reagents.

The linear DNA expression cassette (Section 0) (1.4 μg) was added to the tungsten and incubated for 1 min followed by the addition of 100 μL 2.5 M $CaCl_2$ and a further 1 min incubation. With the tungsten still maintained in suspension, 40 μL of freshly prepared 0.1 M spermidine was added and mixed for 3 min. The tungsten/DNA mixture was incubated on ice and allowed to settle for at least 10 min before collection of the pellet by centrifugation. The pellet was washed in 250 μL absolute ethanol then resuspended in 80 μL absolute ethanol.

2.7.1.2 Preparation of conidia for transformation

Freshly collected conidial suspension (Section 2.1) was plated onto seven areas of a PDA plate to align with the hepta-adapter outlets (5 x 10^7 conidia per area). The plate was allowed to dry for up to 1 h before bombarding.

2.7.1.3 Biolistic Bombardment

All components of the Biolistic Gene Gun were sterilised with 70% ethanol prior to use. The DNA coated tungsten was divided between seven macro-carrier discs (Bio-Rad, USA) and the hepta-adapter components were assembled according to the manufacturer's instructions. The appropriate rupture disk (Bio-Rad, USA) for bombardment at 1350 psi was used unless otherwise stated.

Following bombardment, plates were incubated for 6-9 h at 28°C then overlaid with 10 mL molten PDA containing 1800 U hygromycin B such that the final concentration of hygromycin B in the combined agar volume was 60 U/mL. Plates were incubated for up to 2 weeks and transformants were picked and streaked onto PDA/hygromycin B

selection plates as they appeared. Transformants which survived the second round of selection were allowed to conidiate on PDA plates (Section 2.1).

2.7.2 Standard protoplasting procedure

The protocol for the transformation of fungal protoplasts was provided by Dr. Nina Aro (VTT, Finland), which was based on the protocol from Penttilä *et al*. (1987).

2.7.2.1 Preparation of protoplasts

Freshly harvested conidia were plated onto cellophane discs laid onto PDA plates, and incubated for 16-20 h at 28°C. L2265 Lysing enzymes from *T. harzianum* (Sigma, USA) were dissolved at 5 mg/mL in $MgSO_4$-sodium phosphate buffer and filtered through a 0.45 μm filter to make lysing buffer. The hyphal mat, from approximately 20 plates, was washed from the cellophane discs into a plastic 150 mm petri dish containing 40 mL lysing buffer and incubated at RT with occasional shaking for approximately 1.5 h or until numerous protoplasts were present on microscopic examination.

The protoplasts were separated from the remaining mycelium by filtration through a sintered glass filter (porosity number 1) and rinsed with fresh $MgSO_4$-sodium phosphate buffer. An equal volume of 0.6 M sorbitol, 0.1 M Tris-HCl pH 7.5 buffer was added to the protoplast suspension and centrifuged at 4000 x *g* for 15 min at RT. Following centrifugation, most of the supernatant was carefully removed and the pellet was washed three times in 1.2 M sorbitol, 10 mM Tris-HCl pH 7.5 buffer. After the final wash, the protoplasts were resuspended in 400 μL of 1.2 M sorbitol, 10 mM Tris-HCl,

10 mM $CaCl_2$ pH 7.5. A sample of protoplasts was removed and a dilution series was plated onto nonselective medium (Section 2.2.5) as regeneration controls.

2.7.2.2 Transformation of protoplasts

Approximately 5 μg of linear DNA was added to 200 μL protoplasts, followed by the slow addition of 50 μL 25% PEG 6000, 50 mM $CaCl_2$, 10 mM Tris-HCl pH 7.5 buffer with gentle mixing and incubated on ice for 20 min. A further 2 mL of 25% PEG 6000 buffer was added and mixed by gentle inversion of the tube, then incubated at RT for 5 min. A final 4 mL of 1.2 M sorbitol, 10 mM Tris-HCl, 10 mM $CaCl_2$ pH 7.5 was added before 100 μL aliquots of the transformed protoplast suspension were added to 7 mL molten overlay agar containing 60 U/mL hygromycin B, and plated immediately onto selective agar plates (Section 2.2.5). The plates were incubated at 28°C for 7 days. Colonies were picked and patched onto second-selection PDA plates, containing 60 U/mL hygromycin B as they appeared.

2.8 Extraction of fungal genomic DNA

Genomic DNA was extracted from fungal mycelia by one of two methods. For samples to be used for mass PCR screening of transformants, the relatively rapid Bio101 FastDNA kit (Q-BIOgene, USA) was used following the manufacturer's recommendations for yeast. This method involves vigorous mechanical disruption of the cell wall which also compromises the quality of genomic DNA released.

Genomic DNA required for Southern blotting and the generation of genomic walking linker libraries (Section 2.5), was extracted using an adaptation of the technique described by Lee *et al*. (www.fgsc.net/fgn35/lee35.pdf). *T. reesei* was grown on

cellophane discs placed on PDA plates for 3 days. Mycelia were harvested, ground to a fine powder under liquid nitrogen, then incubated in Lysis Buffer (50 mM Tris-HCl, 50 mM EDTA, 3% SDS and 1% 2-mercaptoethanol) for 1 h at 65°C. An organic extraction step using an equal volume of buffer-saturated phenol (Invitrogen, USA) removed the bulk of cellular debris. Further organic extraction of genomic DNA involved one chloroform:phenol (1:1) step and one chloroform:isoamyl alcohol (24:1) extraction step, which were carried out using Eppendorf (Germany) 15 mL phase lock gel tubes (light) as recommended by the manufacturer. DNA concentration was determined by absorbance reading at 260 nm on an Eppendorf BioPhotometer.

2.9 Extraction of fungal total RNA

The mycelium from liquid cultures were collected by centrifugation (4400 x *g* for 10 min at 4°C), washed in cold DEPC-treated 0.9% (w/v) NaCl, then frozen under liquid nitrogen and stored at –80°C. Total RNA was extracted in RNase-free plasticware using Trizol Reagent (Invitrogen, USA) based upon the manufacturers instructions. The frozen mycelial pellet was ground to a fine powder under liquid nitrogen and approximately 100 mg was added to 1 mL Trizol. The large amount of cell debris was removed by an additional centrifugation at 12,000 x *g* for 10 min at 4°C, prior to the chloroform extraction step. Eppendorf 1.5 mL Phase Lock Gel tubes (light or heavy) were used for organic extraction steps. The RNA pellet was dissolved in DEPC-treated H_2O with a 10 min incubation at 50°C to assist in the complete dissolution of the pellet. Total RNA quality was judged by the appearance of a 2 μL sample separated on a 1% agarose gel stained with ethidium bromide, in conjunction with an $OD_{260nm/280nm}$ ratio of greater than 1.7 for RNA samples diluted 1 in 100 in TE buffer. RNA concentration was determined by the $OD_{260\ nm}$ reading.

2.10 RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) PCR

Total RNA was extracted from the mycelial mass of a 54 h old *T. reesei* culture grown in CLS medium as described in Section 2.9. Rapid amplification of 5' and 3' cDNA ends was carried out using the First Choice RLM-RACE (RNA ligase Mediated Rapid Amplification of cDNA Ends) Kit (Ambion, USA) according to the manufacturer's instructions with the following exceptions. For the reverse transcription step in the 5' RLM-RACE PCR, 16 pmol of the primer hexprobe.revpr 5'-CGTGCAAGCATACGCTAACG was used in place of the recommended 100 pmol random decamers to reduce the level of non-specific products that created smearing in subsequent PCR steps. The *hex1* cDNA generated was amplified in a single PCR step rather than nested PCR. The PCR reaction contained one unit of Qiagen Hotstart Taq Polymerase, 10 pmol hex5'RACE.inpr 5'-CGTCGCGAGTGATTAGCATC, 10 pmol 5' RACE outer primer (Ambion, USA), 1x Qiagen Hotstart Buffer, 20 μM dNTP's, 2 μl reverse transcription reaction and H_2O to 50 μl. The PCR conditions were 1x (95°C, 15 min), 35x (94°C, 30 s; 60°C, 30 s; 72°C, 1 min), 1x (72°C, 7 min).

The conditions for the 3' outer nested PCR were as recommended by the manufacturer and contained 10 μM of the gene specific outer primer hex3'RACE.outpr 5'-CTTCAAGCAGTACCGCGTCC, one unit of Qiagen Hotstart Taq Polymerase and 1x Qiagen Hotstart Buffer. The conditions were 1x (95°C, 15 min), 35x (94°C, 30 s; 60°C, 30 s; 72°C, 30 s), 1x (72°C, 5 min). The 3' inner PCR was as recommended with one unit of Qiagen Hotstart Taq Polymerase, 10 μM gene specific inner primer hex1fwd2.gwpr, and 1x Qiagen Hotstart Buffer. The PCR conditions were 1x (95°C, 15 min), 35x (94°C, 30 s; 60°C, 30 s; 72°C, 30 s), 1x (72°C, 5 min). PCR products were

cloned using the TA cloning kit (Invitrogen, USA), and recombinant plasmids were isolated and sequenced.

2.11 Nucleic acid hybridisation techniques

2.11.1 Generation of DNA probes

Digoxigenin (DIG) labelled DNA probes were used for northern blotting and Southern blotting detection procedures. The probes were generated by a standard PCR reaction using the PCR DIG labelling mix (Roche, Germany) as instructed by the manufacturer.

The *hex1* hybridisation probe was amplified from *T. reesei* chromosomal DNA by PCR using the primers hexprobe2.fwdpr and hexprobe2.revpr (Table 2-2) with a 60°C annealing temperature. The 616 bp probe was complementary to a portion of exon 3 of the *hex1* gene. The 1.6 kb actin gene (*act1*) probe was amplified from *T. reesei* genomic DNA using the primers actin.fwdpr and actin.revpr (Table 2-2) with a 56°C annealing temperature. The 720 bp *cbh1* probe was amplified from *T. reesei* genomic DNA with the primers TSTR.pr and cbhIfusion.seqpr with a 55°C annealing temperature. The 240 bp *dsRed* probe was amplified from the pTimer plasmid (BD Biosciences Clontech, USA) using the primers dsredprobe.fwdpr and dsredprobe.revpr (Table 2-2) with a 50°C annealing temperature step. Following PCR and separation by electrophoresis, the DNA probes were excised from the agarose gel (Section 2.4).

2.11.2 Preparation of genomic DNA for Southern blotting

Good quality genomic DNA (4 μg) was digested overnight at 37°C with relevant restriction enzymes in the appropriate buffers. The digested DNA was electrophoresed

on a 1% (w/v) agarose gel at 80V for approximately 90 min. A picture was taken under UV illumination and the genomic DNA was transferred onto a membrane as described below (Section 2.11.4).

2.11.3 Detection of RNA transcripts by northern blotting

2.11.3.1 Preparation of RNase–free solutions

All clean glassware was sterilised in an autoclave and later baked overnight at 180°C. The electrophoresis tank was washed thoroughly in Pyroneg and rinsed in RO-H_2O followed by two rinses in diethylpyrocarbonate (DEPC) treated H_2O and a rinse in 100% molecular biology grade ethanol to aid drying. Work benches and pipettes were washed with Pyroneg then wiped down with 70% ethanol. Gloves were changed frequently.

All water and most solutions used were treated with DEPC to inactivate any RNase activity. DEPC was added at 0.1% v/v concentration and solutions were incubated overnight at RT before autoclaving at 121°C for 30 min. Other solutions were either purchased RNase-free from the supplier or prepared from RNase-free chemicals with DEPC-treated solutions. For example, solutions containing Tris.

MOPS buffer was produced as a 10x stock solution using DEPC-treated H_2O and contained 200 mM MOPS, 50 mM sodium acetate and 10 mM EDTA with pH adjusted to 8.0 with NaOH. For some later experiments, 10x MOPS buffer was purchased from Eppendorf (Germany). RNA loading buffer contained 250 µL deionised formamide, 83 µL 37% formaldehyde, 50 µL 10x MOPS buffer and 10 µL 2.5% DEPC-treated bromophenol blue. For some later experiments RNA Gel loading buffer was purchased

from Eppendorf (Germany). Solutions used for blotting, hybridisation and detection are described in Section 2.1 and were prepared to be RNase-free.

2.11.3.2 Preparation of RNA for northern blotting

Total RNA (20 µg) was mixed with three volumes of RNA loading buffer and denatured at 65°C for 10 min. Samples were electrophoresed on a 1.2% agarose/formaldehyde gel as described in (Sambrook and Russell, 2001) at 70 V for 2 h. The gel was stained in 0.5 µg/mL acridine orange in DEPC-treated H_2O for 10 min, then destained for 15 min in DEPC-treated H_2O and a picture taken under UV illumination. Transfer of RNA onto the membrane and detection of gene-specific RNAs is described in Section 2.11.4. The hybridisation temperature for detection of *act1*, *cbh1* and *dsRed* specific RNAs was 56°C. The hybridisation temperature for *hex1* specific RNAs was 60°C. The intensity of expression was normalised against actin transcript levels.

2.11.4 Blotting and detection procedure

Transfer of nucleic acids from agarose gels to positively-charged nylon membrane (Roche, Germany) was carried out using the BioRad vacuum blotter (BioRad, USA) with 10x SSC buffer as recommended by the manufacturer. Nucleic acids were fixed onto the membrane by baking at 120°C for 20 min. Hybridisation and detection of specific genes or mRNAs were carried out using the DIG Luminescent Detection Kit for Nucleic Acids (Roche, Germany) based on the manufacturer's instructions. Membranes were incubated in 20 mL reconstituted DIG-Easy Hyb granules (Roche, Germany) in a Hybaid oven at the relevant hybridisation temperature for 1 h (see above). The DIG-labelled DNA probe was denatured at 100°C for 10 min then added to

fresh Hybridisation buffer (20 ng/mL), prewarmed to the hybridisation temperature. The probe was allowed to hybridise to the blot overnight. Following hybridisation, the blot was washed twice for 5 min with 2x SSC, 0.1% SDS at RT and twice for 15 min with 0.1x SSC, 0.1% SDS at RT before being rinsed in washing buffer. The membrane was blocked in blocking solution for 1 h at RT then incubated for 30 min in Anti-Digoxigenin-AP antibody (Roche, Germany) diluted 1 in 10 000 in fresh blocking solution. Excess antibody was removed by two 15 min washes at RT in washing buffer before a 1-2 min incubation in Alkaline Phosphatase buffer. The membrane was inserted in a hybridisation bag with approximately 500 µL ready-to-use CDP-star (Roche, Germany). The substrate was spread evenly across the membrane and the excess squeezed from the bag. The membrane was exposed to CL-Xposure film (Pierce, USA) and developed using Kodak GBX developing solutions.

2.12 Extraction of fungal proteins

2.12.1 Extraction of soluble cellular proteins

The extraction of intracellular soluble proteins was based on the protocol by Kurzatkowski *et al.* (1996). Mycelia from liquid cultures were collected by centrifugation (4400 x *g* for 10 min at 4°C), washed in cold DEPC-treated 0.9% (w/v) NaCl, then frozen under liquid nitrogen and stored at –80°C. The frozen mycelial pellet was ground to a fine powder under liquid nitrogen and approximately 0.5 g was added to 5 mL of chilled citrate buffer (Table 2-1) containing 1 mM PMSF and 0.1% protease inhibitor cocktail (Roche, Germany). The sample was mixed by vortexing and incubated at 4°C for approximately 1 h. Following the incubation, the soluble fraction

was separated from the insoluble material by centrifugation at 12 000 x *g* for 20 min at 4°C. The soluble fraction was stored at 4°C and used within two days.

2.12.2 Extraction of insoluble cellular proteins

For the extraction of total proteins from the cell, the procedure developed by Grinyer *et al*. (2004) was used. The insoluble cell pellet generated in Section 2.12.2 was washed with citrate buffer then resuspended in 5 mL of extraction buffer containing 7 M urea, 2 mM thiourea, 1% C7Bz0, 80 mM citric acid, 5 mM tributyl-phosphine, 1 mM PMSF and 0.1% (v/v) protease inhibitor cocktail. The suspension was vortexed and incubated at 80°C for 5 min. The cellular pellet was separated from the soluble fraction by centrifugation at 12 000 x *g* for 20 min at RT. The supernatant containing the "insoluble" fraction of cellular proteins was stored at RT and used within two days.

2.13 Detection of DsRed protein by Western blotting

Liquid cultures were centrifuged at 4400 x *g* for 10 min at 4°C and the supernatants were collected and stored at 4°C. Supernatant proteins were separated on denaturing 4-20% gradient gels (Life Technologies, USA) run at 150 V for 1.5 h in 1x glycine buffer (25 mM Tris base, 250 mM glycine, 0.1% SDS). The gel was soaked in transfer buffer (48 mM Tris base, 39 mM glycine, 0.037% SDS, 20% methanol) for 10 min then transferred onto nitrocellulose membrane using the BioRad transfer system for 2 h at 300 mA. Following transfer, the membrane was rinsed in 5% skim milk buffer in 1x TBS, 0.05% Tween 20 then blocked in this buffer for 30 min. The membrane was probed with DsRed monoclonal antibody (BD Biosciences Clontech, USA) diluted 1 in 200 in skim milk buffer overnight. The membrane was washed three times for 5 min in 1x TBS and probed with rabbit anti-mouse IgG (Calbiochem, USA) diluted 1 in 800 in

milk buffer for 1 h. Excess antibody was removed by three 5 min washes in 1x TBS before the membrane was covered with ready made Alkaline Phosphatase substrate (Promega, USA). The membrane was incubated in the dark until no further bands appeared and was rinsed in 1x TBS to stop the reaction.

2.14 N-terminal sequencing of *Trichoderma reesei* HEX1 proteins separated by two dimensional gel electrophoresis

Cell envelope proteins of *T. reesei* grown in CLS medium were previously separated by 2D gel electrophoresis (Lim *et al*., 2001). Two spots identified as HEX1 were cut from the gel, eluted in gel elution buffer at 37°C overnight, transferred to PVDF and washed using a ProSorb cartridge. The sample was then subjected to 10 cycles of N-terminal sequencing by automated Edman degradation using an Applied Biosystems 494 Precise Protein Sequencing System at the Australian Protein Analysis Facility (APAF, Macquarie University, Australia).

2.15 Fluorescence measurements of supernatants

Culture supernatants (200 µL per well) were loaded into a 96 well clear plate (Greiner, USA) and fluorescence was recorded on the BMG FLUOstar Galaxy plate reader (Australia). Green fluorescence was detected using the 480 nm excitation filter and 492 nm emission filter. Red fluorescence was detected using the 540 nm excitation filter and 584 nm emission filter. Fluorescence readings were carried out in duplicate and the supernatant from the culture of a nontransformant was used to determine the level of background fluorescence in the samples. This background fluorescence changed over a

period of days due to the changing composition of the culture medium. Therefore, the intensity of fluorescence is provided as the fluorescence ratio (ie. mean fluorescence of the transformant/mean fluorescence of the nontransformant).

2.15.1 Quantitation of DsRed fluorescence

2.15.1.1 Expression of DsRed in *E. coli*

DsRed derived from crude cellular extracts of *E. coli* expressing DsRed, was used as a positive control in Western blots and for estimating the concentration of DsRed secreted into the supernatant by *T. reesei* transformants. *E. coli* transformed with pUC18-*dsRed* (Section 2.6.1) was grown in 5 mL of LB until an OD_{600} of 0.5-1.0 was reached. The expression of DsRed was induced by the addition of 0.4 mM IPTG. The culture was incubated at 37°C for a further 2 h then incubated at RT for 16 h in the dark. The 2 h induction period produced a sufficient quantity of DsRed protein, however the fluorophore conformation was immature and required an extended incubation to assume the fluorescent conformation (Jakobs *et al*., 2000). The *E. coli* cells were collected by centrifugation and resuspended in PBS before lysis by sonication. Cellular debris was removed by centrifugation and the supernatant containing the DsRed was stored at –20°C.

2.15.1.2 Determination of DsRed concentration in *E. coli* cell extract

To deduce the concentration of DsRed within the crude *E. coli* cellular extract, a 2 μL, 5 μL and 10 μL sample of the extract was separated on a denaturing 4-20% precast SDS-PAGE gradient gel (Section 2.13) alongside a series of BSA protein concentration standards (0.25 mg/mL to 2 mg/mL). An image of the coomassie stained gel was

acquired using the ChemiImager System 5500 (Alpha Innotech, USA). The concentration of the DsRed protein band was calculated by comparison with the BSA standards by densitometry analysis using the ChemiImager AlphaEase FC software.

2.15.1.3 DsRed protein concentration and the fluorescence ratio

The relationship between the fluorescence ratio and DsRed concentration in supernatants of cultures grown in cellulase medium was estimated using the crude *E. coli* cell extract (Section 2.15.1.1). Several dilutions of the *E. coli* extract into the soluble fraction of cellulase medium were carried out. *E. coli* extract from nontransformed *E. coli* was also serially diluted as a control for background fluorescence. The fluorescence intensity of the dilution series of *E. coli* extract in culture medium was recorded as above (Section 2.14) and the fluorescence ratio calculated. The fluorescence ratio was plotted against DsRed concentration.

2.16 Confocal fluorescence microscopy

An Olympus IX70 confocal fluorescence microscope fitted with an Olympus LG-PS2 lamp, U-RFL-T burner and Argon and Helium/Neon lasers were used. For capturing green fluorescence, the Argon laser intensity was set at 21 V and barrier filter 510F was used. For capturing red fluorescence, the Helium/Neon laser intensity was set at 66 V and barrier filter BA585 was used. Images were captured using Fluoview software (version 4.3) with a Kalmann of 3 and step size of 0.2 μm for z-stack images. Z-stack images were compiled into an extended view image or annotated into a three-dimensional movie using Fluoview software. Images and movies were also manipulated using Image J software available from http://rsb.info.nih.gov/ij/ . Figures were edited

and assembled using Adobe Photoshop Elements (version 2.0) and Adobe Illustrator (version 10).

2.16.1 Sample preparation for microscopy

Fungal mycelia grown in liquid cultures were mounted on slides within double sided adhesive Gene Frame wells (ABgene, UK) whereby approximately 3 μL of the fungal culture was mixed with 20 μL of 30% glycerol in PBS and covered with a coverslip before observation.

To view Woronin bodies and hyphal morphology, the mycelia was fixed prior to observation. Approximately 100 μL of culture containing mycelia was collected by centifugation and the supernatant was replaced with 200 μL 4% paraformaldehyde in PBS. Mycelia were incubated at RT for at least 30 min then washed three times in PBS and mounted as described above.

Cultures were grown directly on a thin layer of growth medium on the microscope slide for the observation of mycelia grown on solid medium. The medium was made up to 2.5% agar concentration to encourage hyphal growth on the surface and not through the medium. The slides were prepared as follows. Three microscope slides were placed side-by-side. The outer slides were raised slightly by resting on 2-3 layers of masking tape on the bench. A drop of molten medium was placed on the middle slide and a fourth slide was immediately placed across the top such that it rested on the two outer slides. Once the medium had set, the top slide was removed. The medium was inoculated with conidia and incubated in a humid chamber until the desired level of growth was reached.

Chapter 3 Isolation of the *hex1* gene from *Trichoderma reesei* and *Ophiostoma floccosum*

3.1 Introduction

3.1.1 The *hex1* gene from filamentous fungi

Whilst there is growing research devoted to the physiology and functioning of the Woronin body of filamentous fungi, there are only some reports on the expression and structure of the *hex1* gene which codes for the major component of this unique fungal organelle. The *hex1* gene sequence from *N. crassa* was identified from the *N. crassa* genome database after the HEX1 protein was isolated (Jedd and Chua, 2000; Tenney *et al.*, 2000). Prior to this, no other *hex1* gene sequence had been identified. Comparison of genome databases for other filamentous fungi such as *A. nidulans*, *M. grisea* and *Botrytis cinerea* with the *N. crassa hex1* sequence resulted in the identification of similar genes (Tenney *et al.*, 2000), which all encode for a highly conserved HEX1 protein.

Only a few further investigations have followed since the confirmation of the role of the Woronin body as a septal plug and the identification of its major constituent, the HEX1 protein (Jedd and Chua, 2000; Tenney *et al.*, 2000). The unique and highly stable nature of the Woronin body core (Jedd and Chua, 2000) prompted the resolution of the crystal structure for HEX1 and the Woronin body core (Yuan *et al.*, 2003) (refer to Section 1.7.3). In addition, the role of the HEX1 protein as part of a survival mechanism for the fungal mycelium motivated a study on HEX1 expression during phytopathogenic attack by the fungus *Magnaporthe grisea* on rice crops. The crucial role of HEX1 in

maintaining mycelium integrity following hyphal damage suggested that HEX1 may be a useful target for a biocontrol mechanism. However, it was found that during mycopathogenic attack by *M. grisea,* HEX1 was expressed in germinating conidia and in vegetative cells but only very weakly in the appressoria, which are responsible for the invasive growth (Asiegbu *et al*., 2004). This recent report was the first to address the expression and regulation of the *hex1* gene.

The interest in *T. reesei* HEX1 described in the present work has stemmed from two angles: first, to investigate the *hex1* promoter, which drives the expression of this highly abundant protein, and to assess its usefulness for expression of selected proteins. Secondly, to explore further the role of HEX1 in maintaining mycelial integrity and to investigate how this mechanism may be manipulated to enhance the efficiency of *T. reesei* as a versatile protein expression system. Prior to this work, the *hex1* gene had not been isolated from any economically-relevant filamentous fungus. This chapter describes the isolation and characterisation of the *hex1* genes from *T. reesei* and the fungus *O. floccosum,* which is currently being developed as an alternative expression system in our laboratory.

3.1.2 The filamentous fungus *Ophiostoma floccosum*

Ophiostoma floccosum is a filamentous fungus that is one of the major causes of sapstain in softwood logs. Sapstain occurs once the fungal hyphae that have penetrated deep into the parenchyma and resin canals of wood, produce melanin. The resulting dark stained timber leads to significant economical losses for the forestry industry. The logs are infected with *Ophiostoma* on harvesting and the colonies penetrate deeper into the timber on transportation under humid, warm conditions such as those experienced

within a ship's hull. Albino strains of *Ophiostoma* have been selected and are currently used as a form of biocontrol for the dark staining fungi by inoculating wood chips with albino strains on harvesting and processing (Farrell *et al*., 1993). The albino strains then become the dominant colonisers of the wood and prevent sapstains from occurring (Held *et al*., 2003). An additional feature of inoculating chips with albino strains is that the colonisation of the fungus also reduced the pitch and resin content of the wood, thereby reducing the quantity of bleach chemicals required to make whiter paper products (Farrell *et al*., 1993).

The growth characteristics of *O. floccosum* potentially make this fungus useful for a variety of biotechnological applications. For example, the ability of hyphae to penetrate deep into the wood may serve as an ideal applicator of enzymes useful in the processing of wood chips for paper making. Alternatively, *O. floccosum* can also be grown in liquid culture for enzyme production.

3.2 Results and discussion

3.2.1 The *hex1* gene sequence of *Trichoderma reesei* and *Ophiostoma floccosum*

A fragment of approximately 600 bp in length was amplified from the genomic DNA of *T. reesei* and a fragment of 747 bp in length was amplified from the genomic DNA of *O. floccosum,* using degenerate primers designed against fungal genes similar to the *N. crassa hex1* sequence. Genomic DNA was digested with restriction enzymes and ligated to compatible linkers as described in Section 2.5. This linker library was used for chromosome walking PCR and the complete *hex1* encoding sequence and flanking

Figure 3-1. Sequence of the *hex1* gene of *T. reesei*. The putative regulatory motifs are highlighted in yellow for CAAT boxes, blue for CDRE, purple for AGAA boxes, green for CCAAT boxes and grey for CAC rich regions. A putative GATA box is boxed. The transcription start sites are marked with solid triangles and the polyadenylation sites are marked with open triangles. The sequence in lower case signifies intron sequence. Solid underline marks two regions of the introns which are highly complementary. Bold print marks conserved 5' intron splice sites and dashed underline marks a putative Kozac sequence.

regions were amplified, then sequenced (Figure 3-1). The gene sequences were subjected to analysis using the GenScan programme (Burge *et al*., 1997) on BioManager (www.angis.org.au) to predict the positions of exons. In *T. reesei*, the 783 bp ORF contained a single 109 bp intron and the *hex1* gene of *O. floccosum* contained a single 323 bp intron in a similar position following the highly conserved sequence of exon 2. The sequences of the genes isolated in this work are available from Genbank under accession numbers **AY288289** for *T. reesei* and **AY876925** for *O. floccosum*.

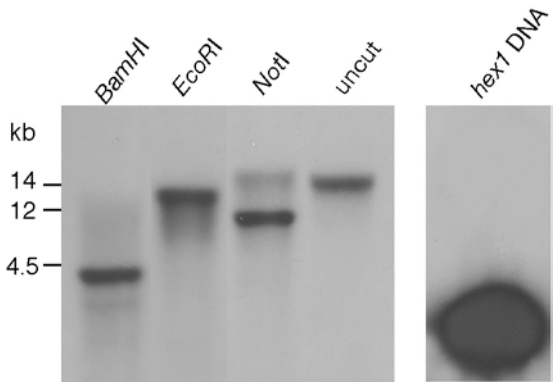

Figure 3-2. Southern blot of *T. reesei* strain VTT-D-79125 probed for *hex1*.

A Southern blot was performed in order to determine the number of copies of the *hex1* gene within the genome of *T. reesei*. Genomic DNA was digested with the restriction endonucleases *BamH*I, *EcoR*I and *Not*I, which do not have cleavage sites within the *hex1* gene probe sequence. A single band from all the digests strongly suggested that *hex1* is a single copy gene within the *T. reesei* genome (Figure 3-2).

3.2.2 Alternative forms of *hex1* transcript in *Trichoderma reesei*

Preliminary analysis of the expression of *hex1* in *T. reesei* by northern blotting indicated that greater than one transcript size existed for the gene (discussed later). To investigate

this further, RLM-RACE PCR was used to identify the transcription start sites and termination signals of the *hex1* transcripts from *T. reesei*. In this procedure only full length transcripts with capped 5' ends are successfully tagged with a linker sequence for amplification by RT-PCR (Section 2.10). PCR amplification of the 5' terminus of *hex1* cDNAs from a *hex1* cDNA library produced three bands of approximate sizes of 400 bp, 550 bp and 1000 bp (Figure 3-3). Each strong band had a mirror band which ran slightly slower than the main band. The mirror bands were excised from the gel separately and sequenced and were found to be identical to the main band. This indicates that secondary structure of the DNA fragment was responsible for this double banding effect on agarose gels.

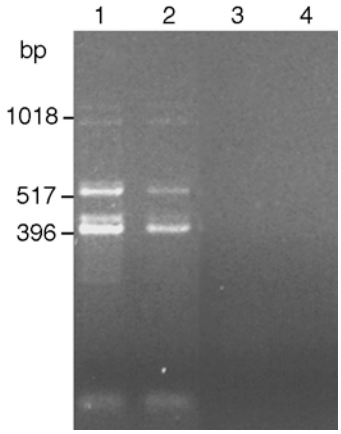

Figure 3-3. The 5' RLM-RACE PCR products amplified from full length *hex1* cDNAs. 1. VTT-D-79125 cDNA linker library. 2. Transformant h24 (see Chapter 4) cDNA linker library. 3. No linker VTT-D-79125 control. 4. No cDNA control. Refer to Chapter 2 for details in methodology.

The three RLM-RACE PCR products amplified from the 5' ends of *hex1* transcripts confirmed that alternative transcripts are transcribed from the single copy *hex1* gene. Two general transcription start point (*tsp*) regions were identified and alternative splicing of the transcripts resulted in three different isoforms of *hex1* mRNA (Figure 3-4). The *tsp* of several clones was in the region between – 327 to –275 (from A of the

start codon as position +1). From these transcription start sites the transcript undergoes alternative splicing at the 3' end of the intron within the coding sequence. Cloned transcript n198 (*tsp* –288) contained a 109 bp intron and the coding sequence for a histidine-rich variable region of HEX1 (discussed further in Section 3.2.3). The intron in cloned transcript type n197 (*tsp* –285) has a shift in the 3' splice site resulting in an extended 243 bp intron that encompasses the histidine-rich region producing a coding region that excludes most of the variable region of *hex1*. The third type of transcript, represented by clone n205, is different in that the *tsp* is at –1337 and it has a 506 bp intron within the 5' untranslated region (UTR). In *N. crassa,* the 5' ends of *hex1* cDNAs were concentrated at 360-420 bp upstream of the start codon with a single intron within the coding region which demonstrated no alternative splicing (Tenney *et al*., 2000). The *N. crassa hex1* cDNA is most similar to *T. reesei hex1* transcript type n197.

The yield of PCR product decreased with increasing size of the product such that the 1000 bp product was just visible from a standard 50 μL reaction. It was not possible to amplify any products greater than the 1000 bp fragment but this does not eliminate the possibility of larger cDNAs transcribed from further upstream. Amplification of the 3' end of *hex1* transcripts using a poly T primer produced only a single PCR product which was subsequently cloned. Two recombinants were sequenced and showed that the polyadenylation sites are 243 bp and 245 bp downstream from the stop codon (Figure 3-1).

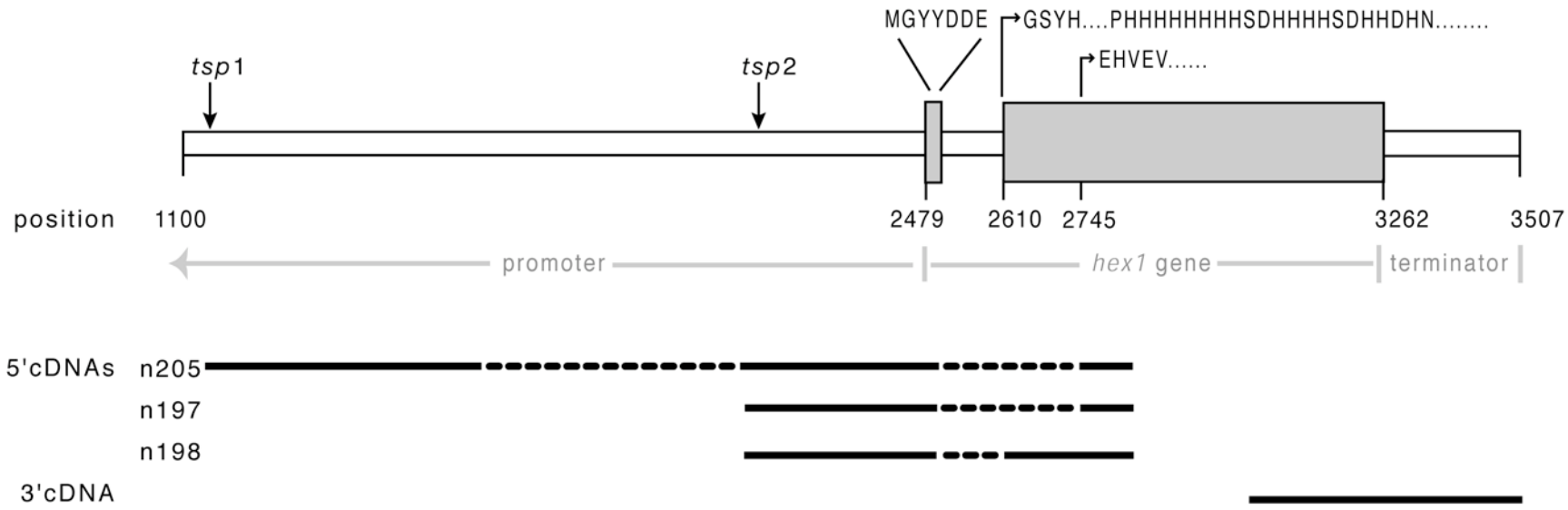

Figure 3-4. Schematic representation of the 5' and 3' ends of alternative *hex1* cDNAs. The shaded boxes represent coding sequence of the *hex1* gene and the relative positions coincide with the sequence in Figure 3-1. The residues MGYYDDE are encoded by exon 2 of the *hex1* gene and the protein encoded by exon 3 either commences with residues GSYH…, and includes a histidine-rich stretch, or commences with residues EHVEV… . The solid line represents cDNA sequence and dashed lines show the intronic regions. Intron 1 lies within the 5' UTR and intron 2 is of variable length and lies within the coding region.

3.2.3 Alternative forms of HEX1 protein

The three transcript types from *T. reesei* theoretically code for two forms of HEX1 protein, a 25 kDa form or a 19 kDa form. All of the HEX1 proteins identified from filamentous fungi have a highly homologous first coding exon followed by a variable region that leads into the conserved remainder of the protein (Tenney *et al.*, 2000) (Figure 3-5). This pattern also exists for the HEX1 protein of *O. floccosum* and the two forms of HEX1 from *T. reesei* with the length of the variable region depending on the alternative splicing of the internal intron. In *T. reesei* the 25 kDa form of HEX1 is encoded by transcript type n198 and contains a long histidine-rich stretch. The 19 kDa form of HEX1 could be translated from transcript type n197 or n205.

The *A. nidulans* HEX1 protein also has an extended variable region that is histidine-rich. In *M. grisea*, two cDNAs for *hex1* have been documented (Tenney *et al.*, 2000). These cDNAs encoded two forms of HEX1 protein, with and without a histidine-rich variable region (Tenney *et al.*, 2000), similar to that described here for *T. reesei*. The *M. grisea hex1* cDNA that encodes HEX1 without the histidine-rich stretch (Genbank **AF170544**), also has a 111 bp intron spliced from the 5' UTR of the transcript, and is similar to *T. reesei hex1* transcript type n205. The conservation of this pattern of expression and splicing suggests it has functional importance. However, it is currently unknown why these alternative forms occur.

To identify if the two potential forms of HEX1 exist as proteins, we returned to previous work in which *T. reesei* cell membrane-associated proteins were separated on two dimensional gels (Lim *et al.*, 2001). Two spots, d5 and d6, which had previously been identified as HEX1 were excised from one of the replicate two dimensional gels

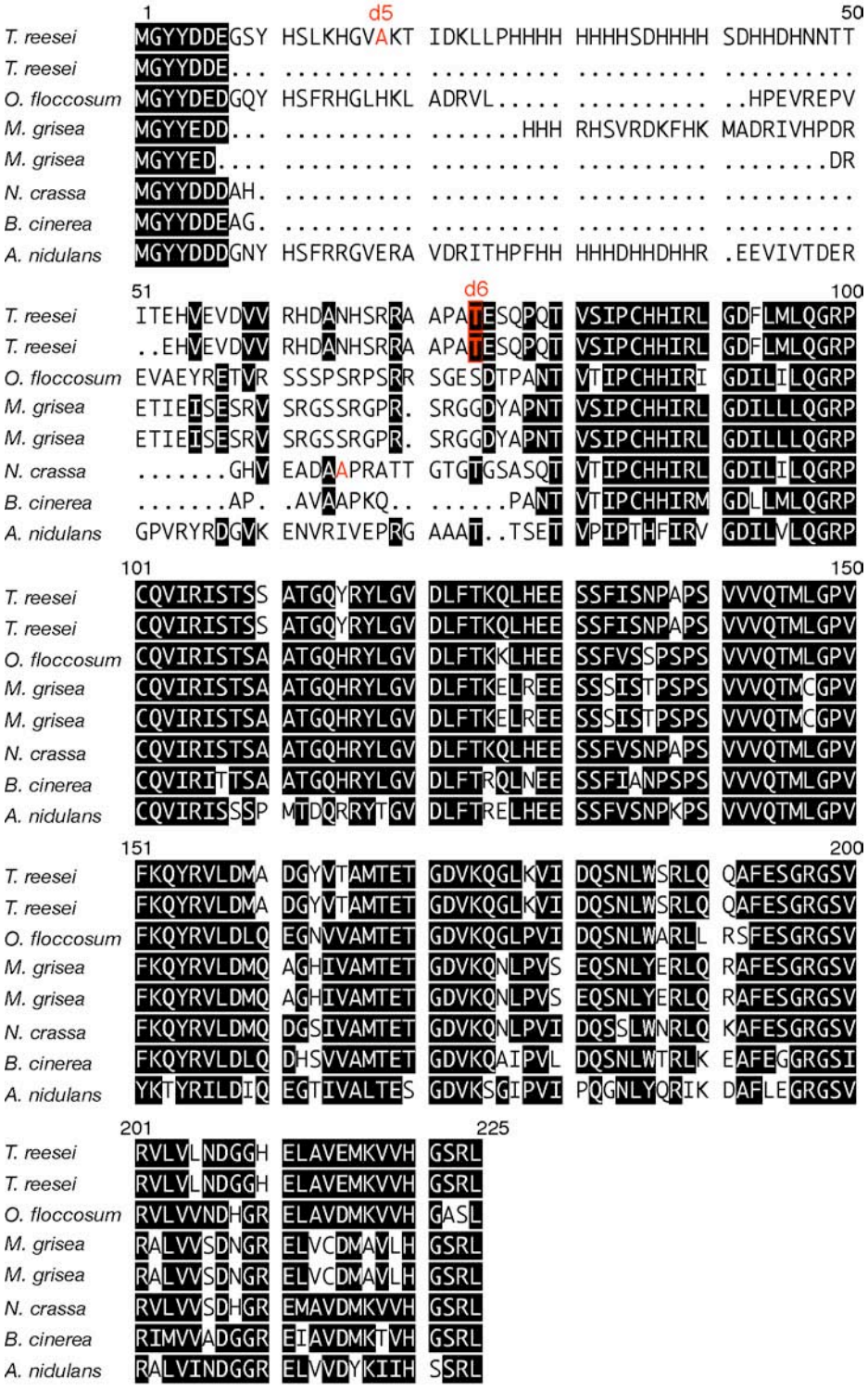

Figure 3-5. Alignment of HEX1 proteins from filamentous fungi. Residues in reverse font represent conserved residues. Red font marks the N-terminus of the HEX1 protein that has undergone post-translational cleavage. Positions marked d5 and d6 show the N-terminus of *T. reesei* HEX1 proteins excised from a two dimensional gel conducted in a previous study (see text).

and subjected to N-terminal sequencing. Each spot represented an alternative form of HEX1 with differently cleaved N-termini (Figure 3-5). Comparison with the *hex1* cDNAs suggests that the two forms of HEX1 protein may be translated from the alternative transcripts. HEX1 from spot d5 must have been derived from transcript type n198, but HEX1 of spot d6 may have been derived from either transcript type n197, n198 or n205. The protein from spot d5 indicated cleavage of approximately 18 residues from the histidine rich form of HEX1 and spot d6 has had either 28 residues cleaved from the smaller form or alternatively, 73 residues from the histidine-rich form. The size of *N. crassa* HEX1 is 19 kDa and it undergoes cleavage of 16 residues from the N-terminus yielding a 17 kDa form of HEX1 with an unknown function (Tenney *et al.*, 2000). There is no recognised conserved cleavage sequence at any of these sites. These results confirm that the mRNA corresponding to the histidine-rich form of HEX1 is translated and that the HEX1 protein exists in at least two forms due to post-translational cleavage of the N-terminus.

The histidine-rich region of the *T. reesei* HEX1 protein is unusual and is unlikely to be folded within the protein due to the highly charged composition. A BLAST search (www.ncbi.nlm.nih.gov/BLAST/) of the protein database with the HEX1 histidine stretch returned several other proteins which contain similar histidine-rich motifs. Generally, the proteins revealed by the search bound metal ions. Examples include the metal-binding polypeptide of *Helicobacter pylori* (NP208218), metal transporter of *Pseudomonas aeruginosa* (NP249988), magnesium ion dependent protein kinase of human and mouse (Q9UBE8, NP032728), and several transcription factors including *Arabidopsis* zinc finger transcription activator (NP200497). A histidine-rich region

(seven repeats of His-X) is also present in the *T. reesei* transcription activator ACEII adjacent to the zinc binuclear cluster DNA-binding domain (Aro *et al.*, 2001).

3.2.4 Further *in silico* analysis of *Trichoderma reesei* HEX1 protein

The HEX1 protein was subjected to various computational analyses in order to gain some insight into the regulation of *hex1* expression and to further understand the features of HEX1 which are relevant to its physiological functioning in the fungal mycelium.

The leading six amino acids of the N-terminus (MGYYD/ED) are conserved in all known HEX1 sequences, which suggests a functional role. However, this peptide is represented as a free end in the molecular structure of HEX1 solved for *N. crassa* (Yuan *et al.*, 2003) with no involvement in secondary structure. A BlastX search for short, nearly exact matches to the N-terminal sequence highlighted several other proteins which contained a very similar motif. These included the *Mycoplasma mycoides* prolipoprotein (NP 975404), *Thermoplasma volcanium* sugar kinase (NP 110646), *Arabidopsis thalina* jacalin lectin family protein and the receptor lectin kinase 3 (AAD00733). The two tyrosine residues of this conserved region are potential phosphorylation sites. There is a conserved serine/threonine rich region of five residues from position 107, which corresponds to an exposed loop on the solved *N. crassa* HEX1 structure (Yuan *et al.*, 2003). These residues are also potential phosphorylation sites.

T. reesei HEX1 has a conserved peroxisome targeting signal (-SRL) on the C-terminus that has been previously characterised in the *N. crassa* HEX1 protein (Jedd and Chua, 2000). However, the residues –ASL on the C-terminus of the *O. floccosum* HEX1 do

not conform to the known peroxisome targeting tripeptides S/A/C-K/R/H-L characterised in mammalian cells (Gould *et al*., 1989). This difference may suggest that -ASL is a novel peroxisome-targeting signal in fungi. There is conservation of the residues within HEX1 of *T. reesei* and *O. floccosum* that correspond to residues involved in the Group I, II and III interactions described for HEX1 polymerisation (Yuan *et al*., 2004) (refer to Section 1.7.3). The highly conserved features of HEX1 throughout the filamentous fungi implies that the reported mechanism of Woronin body formation in *N. crassa* (Yuan *et al*., 2003) is likely to be applicable across species.

The HEX1 protein sequences were analysed using the calmodulin target database at http://calcium.uhres.utoronto.ca/ctdb/flash.htm and the region spanning residues 188 to 203 were nominated as putative calmodulin binding sites based on the hydropathy, residue weight and charge and the propensity to form an alpha helix. This region coincides with alpha helices H2 and H3 of the HEX1 structure solved for *N. crassa* (Yuan *et al*., 2003). HEX1 of *Aspergillus oryzae* also possesses a putative calmodulin binding domain at the C-terminus which has been shown to bind calmodulin *in vitro* (Juvvadi *et al*., 2004). Calmodulin is a cytoplasmic calcium ion receptor present in all eukaryotic cells. On binding calcium, calmodulin (CaM) is responsible for triggering numerous biochemical cascades in cell signalling, circadian rhythms, cell polarisation and regulation of ion gated membrane channels (Yang *et al*., 2003; Kortvely *et al*., 2004). The presence of a CaM binding site on the HEX1 protein may suggest a role of CaM in the formation of Woronin bodies in fungi. The Ca^{++}/CaM complex typically binds enzymes, predominately multifunctional Ca^{++}/CaM-dependent serine/threonine kinases and the phosphatase, calcineurin (Joseph *et al*. 2002; The CaBP Data Library at http://structbio.vanderbilt.edu/cabp_database/general/targets/calmod_targ.html).

However, HEX1 has not been shown to demonstrate any enzymatic activity, so the purpose of calmodulin binding remains unknown.

Analysis of the HEX1 proteins has led to the identification of several features, including putative phosphorylation sites and calmodulin binding sites, which may be important in the cellular role of HEX1 but were not explored any further for the purposes of this thesis.

3.2.5 Conserved putative regulatory motifs within the *hex1* promoter region

The discovery of alternative *tsp*s and an intron within the 5' UTR of the *T. reesei hex1* promoter suggested that the mechanism of regulation may be complicated. The promoter regions of the *O. floccosum* and *T. reesei hex1* genes were analysed for putative conserved regulatory motifs in order to gain insight into the regulation of the expression. The 5' flanking region of *hex1* from *T. reesei* and *O. floccosum* were compared to the 5' flanking regions of *hex1* from *N. crassa* and *M. grisea* for small conserved sequences using the DNA-pattern programme at the site http://embnet.cifn.unam.mx/rsa-tools. *T. reesei,* like *O. floccosum*, *N. crassa* and *M. grisea* have a conserved putative ribosome binding site (Kozac sequence) adjacent to the translation start codon (GCCAAGATGG). CAC repeats appear throughout the promoter region for *T. reesei* (Figure 3-1). These repeats are also present within the *M. grisea and N. crassa hex1* promoters but the positioning is not consistent and if they possess a functional role, it is yet to be defined. There is a putative consensus binding motif (GTGGCGCC) at position 149 in the *T. reesei* sequence for the calmodulin/calcineurin activated transcription factor Crz1p (Staphopoulos *et al.*, 1997;

Cyert, 2003). Analysis of the promoter failed to recognise any TATA-like box upstream of the start codons. In the *T. reesei* gene there are four putative CAAT boxes, two at positions 657 and 678, which are approximately 460 and 480 bp upstream of the *tsp* at position 1140, and two within the 5' UTR intron (Figure 3-1). Also within the *T. reesei hex1* 5' UTR intron are alternating purine- and pyrimidine-rich regions expanding 38 bp to 107 bp, which are frequently interrupted resulting in approximately 70-80% CT or GA. Pyrimidine-rich areas frequently occur upstream from the *tsp* in fungal promoters (Punt *et al.*, 1988), but these are usually uninterrupted and not adjacent to a purine rich region. The purine and pyrimidine rich regions may indicate strong secondary structure within the 5' UTR of *hex1* and may have implications for *cis* regulation of gene expression by positioning of the nucleosome (Brahmachari *et al.*, 1997; Raghavan *et al.*, 1997). There are also three putative AGAA boxes and six putative consensus sites for CreA binding which lie within the 5' UTR intron of *T. reesei*, yet upstream from the *tsp* for transcript types n197 and n198. These sites are involved with promoter regulation by undescribed *cis* acting motifs (Mach *et al.*, 2003) and carbon repression (Kulmburg *et al.*, 1993) respectively. Within the 5' UTR intron are also two putative CCAAT motifs, which have been shown to induce transcription and enhance promoter activity by remodelling of nucleosomes (Narendja *et al.*, 1999; Steidl *et al.*, 1999; Zeilinger *et al.*, 2003).

3.2.5.1 Analysis of the intron sequences

The presence of introns, which contain promoter regulatory motifs, within the 5' UTR of fungal genes is rare and only a few cases have been described so far. There is an intron within the 5' UTR of *A. nidulans* glyceraldehyde-3-phosphate dehydrogenase gene (Punt *et al.*, 1988), in the endoglucanase III gene of *T. reesei* (Saloheimo *et al.*,

1988) and in the 1,3-β-glucosidase gene from *T. atroviride* (Donzelli *et al.*, 2001). A 229 bp intron appears in the 5' UTR of the *T. reesei ace1* gene and contains a start codon and a potential open reading frame for 32 aa (Saloheimo *et al.*, 2000). The 188 bp intron within the 5' UTR of 1,3-β-glucosidase gene from *T. atroviride* is 78 bp upstream of the start codon and contains two putative TATA boxes as well as transcription factor binding sites such as STRE motifs, CREI binding elements, conidiation responsive elements, and four CAAT boxes (Donzelli *et al.*, 2001). The authors suggested that these elements are likely to play a role in the regulation of gene expression. A region within an intron of the 5' UTR of the *A. oryzae* thiamine synthesis gene (*thiA*), forms part of a ribo-switch-like domain, which controls mRNA splicing and gene expression through direct interaction of the motif with the metabolite thiamine (Kubodera *et al.*, 2003). Therefore, the introns of the 5' UTR of genes can be directly involved in gene regulation. The high concentration of putative fungal promoter regulatory motifs within the intron of the 5' UTR of the *T. reesei hex1* gene suggests that this region is important for the regulation of expression.

Analysis using BioManager (http://biomanager.angis.org.au) MEME programme of 48 introns from fungal genes showed conserved regions within the introns. The conserved splice site GTAHGYT marks the 5' boundary of both *T. reesei hex1* introns. The intron within the coding sequence also contains the highly conserved UACUAAC box, found in *Saccaromyces cerevisiae* (Parker *et al.*, 1987) and *T. atroviride* (Donzelli *et al.*, 2001) which is the recognition site for binding of snRPS that lead to Lariat formation and splicing of the intron. The intron within the coding sequence for *T. reesei hex1* contains the palindromic sequence, GTCTGGCCAGAC, and the intron of the *O. floccosum hex1* gene contains a stretch of 15 nucleotides from position 803 that are

complementary to nucleotides from position 837 which may be involved in hairpin formation.

3.2.6 Expression of *Trichoderma reesei hex1* during growth

HEX1 has been shown to be an abundant component within protein preparations extracted from the *T. reesei* cell envelope (Lim *et al.*, 2001). In order to explore the expression of *hex1* in relation to growth and culture medium, RNA was extracted from *T. reesei* cultures grown in CLS medium (Section 2.9) which supports the synthesis of cellulolytic enzymes, minimal glucose medium, which typically represses the synthesis of enzymes regulated by carbon catabolite repression, and minimal glycerol medium, which is considered a neutral carbon source in relation to cellulase expression (Margolles-Clark *et al.*, 1997). The northern blots were probed for *hex1,* the main cellulase gene *cbh1,* and actin.

Northern blots (Figure 3-6) show the presence of *hex1* transcripts on cellulase-inducing, cellulase-repressing (glucose) and "neutral" conditions, indicating that *hex1* expression is not directly regulated by the carbon source of growth medium. When the level of *hex1* transcripts is compared to the growth curve of the cultures, it is apparent that the comparatively extended period of high expression of *hex1* on glycerol medium occurs during an extended period of steady growth. Likewise, the highest expression in CLS medium at 24 h is also during a period of steady growth (if the graph is extrapolated to zero weight at inoculation) (Figure 3-6). In glucose medium, the peak level of transcript was at 54 h but the overall level of expression was low. Thus it appears that *hex1* expression is highest during the active growth phase of a culture, which implies that the *hex1* promoter is a growth-rate dependent promoter.

A peak in *hex1* transcript level, of comparable intensity to that seen in glycerol and CLS medium, was not present in the glucose-grown samples. This may be attributed to the very rapid growth of the mycelium between 24 h and 54 h, but not concurrent with the sampling time points. It may be speculated that a sample taken at approximately 37 h would demonstrate a peak in *hex1* expression comparable to CLS at 24 h and glycerol at 54 h. Alternatively, the relatively low level of *hex1* message in glucose medium may be indicative of some repression of the *hex1* promoter occurring which may be effective via the putative CREI binding sites within the intron of the 5' UTR (Section 3.2.5). However, conserved CREI binding sites are not necessarily indicative of a promoter that is subjected to glucose repression since not all the CREI sites upon the *cbh1* promoter bind CREI (Takashima *et al.*, 1996). Some evidence suggests that CREI may have an additional role to carbon repression of genes (Ilmén *et al.*, 1996b). In addition, the repression of *hex1* expression on glucose medium is not congruent to the level of HEX1 protein isolated from the cell envelope of *T. reesei* grown on glucose (Lim *et al.*, 2001).

The northern blots are not directly comparable to each other due to different exposure times required for each probe. At peak HEX1 expression in CLS medium, the level of *hex1* transcript was several thousand fold lower than the level of *cbh1* transcript on induction. However, at the early time point of 24 h and under cellulase inducing conditions, *hex1* is expressed earlier and more intensely than *cbh1*. In glycerol medium, when *cbh1* is not expressed, *hex1* is expressed strongly.

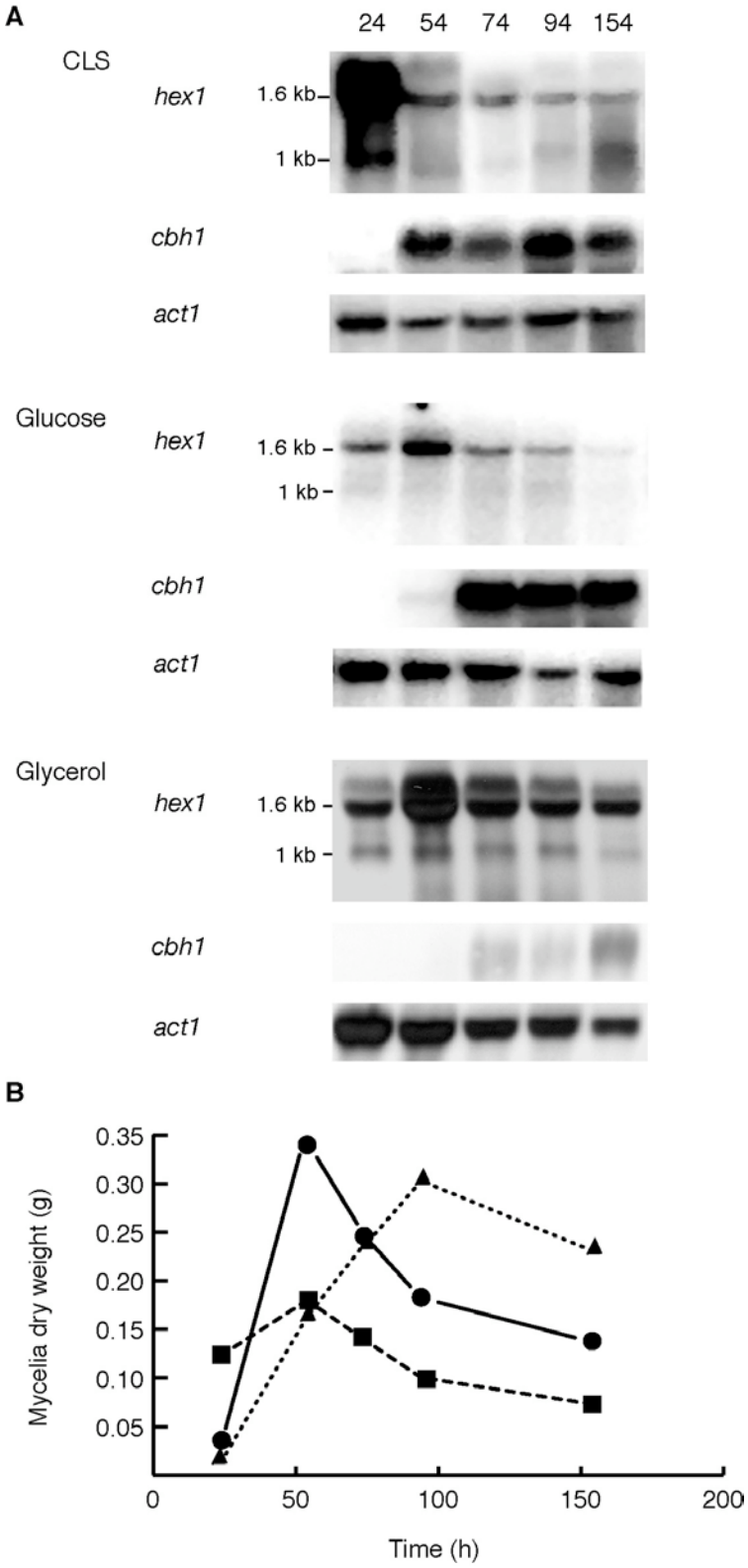

Figure 3-6. The growth rate and expression of *hex1* and *cbh1* over a 154 h culture period in various media. A. A blot of total RNA (20 µg) was hybridised with the gene specific probe indicated along the left, including actin (*act1*) as a loading control. The blots were exposed for different times in order to show a signal thus the intensities between different blots are not comparable to each other, except for those in glycerol medium. B. Mycelium dry weight over the culture period when grown in CLS medium (dashed line), glucose medium (solid line) or glycerol medium (dotted line).

The northern blots reveal that the dominant *hex1* transcript is approximately 1.6 kb in size, which corresponds to transcript type n205 with a predicted size of 1617 kb. This transcript does not code for the histidine-rich form of HEX1 protein. The other band, running at approximately 1 kb, was expressed less intensely but mirrored the expression pattern of the dominant 1.6 kb transcript in CLS, glycerol medium and glucose medium. This 1 kb transcript corresponds to transcript type n197 which also does not code for the histidine-rich form of HEX1. A 1.2 kb sized transcript corresponding to transcript type n198, which does encode the histidine-rich form of HEX1, was not detected on these northern blots. The constant difference between the level of the 1.6 kb and 1 kb transcripts indicates that there is differential regulation of each *tsp*.

A band, approximately 1.8 kb in size, was detected by the *hex1* probe on the northern blots. This band does not correspond to either of the transcript types identified by RLM-RACE PCR but the possibility of additional *tsp* further upstream from position 1141 can not be discounted.

The appearance of *cbh1* transcripts from 74 h in glucose medium was probably due to the depletion of the repressor (glucose) in the medium, as seen previously (Ilmén *et al*., 1997). The exhaustion of glucose was met with the commencement of die off within the culture as demonstrated by the declining mycelium dry weight. The cellular debris, including the polysaccharides of the cell wall, may act as inducers to cellulase expression (Kubicek *et al*., 1998). In glycerol medium, even though there was no active repression of *cbh1* expression, there was a minimal amount of free inducing compounds in the medium due to the prolonged active growth phase and reduced level of cell death.

3.2.7 Post-translational modifications of HEX1

Numerous spots of variable pI and mass were identified as HEX1 on 2D gels (Lim *et al*., 2001). We have shown that three alternative transcripts are transcribed from the *hex1* gene, which encode two forms of HEX1 protein. Each of these forms of HEX1 protein may then undergo posttranslational cleavage of the N-terminus to produce two isoforms of HEX1 protein of different size. However, this does not explain the numerous spots separated on the two dimensional gels.

3.2.7.1 Interpretation of *hex1* gene sequence data in relation to a previous study of HEX1

The multiple protein spots identified as HEX1 (Lim *et al*., 2001) may have been due in part, to proteolysis, even though a protease inhibitor cocktail was used during extraction of the cell membrane-associated proteins, and the denaturing agent thiourea, which can have an inhibitory effect on proteases (Castellanos-Serra *et al*., 2002), was present in the sample loading buffer. Many of the spots identified as HEX1 had a molecular weight larger than that predicted based on the HEX1 sequence. HEX1 polymerises into a crystalline lattice bound by hydrogen and salt bridge interactions (Yuan *et al*., 2003). The polymer of HEX1 may not have been completely denatured before separation on the two dimensional gels. This is possible because the HEX1 matrix of a Woronin body is a stable complex (Jedd and Chua, 2000). Previous investigators have found that the dense HEX1 matrix of *N. crassa* Woronin bodies was resistant to treatments including salt washes with 1 M NaCl, lysis buffer containing HEPES and EGTA, and boiling in SDS-PAGE gel loading buffer, such that the crystalline structure of the Woronin body core was still present under microscopic examination. HEX1 also failed to solubilise in 1% Triton X-100 or 1 M urea (Jedd and Chua, 2000). HEX1 dimers and tetramers also

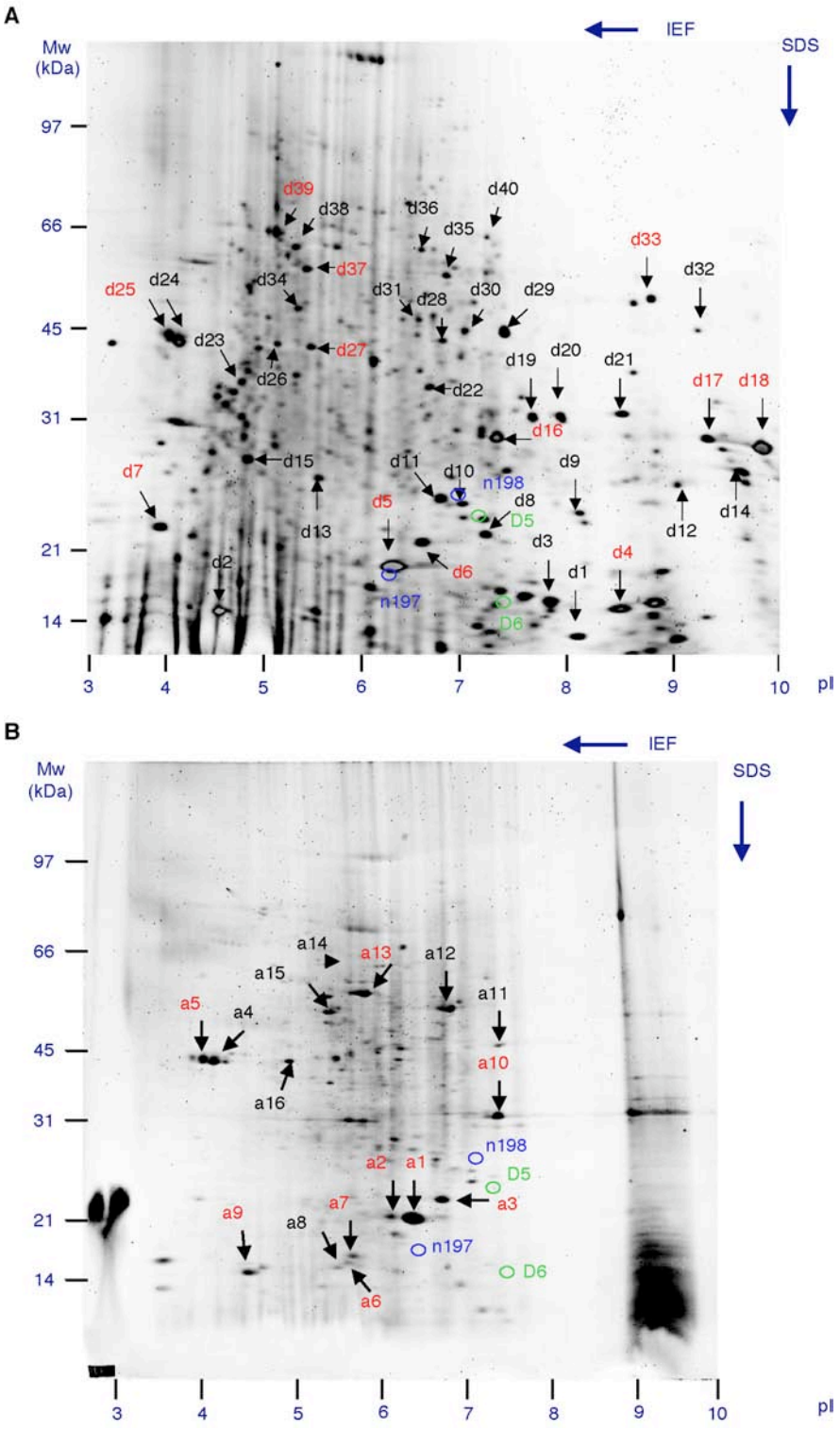

Figure 3-7.Two dimensional gels showing proteins separated from the cell membrane of *T. reesei* grown in CLS medium (A) or glucose medium (B). Spots identified as HEX1 are labelled in red and the hypothetical proteins from *hex1* transcripts and N-terminal sequencing are shown in blue and green respectively. Adapted from Lim *et al.* (2001).

remained intact after boiling in 2% SDS, 140 mM β-mercaptoethanol (Tenney *et al.*, 2000). The third possible factor which would contribute to the number of spots identified as HEX1, is post-translational modification which can cause varying pIs and mass differences. The stable HEX1 dimers and tetramers purified from *N. crassa* Woronin bodies were denatured to individual units after treatment with phosphatase, indicating that phosphorylation of HEX1 was important for the stable oligomer structure of the Woronin body core (Tenney *et al.*, 2000). Therefore, phoshorylation of the several threonine, serine and tyrosine residues on HEX1 (discussed in Section 3.2.4) could cause a shift in protein pI (Sickmann *et al.*, 2001). There are also two potential N-linked glycosylation sites at residues 47 and 65.

A range of post-translational modifications to the HEX1 protein, including those discussed above, may explain the difference between the actual positions of spots d5 and d6 on the 2D gels and the theoretical positions deduced from the predicted pI and molecular weight of the primary amino acid sequence (Figure 3-7A).

3.3 List of main findings

- *hex1* is a single copy gene in *T. reesei* from which three alternative RNAs are transcribed.

- The dominant mRNA type has a *tsp* –1337 bp from the start codon, contains a 506 bp intron within the 5' UTR and a single intron after a highly conserved exon 2.

- The three transcript types code for two forms of HEX1 protein which are different according to the length of a variable region which includes a histidine-rich stretch.

- The intron within the 5' UTR contains numerous putative promoter regulatory motifs indicating complex regulation.

- The *hex1* promoter is a growth rate dependent promoter which drives the strongest transcription during the active growth phase of a culture.

- The *hex1* promoter is not likely to be directly affected by the carbon source of the growth medium.

- *The hex1* promoter is more active than the *cbh1* promoter after 24 h in medium that induces *cbh1* expression.

- The *hex1* promoter is more active than the *cbh1* promoter in conditions that are noninducing or repressive of the *cbh1* gene.