Chapter 3

Aspergillus niger genome wide analysis reveals a large number of novel α-glucan acting enzymes with unexpected expression profiles

Xiao-Lian. Yuan*, Rachel M. van der Kaaij*, Cees A.M.J. van den Hondel, Peter J. Punt, Marc J.E.C van der Maarel, Lubbert Dijkstra, Arthur F.J. Ram

* These authors contributed equally to this study

Submitted
Abstract

The filamentous ascomycete *Aspergillus niger* is well known for its ability to produce a large variety of enzymes for the degradation of plant polysaccharide material. A major carbon and energy source for this soil fungus is starch, which can be degraded by the concerted action of α-amylase, glucoamylase and α-glucosidase enzymes, members of the glycoside hydrolase (GH) families 13, 15 and 31, respectively. In this study we have combined analysis of the genome sequence of *A. niger* CBS 513.88 with microarray experiments to identify novel enzymes from these families and to predict their physiological functions. We have identified 17 previously unknown family GH13, 15 and 31 enzymes in the *A. niger* genome, all of which have orthologues in other aspergilli. Only two of the newly identified enzymes, a putative α-glucosidase (AgdB) and an α-amylase (AmyC), were predicted to play a role in starch degradation. The expression of the majority of the genes identified was not induced by maltose as carbon source, and not dependent on the presence of AmyR, the transcriptional regulator for starch degrading enzymes. The possible physiological functions of the other predicted family GH13, GH15 and GH31 enzymes, including intracellular enzymes and cell wall associated proteins, in alternative α-glucan modifying processes are discussed.

Introduction

*Aspergillus niger* is a saprophytic fungus well known for its production and secretion of a variety of hydrolytic enzymes contributing to its ability to degrade plant polysaccharides such as cellulose, hemicellulose, pectin, starch and inulin (De Vries and Visser 2001; Tsukagoshi et al. 2001; Yuan et al. 2006). Starch is the most abundant storage carbohydrate in the plant kingdom and is present in tubers, seeds and roots of a variety of crop plants including cereals, potatoes and manioc (Peters 2006). Starch is composed of two different molecules: (i) amylose, an unbranched, single chain of α-(1,4)-linked glucose residues and (ii) amylpectin, consisting of a α-(1,4)-linked glucose chain with α-(1,6)-branches on every 12-25 glucose residues along the α-(1,4)-linked backbone (Robyt 1998). The degradation of starch is performed by a variety of enzymes, which are divided over three Glycoside Hydrolase (GH) families based on their sequence similarity (http://www.cazy.org) (Coutinho and Henrissat 1999). The first step in starch degradation is the endo-hydrolysis of the long polysaccharide chains into shorter maltooligosaccharides and α-limit dextrans by α-amylases (EC 3.2.1.1). α-Amylases belong to family GH13, a large family containing various hydrolysing and transglycosylating enzymes, mostly acting on α-(1,4) or α-(1,6)-glycosidic bonds. Members of family GH13 have a (β/α)8 barrel structure and can be recognized by four highly conserved amino acid regions containing three catalytic residues (MacGregor et al. 2001; Nakajima et al. 1986). After endo-hydrolysis, subsequent steps in starch degradation involve exo-acting enzymes releasing glucose. This reaction is performed by glucoamylase type enzymes of family GH15 (EC 3.2.1.3), a relatively confined family with regard to enzyme specificity, as all its studied members hydrolyse either α-(1,4)
or α-(1,6)-bonds to release β-glucose from the non-reducing end of maltooligosaccharides. Most GH15 enzymes described thus far possess a starch binding domain (SBD) (Sauer et al. 2000), a discrete C-terminal region of the protein that binds to starch and facilitates hydrolysis (Southall et al. 1999). Additionally, α-(1,4)-glucosidases of family GH31 may release α-glucose from the non-reducing end of starch (EC 3.2.1.20). This family also harbours other enzyme specificities such as α-xyllosidase activity.

Several A. niger enzymes involved in starch degradation, and their corresponding genes, have been characterized and isolated. A. niger glucoamylase GlaA (family GH15) is an important enzyme for the modification of starch in the food industry (Boel et al. 1984; van Dijck et al. 2003). Additionally, one GH31 α-glucosidase (AglA, renamed AgdA) (Nakamura et al. 1997) has been characterized previously, as well as three family GH13 α-amylases: acid amylase AamA, and the almost identical AmyA and AmyB (Boel et al. 1990; Korman et al. 1990). The transcriptional regulation of the genes encoding starch degrading enzymes has been studied in several aspergilli (Nakamura et al. 2006). In general, their expression is high on starch and induced by the presence of (iso)maltose (Tsukagoshi et al. 2001; Kato et al. 2002a). The presence of the inducer activates the Zn(II)2Cys6 transcription factor AmyR which binds to CGGN8(C/A)GG sequences in the promoter regions of AmyR target genes thereby activating their transcription (Petersen et al. 1999; Gomi et al. 2000; Tani et al. 2001; Ito et al. 2004).

Recent studies have indicated that some GH13 enzymes in fungi may be involved in the synthesis or modification of α-glucan in the fungal cell wall, rather than in starch degradation. The cell wall of aspergilli contains four major classes of polysaccharides: chitin, α-glucan, β-(1,3)-glucan and galactomannan (Fontaine et al. 2000; Beauvais and Latgé 2001). The α-glucan fraction identified in A. niger consists of two types of molecules: a linear polymer with alternating α-(1,3)/(1,4)-glycosidic bonds called nigeran (Barker and Carrington 1953) and pseudonigeran, a linear α-(1,3)-glucan molecule with some (3-10%) α-(1,4)-linkages (Johnston 1965; Horisberger et al. 1972). Synthesis of α-glucan is thought to be carried out by α-glucan synthase enzymes encoded by ags genes. The first putative α-D-glucan synthase encoding gene (ags1) was identified in Schizosaccharomyces pombe (Hochstenbach et al. 1998). The ags1 gene encodes a large, three-domain protein. In addition to the multi-pass transmembrane domain in the C-terminal part of the protein, two predicted catalytic domains are present. The middle domain shows strong similarity to glycogen and starch synthases in Glycosyltransferase family (GT) 5 and is predicted to be involved in the synthesis of α-glucan. The N-terminal part of the protein is similar to α-amylases and belongs to family GH13. This part of the protein is predicted to be localized extracellularly, and might be involved in connecting two α-(1,3)-glucan chains (Grün et al. 2005). Apart from the α-glucan synthases, two types of family GH13 enzymes were recently identified in fungi to play a role in fungal cell wall biosynthesis. Marion et al. (Marion et al. 2006) provided evidence for the involvement of a putative α-amylase (Amy1p) in the formation of α-(1,3)-glucan in the cell wall of Histoplasma capsulatum. An AMY1 knockout was unable to produce α-(1,3) glucan and showed reduced virulence. In dimorphic fungi like H. capsulatum, cell wall α-glucan is a known virulence factor (Rappleye et al. 2004;
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Rappleye and Goldman 2006). The second α-amylase-like enzyme Aah3p was first studied in S. pombe (Morita et al. 2006). Disruption of aah3 encoding a GPI-anchored protein resulted in a hypersensitivity towards cell wall degrading enzymes and an aberrant cell shape, indicating that normal cell wall biosynthesis was affected. Disruption of a homologous gene (agtA) in A. niger also affected cell wall stability (van der Kaaij et al. 2007a).

In this study, we surveyed the A. niger genome sequence (Pel et al., 2007) to identify for the first time all GH13, GH15 and GH31 family members present in this important industrial source for amylolytic enzymes. This resulted in identification of a surprisingly large number of previously unknown enzymes. By studying their phylogeny, the presence of specific protein features and synteny with other Aspergillus species, members of each GH family could be divided into several groups. Additionally, we studied the transcriptional regulation of the genes encoding these proteins in a wild type A. niger strain, and in a derived amyR deletion strain, during their growth on xylose and maltose. Only few of the identified proteins were induced by maltose. Expression of many of the identified groups of enzymes, including the homologues of both S. pombe Aah3p and H. capsulatum Amy1p, was not induced by maltose and was not dependent on the presence of AmyR. The possible involvement of these enzymes in cell wall α-glucan synthesis and modification is discussed.

Material and methods

Database mining of A. niger genome and analysis of predicted proteins

The full genome sequence of A. niger strain CBS 513.88, has been deposited at the EMBL database with accession numbers AM270980-AM270998 (Pel et al. 2007) and was used for database mining. The nucleotide accession numbers of A. niger genes as listed in Table 1 refer to this database. Hidden Markov Model (HMM) profiles were built with the HMMER package (Durbin and Eddy 1998) (http://hmmer.wustl.edu/) based on the amino acid sequences of known members of GH13, GH15 and GH31. Proteins belonging to these families, originating from the different kingdoms of life, were retrieved from the CAZy website at http://www.cazy.org (Coutinho and Henrissat 1999), and the protein sequences were extracted from the GenBank/GenPept database at http://www.ncbi.nlm.nih.gov/entrez/ and Swiss-Prot database at http://www.expasy.org/sprot/. The A. niger genome was searched with the HMM profiles using the WISE2 package (Birney et al. 2004) (http://www.ebi.ac.uk/Wise2/).

The presence of signal peptidase cleavage sites, glycosylphosphatidylinositol (GPI) attachment sites and starch binding domains (SBD) in the obtained sequences were predicted by web-based tools at URL: http://www.cbs.dtu.dk/services/SignalP/ (Bendtsen et al. 2004), URL: http://mendel.imp.ac.at/sat/gpi/gpi_server (Eisenhaber et al. 2004), and URL: http://www.ncbi.nlm.nih.gov/BLAST/ (Marchler-Bauer and Bryant 2004), respectively.

Multiple sequence alignments of GH13, 15 and 31 family members were performed using DNAMAN version 4.0 (Lynnon BioSoft, Canada). The alignments were
based on the full length of the predicted proteins, except in case of predicted α-glucan synthases for which only the N-terminal part, encoding the family GH13 domain, was used for the alignment. The phylogenetic relationship was calculated by using Optimal Alignment (Thompson et al. 1994) with gap opening penalty and gap extension penalty of 10 and 0.05, respectively. A bootstrapped test of phylogeny was performed by the Neighbor-Joining method using 1000 replicates. Wherever possible, one protein with described activity was included for each of the groups identified based on phylogenetic analysis.

**Strains and transformations**

The wild type *A. niger* strain used in this study is N402, a *cpsA1* derivative of *A. niger* van Tieghem (CBS 120.49, ATCC 9029) (Bos et al. 1988). Strain AB4.1 is a *pyrG* negative derivative of N402 (van Hartingsveldt et al. 1987) and was used to construct the *amyR* disruption strain. *A. niger* strains were grown in *Aspergillus* minimal medium (MM) (Bennett and Lasure 1991), or *Aspergillus* complete medium (CM) consisting of MM with the addition of 0.5% (w/v) yeast extract and 0.1% (w/v) casamino acids. Growth medium was supplemented with 10 mM uridine (Serva, Germany) when required. Transformation of *A. niger* AB4.1 was performed as described earlier (Punt and van den Hondel 1992) using lysing enzymes (L1412, Sigma, U.S.A.) for protoplastation. The bacterial strain used for transformation and amplification of recombinant DNA was *Escherichia coli* XL1-Blue (Stratagene, U.S.A.). Transformation of XL1-Blue was performed according to the heat shock protocol (Inoue et al. 1990).

**Disruption of the maltose utilization activator gene *amyR* in *A. niger***

Plasmid pJG01 containing the *A. niger* *amyR* gene as a 4.3 kb *NsiI* fragment in pGEM11 was kindly provided by P. van Kuyk (Wageningen University, the Netherlands) and used to disrupt the *amyR* gene. The construction of the *amyR* deletion cassette was performed as follows. The *BamHI-EcoRI* fragment and *NsiI-SalI* fragment flanking the *amyR* ORF at the 5’ and 3’ region respectively were isolated from pJG01. The isolated *NsiI-SalI* fragment was cloned into pUC19 to obtain plasmid pAmyRF3. Subsequently, a *BamHI-SalI* fragment carrying the *A. oryzae pyrG* gene, obtained from plasmid pAO4-13 (de Ruiter-Jacobs et al. 1989) was inserted into pAmyRF3 which resulted in plasmid pAmyRF3-pyrG. The *BamHI-EcoRI* fragment isolated from pJG10 was ligated into pAmyRF3-pyrG resulting in the *amyR* deletion plasmid (pAmyRF). Prior to transformation to AB4.1, pAmyRF was linearized with *EcoRI*. Uridine prototrophic transformants were selected by their ability to grow on MM without uridine. After two rounds of purification, transformants were tested for their ability to grow on starch. Approximately 10% of the *pyrG* transformants showed defective growth on MM agar plates containing starch as sole carbon source. Six independent putative *amyR* deletion strains (YvdM1.1-1.6) with identical phenotypes were obtained. Southern blot analysis confirmed proper deletion and a single integration of the *amyR* disruption cassette
at the amyR locus. Strain YvdM1.1 was used for further analysis and we will refer to this strain as the ΔamyR strain in the remaining of this paper.

**Culture conditions, RNA preparation, microarray analysis and data analysis**

RNA extracted from the *A. niger* ΔamyR strain and its parental strain (N402) grown on different carbon sources were used for microarray experiments using custom made ‘dsmM_ANIGERa_coll’ Affymetrix GeneChip® Microarrays kindly provided by DSM Food Specialties (Delft, The Netherlands). All experiments for each growth condition (culturing the mycelia, RNA extractions and microarray hybridizations) were performed twice as independent biological experiments.

*A. niger* spores (2 x 10⁶ spores ml⁻¹) were inoculated in 250 ml MM supplemented with 2% (w/v) xylose (Sigma) and 0.1% (w/v) casamino acids and grown for 18 h at 30 °C on a rotary shaker at 300 rpm. The mycelium was harvested by suction over a nylon membrane and washed with MM without carbon source. Aliquots of 1.6 g wet weight of mycelium were transferred to 300 ml Erlenmeyer flasks containing 70 ml MM supplemented with 1% (w/v) carbon source (maltose (Sigma) or xylose) and incubated at 30 °C for a further 2 or 8 h. The pH of all cultures grown for 2 h was equal to pH at the time of transfer (pH 6.2). Cultures grown for 8 h were buffered at pH 4 by the addition of 100 mM of citric acid/sodium citrate to allow comparison between the N402 and the ΔamyR strain at this time point as medium of the ΔamyR strain did not acidify as quickly as the medium of the N402 strain. The mycelium was harvested over Miracloth filter, frozen in liquid nitrogen and stored at -80 °C. Total RNA was isolated from mycelia using TRIzol reagent (Invitrogen) and RNA quality was verified by analyzing aliquots with glyoxal/DMSO gel electrophoresis and Agilent Bioanalyzer “Lab on chip” system (Agilent Technologies, U.S.A.).

Processing, labeling and hybridization of cRNA to *A. niger* Affymetrix GeneChips were performed according to the corresponding Affymetrix protocols for “Eukaryotic Target Preparation” and “Eukaryotic Target hybridization”. For probe array washing and staining, the protocol “Antibody Amplification for Eukaryotic Targets” was followed. Hybridized probe array slides were scanned with Agilent technologies G2500A Gene Array Scanner at a 3 μm resolution and a wavelength of 570 nm. Affymetrix Microarray Suite software MAS5.0 was used to calculate the signal and p-values and to set the algorithm’s absolute call flag, which indicates the reliability of the data points according to P (present), M (marginal) and A (absent). The data on each chip were globally scaled to an arbitrary target gene intensity of 500. The complete microarray data were deposited into ArrayExpress with an accession E-TABM-324 at http://www.ebi.ac.uk/miamexpress.

The pre-scaled data from each hybridization experiment was then normalized for statistical analysis using Genespring 7.0 software (Silicon Genetics, U.S.A.). The per chip normalization was performed to ensure that the overall characteristic of the expression distribution such as median should be the same for all the chips. For the genome wide analysis, we focused on maltose induced genes and therefore a pre-filtering of data was performed to select for genes whose detection calls are present in both maltose duplicate samples in the wild type strain (N402). The selected dataset was used to perform a 1-way
Identification and transcriptional regulation of starch and α-glucan acting enzymes

ANOVA analysis under the test type of “parametric test, don’t assume variances equal”. Fold changes in expression between two different conditions were then computed for genes with \( p < 0.08 \) based on 1-way ANOVA analysis.

**Results**

**Identification of glycoside hydrolase family 13, 15 and 31 genes in the *A. niger* CBS 513.88 genome sequence**

\( \alpha \)-Amylases, glucoamylases and \( \alpha \)-glucosidases, members of families GH13, 15 and 31, respectively, are the three main types of enzymes involved in breakdown of starch by aspergilli (Tsukagoshi et al. 2001). To identify all genes encoding enzymes that might play a role in starch utilization, or other \( \alpha \)-glucan modifying processes in *A. niger*, the genome of *A. niger* CBS 513.88 was searched with HMM profiles based on known enzymes from families GH13, 15 and 31. This resulted in the retrieval of a total of 27 protein sequences including 17 previously unknown proteins, listed in Table 1A.

Two approaches were combined to predict putative functions in cellular processes for this surprisingly large number of newly identified proteins. First, phylogenetic trees were constructed using the GH13, GH15 and GH31 family members identified in the *A. niger* genome, as well as functionally characterized proteins from other organisms with similarity to the identified *A. niger* proteins (Fig. 1). Second, using DNA microarrays, the expression of all the *A. niger* genes encoding GH13, GH15 and GH31 enzymes was examined in both the *A. niger* wild type strain N402 and the \( \Delta \text{amyR} \) strain derived, after growth on different carbon sources (Fig. 2 and Table S1). Both the N402 and the \( \Delta \text{amyR} \) strains were pregrown in xylose for 18 h, and mycelia were transferred to either xylose or maltose media and grown further for 2 h or 8 h. Expression levels were determined based on geometric mean data of biological duplicate samples. We will discuss each enzyme family in detail and combine the findings in *A. niger* with the predicted proteins present in the genomes of *A. fumigatus* (Nierman et al. 2005), *A. nidulans* (Galagan et al. 2005) and *A. oryzae* (Machida et al. 2005).
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Fig. 1. Bootstrapped phylogenetic tree of A. niger GH13 (A) and GH31 (B) enzymes and several closest homologues from other species. Newly identified proteins in the genome of A. niger are shown in bold. A description of each protein is listed in Tables 1A,B. Bootstrap values are indicated on the node of each branch. The tree was created with DNAMAN 4.0 using gap and extension penalties of 10 and 0.5 respectively. The scale bar corresponds to a genetic distance of 0.05 substitution per position.
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Table 1A. All members of family GH13, GH15 and GH31 identified in the genome sequence of *A. niger* CBS 513.88 using HMM profiles. The newly identified proteins are indicated in bold.

| Accession no. | Gene   | Family | Enzyme activity        | Features | AmyR binding sites | Proposed biological function              | Ref  |
|--------------|--------|--------|------------------------|----------|--------------------|-------------------------------------------|------|
| An11g03340   | aamA   | GH13   | acid α-amylase         | SS       |                    | Starch degradation                        | 1    |
| An12g06930   | amyA   | GH13   | α-amylase              | SS       | +970; +252         | Starch degradation                        | 2    |
| An05g02100   | amyB   | GH13   | α-amylase              | SS       | +252               | Starch degradation                        | 2    |
| An04g06930   | amyC   | GH13   | α-amylase              | SS       | +787; +664; -531   | Starch degradation                        | 3    |
| An09g03100   | agtA   | GH13   | α-glucanotransferase   | SS, GPI  |                    | Cell wall α-glucan synthesis              | 4    |
| An12g02460   | agtB   | GH13   | α-glucanotransferase   | SS, GPI  | +810               | Cell wall α-glucan synthesis              | 4    |
| An15g07800   | agtC   | GH13   | putative α-glucanotransferase | SS, GPI  |                    | Cell wall α-glucan synthesis              | 4    |
| An02g13240   | agdC   | GH13   | putative α-glucosidase | +368     |                    | Unknown                                   |      |
| An13g03710   | agdD   | GH13   | putative α-glucosidase | +504; -32 |                | Unknown                                   |      |
| An01g13610   | amyD   | GH13   | putative α-amylase     | -76      |                    | Cell wall α-glucan synthesis              |      |
| An09g03110   | amyE   | GH13   | putative α-amylase     | -76      |                    | Cell wall α-glucan synthesis              |      |
| An01g06120   | gdbA   | GH13   | glycan debranching enzyme | -487; +393 |                | Glycogen metabolism                       |      |
| An14g04190   | gbeA   | GH13   | glycan branching enzyme |         |                    | Glycogen metabolism                       |      |
| An04g09890   | agsA   | GH13   | putative α-glucan synthase | SS       | +287               | Cell wall α-glucan synthesis              | 5    |
| An15g07810   | agsB   | GH13   | putative α-glucan synthase | SS       | -973; +622, -185  | Cell wall α-glucan synthesis              | 5    |
| An12g02450   | agsC   | GH13   | putative α-glucan synthase | SS       |                    | Cell wall α-glucan synthesis              | 5    |
| An02g03260   | agsD   | GH13   | putative α-glucan synthase | SS       |                    | Cell wall α-glucan synthesis              | 5    |
| An09g03070   | agsE   | GH13   | putative α-glucan synthase | SS       |                    | Cell wall α-glucan synthesis              | 5    |
| An03g06550   | glaA   | GH15   | glucoamylase           | SS, SBD  | -792; -669; +423; -301 | Starch degradation                       | 6    |
| An12g03710   | glaB   | GH15   | putative glucoamylase  | -878     |                    | Starch degradation                        |      |
| An04g06920   | agdA   | GH31   | α-glucosidase          | SS       | +574; +191         | Starch degradation                        | 7    |
| An01g10930   | agdB   | GH31   | putative α-glucosidase | SS       | +904; -334         | Starch degradation                        |      |
| An09g05880   | agdE   | GH31   | putative α-glucosidase | SS       |                    | Protein glycosylation                     |      |
| An18g05620   | agdF   | GH31   | unknown                |          |                    | Unknown                                   |      |
| An07g00350   | agdG   | GH31   | unknown                | SS       | +402               | Unknown                                   |      |
| An09g03300   | axIA   | GH31   | putative α-xylosidase  | SS       | +126               | Xyloglucan degradation                    |      |
| An01g04880   | axIB   | GH31   | putative α-xylosidase  | +430; +138 |                | Xyloglucan degradation                    |      |

*a* SS = predicted N-terminal Signal Sequence; GPI = predicted Glycosylphosphatidylinositol anchor signal; SBD = predicted starch binding domain.

*b* The presence of consensus AmyR binding motif (CGGNN8(A/C)GG) was analysed in the promoter region up to 1 kb upstream of the start codon.

*c* References: 1 (Boel et al. 1990); 2 (Korman et al. 1990); 3 (Van der Kaaij & Yuan. unpubl.); 4 (van der Kaaij et al. 2007); 5 (Damveld et al. 2005); 6 (Boel et al. 1984); 7 (Nakamura et al. 1997).
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Table 1B. Functionally described family GH13 and GH31 members from other organisms, used for the multiple sequence alignments in Fig. 1. For each *A. niger* protein identified, a functionally or biochemically characterized protein with the highest similarity was used in the phylogenetic analysis.

| Accession no. | Name           | Family | Enzyme activity          | Features*          | Biological function               | Organism      | Ref |
|---------------|----------------|--------|--------------------------|--------------------|-----------------------------------|---------------|-----|
| BAA78714      | AndGbe1        | GH13   | glycogen branching enzyme|                    | Glycogen metabolism              | *A. nidulans* | 1   |
| BAA34996      | ScGdb1         | GH13   | glycogen debranching enzyme|                  | Glycogen metabolism              | *S. cerevisiae* | 2   |
| P19571        | BsAmyA         | GH13   | α-amylase                | SS                 | Starch degradation               | *Bacillus sp.* | 3   |
| CAA54266      | BsAglA         | GH13   | α-glucosidase            |                    | Starch degradation               | *Bacillus sp.* | 4   |
| CAA21237      | SpAah1         | GH13   | unknown                  | SS, GPI            | α-Glucan biosynthesis            | *S. pombe*    | 5   |
| CAA91249      | SpAah2         | GH13   | unknown                  | SS, GPI            | α-Glucan biosynthesis            | *S. pombe*    | 5   |
| CAB40006      | SpAah3         | GH13   | unknown                  | SS, GPI            | α-Glucan biosynthesis            | *S. pombe*    | 5   |
| CAA16864      | SpAah4         | GH13   | unknown                  | SS, GPI            | α-Glucan biosynthesis            | *S. pombe*    | 5   |
| ABK62854      | HcAmy1         | GH13   | unknown                  |                    | α-Glucan biosynthesis            | *H. capsulatum* | 6   |
| ABF50883      | AN7345.2       | GH31   | α/B-glucosidase          | SS                 | Starch/cellulose degradation     | *A. nidulans* | 7   |
| ABF50846      | AN7505.2       | GH31   | α-xyllosidase            |                    | Xylan degradation                | *A. nidulans* | 7   |
| BAB39856      | AndAgdB        | GH31   | α-glucosidase            | SS                 | Starch degradation               | *A. nidulans* | 8   |
| AAU87580      | TrAgull        | GH31   | α-glucosidase II         | SS                 | Protein glycosylation            | *T. reesei*   | 9   |
| A45249        | CAMAL2         | GH31   | maltase                  |                    | Maltose degradation              | *C. albicans* | 10  |

* SS = predicted N-terminal Signal Sequence; GPI = predicted Glycosylphosphatidylinositol anchor signal

* References: 1 (Sasangka et al. 2002); 2 (Teste et al. 2000); 3 (Tsukamoto et al. 1988); 4 (Nakao et al. 1994); 5 (Morita et al. 2006); 6 (Marion et al. 2006); 7 (Bauer et al. 2006); 8 (Kato et al. 2002b); 9 (Geysens et al. 2005); 10 (Geber et al. 1992).
Identification and transcriptional regulation of GH13 family members

The HMmer search for family GH13 enzymes in the A. niger genome resulted in the identification of 18 protein sequences of which 10 had not been identified previously (Table 1A). Table 2 displays the four conserved regions typical for family GH13 proteins as identified in these enzymes. A phylogenetic tree was produced combining the A. niger family GH13 enzymes with several functionally characterized GH13 family proteins from other organisms (Fig. 1). The combination of this phylogenetic analysis with a functional annotation of the proteins revealed 6 recognizable subgroups.

Fig. 2. Expression profiles of A. niger family GH13, 15 and 31 enzymes. Accession numbers of the gene names are given in Table 1. Strain and time points after transfer from the preculture are indicated on the right hand side. The numeric values and Present/Absent calls from the expression data are provided as Supplementary Table 1.

Group I consists of 4 extracellular α-amylases. Three of these are the previously characterized extracellular α-amylases acid-amylase (AamA) and α-amylases AmyA and AmyB (Korman et al. 1990). One new extracellular α-amylase was identified and named AmyC. This protein displays high similarity with the known A. niger α-amylases (74% identity to AmyA and AmyB, 65% identity to AamA). The amyc gene is located in a gene cluster also containing an α-glucosidase gene (agdA) and the amyR gene encoding the AmyR transcription factor. Expression of aamA in A. niger N402 was not detectable in xylose media, but was strongly induced in the presence of maltose (Fig. 2). The expression of aamA was reduced to a non-detectable level in the ΔamyR strain. Expression of amyA and amyB was not detected in A. niger N402 in any of the conditions tested (Fig. 2) (see Discussion). The expression of the newly identified amyc gene was relatively low compared to the aamA gene. At 2 h after transfer from the preculture, the expression level of amyc was
Identification and transcriptional regulation of starch and α-glucan acting enzymes

independent of the carbon source, while after 8 h the expression level was reduced 3 fold on xylose compared to maltose. Additionally, the expression on maltose was reduced 2-3 fold in the ΔamyR strain (Table S1). The presence of three putative AmyR binding elements in the promoter region of amyC further suggests that its expression is controlled by AmyR (Table 1A).

The A. niger acid amylase (encoded by aamA) in the CBS 513.88 strain does not contain a predicted Starch Binding Domain (SBD), while its homologues in A. nidulans, A. fumigatus and other aspergilli contain a full length SBD (Table S2). In addition, the presence of a functional SBD in the AamA protein of A. niger N402 was suggested because of the purification of AamA via its SBD from culture fluid of a N402 ΔglaA strain, and subsequent demonstration of the SBD with specific antibodies (pers. comm. M.F. Coeffet-Le Gal and D. Archer, University of Nottingham, UK). We therefore PCR amplified the aamA gene and its 3’ flanking regions using N402 genomic DNA as a template, determined its DNA sequence and compared it to the aamA gene and its 3’ flanking region from CBS 513.88. The comparison revealed that the aamA gene of A. niger strain N402 does include a SBD and that the CBS513.88 strain harbours a deletion of 230 nucleotides right after the part encoding the (β/α)8 barrel, causing a frame shift and the introduction of a stop codon resulting in a truncated protein (Fig. S2). After the deletion, the DNA sequence continues by encoding part of a predicted SBD, but this part is not translated (Fig. S2). The SBD in AamA found in the A. niger N402 strain is also predicted to be present in A. niger strain ATCC 1015 (Baker, 2006).

Group II contains three putative GPI-anchored enzymes, recently identified as α-glucanotransferases, and named AgtA, AgtB and AgtC, respectively (van der Kaaij et al. 2007a). This subgroup of proteins is characterized by the presence of two hydrophobic signal sequences. The N-terminal signal sequence is predicted to serve for translocation to the endoplasmic reticulum (ER) while the C-terminal sequence is predicted to be replaced by a preassembled glycosylphosphatidylinositol (GPI) anchor in the ER (Orlean 1997). The three enzymes cluster together with the α-amylases in the phylogenetic tree (Fig. 1A), but can be distinguished from the α-amylases by their catalytic domains which are clearly different from the consensus sequence for the α-amylase family. In all three proteins, one or two highly conserved histidines in conserved regions I and IV are replaced by other hydrophilic residues (Table 2). Family GH13 members without these histidine residues, which are part of the active site, are very rare (Uitdehaag et al. 1999). Mutation of these residues generally results in reduced activity or altered reaction specificity of the enzymes (Chang et al. 2003; Leemhuis et al. 2004). Interestingly, both the conserved His-residues in Regions I and IV are also missing in all (putative) α-glucan syntheses (Table 2). Other aspergillus genomes harbour two or three Agt homologues (Fig. S1A, Table S2), all sharing the aberrant conserved regions and predicted GPI-anchoring. In the A. niger genome, both agtB and agtC are located next to genes encoding putative α-glucan syntheses, and this arrangement of genes is conserved in other aspergilli. The agtA gene is constitutively expressed in both the wild type strain N402 and the ΔamyR strain under all growth
Group III consists of two putative, intracellular α-glucosidases, named AgdC (An02g13240) and AgdD (An13g03710). The protein sequences contained all residues commonly conserved in the α-amylase superfamily (Table 2). The predicted intracellular proteins lack clear similarity to any previously characterized fungal protein, although similar enzymes are predicted in *A. oryzae*, *A. nidulans* and *A. fumigatus* (Fig. S1A). Their most related functionally characterized homologue is an α-glucosidase from *Bacillus* sp. SAM1606 (Nakao et al. 1994). Expression of *agdC* was low and not induced on maltose, and expression of *agdD* was not observed (Fig. 2).

Group IV contains two putative intracellular α-amylases, named AmyD (An01g13610) and AmyE (An09g03110), which share 55% identity. AmyD was recently characterized as an α-amylase with low hydrolyzing activity on starch and related substrates (Van der Kaaij et al. 2007b). AmyD and AmyE are similar to a recently identified protein Amy1p from *H. capsulatum*, which was shown to be involved in cell wall α-glucan synthesis. The latter protein has not been characterized biochemically. Functionally characterized enzymes with similarity to this cluster therefore included only bacterial enzymes, of which maltodexrose-forming α-amylase of alkaloiphilic *Bacillus* sp. #707 (Tsukamoto et al. 1988) had the highest similarity. No significant expression of *amyD* or *amyE* was detected in our experiments (Fig. 2). Predicted enzymes highly similar to *A. niger*
AmyD and AmyE were also present in other *Aspergillus* species (Fig. S1A, Table S2). Like the *A. niger* amyE gene, the orthologues in *A. nidulans* (AN3309.3) and in *A. oryzae* (AO003001497) are clustered in the genome with genes encoding α-glucanotransferases and α-glucan synthases (Table S2).

The two *A. niger* proteins in Group V could be reliably annotated as enzymes involved in glycogen metabolism: glycogen branching enzyme (GbeA) and glycogen debranching enzyme (GdbA). A homologue for each of these enzymes is present in the other *Aspergillus* genome sequences. Transcriptional analysis in *A. niger* showed that both genes were expressed both on xylose and maltose and that their expression was unaffected in the ΔamyR strain (Fig. 2). Group VI contains the five predicted α-glucan synthase genes (Damveld et al. 2005). The derived proteins are highly similar (66-83%) to each other and all contain the two catalytic domains (GT5 and GH13) characteristic for these proteins. Both agsB and agsC are clustered in the genome with genes encoding the above mentioned α-glucanotransferases agtC and agtB, respectively. In both cases, the direction of transcription of the pair of genes is such that they can be transcribed from their intergenic region. Expression analysis in our microarray data collection, which includes data from various time points on several carbon sources (glucose, maltose, xylose, inulin, sucrose) indicate that agsC and agtB are co-expressed at later growth stage, independent of the carbon source (data not shown). Expression of both agsB and its neighboring gene agtC was not detected in any of the growth conditions. The expression levels of agsA and agsD were very low, or not detectable. AgsE was highly expressed in all experiments independent of AmyR (Fig. 2). As expected, several proteins with high similarity to the *A. niger* α-glucan synthases are predicted from the other *Aspergillus* genomes, although the number of homologues present is highest in *A. niger* (Fig. S1A, Table S2).

**Identification and transcriptional regulation of GH15 family members**

A HMMer search for GH15 family members in the *A. niger* genome returned two predicted proteins: the previously described glucoamylase GlaA (Boel et al. 1984) and an unknown predicted protein named GlaB (An12g03070). The GlaB protein lacks both an N-terminal signal sequence and an SBD, and displays a low similarity of 26 % with GlaA. These two types of family GH15 enzymes are also recognized in other aspergilli studied (Fig. S1B). All predicted proteins similar to GlaB lack both a signal for secretion and an SBD, features typically present in the fungal GH15 enzymes described to date. *A. niger* glaA was expressed on xylose and strongly induced on maltose, as described previously (Fowler et al. 1990). The induction of glaA was AmyR dependent. Expression of glaB was not detected in any of the conditions tested (Fig. 2).

**Identification and transcriptional regulation of GH31 family members**

HMMer searches in the genome of *A. niger* revealed the presence of seven GH31 family members, of which only one was previously identified as an α-glucosidase (*aglA*) (Nakamura et al. 1997). We propose to name all (putative) α-glucosidases in *A. niger* Agd
enzymes, similar to the nomenclature in *A. nidulans*, and to rename AglA as AgdA, to prevent confusion with α-galactosidases (den Herder et al. 1992). The phylogenetic tree (Fig. 1B) shows the presence of (at least) 4 subgroups within family GH31. In the group of the α-glucosidases, AgdA clusters with AgdB, a predicted extracellular protein. AgdB has some similarity to two *A. nidulans* enzymes: AN7345.2 (62% identity) with both α- and β-glucosidase activity (Bauer et al. 2006) and α-glucosidase B (AndAgdB/ AN8953.3, 53% identity), with strong transglycosylation activity (Kato et al. 2002b). However, none of the true orthologues of AgdB has been characterized (Fig. S1C). Transcript analysis revealed that *agdB* was regulated similar to *agdA*, as the strong induction of both genes in the presence of maltose was dependent on the presence of *amyR*. However, where most AmyR regulated genes (*aamA*, *glaA* and *agdA*) reached their highest level of induction after 8 h growth on maltose, the expression level of *agdB* was decreased after 8 h compared to 2 h (Fig. 2).

From the remaining five GH31 family members, AgdE (An09g05880) shows similarity to *Trichoderma reesei* glucosidase II (TrAguII) (Geysens et al. 2005). This type of α-glucosidases is located in the ER where it is involved in the trimming of α-(1,3)-linked glucose residues from N-glycan core structure Glc3Man9GlcNAc2, which may be attached to proteins designated to be secreted (Geysens et al. 2005). AgdE contains a predicted signal sequence which might serve to direct the protein to the ER. The *agdE* gene was constitutively expressed in all conditions tested (Fig. 2), consistent with its predicted function as an α-glucosidase II. Two additional GH31 family members, AgdF (An18g05620) and AgdG (An07g0350), lack similarity to any functionally described proteins. Expression of both genes was not detected in any of the tested growth conditions.

The two final family GH31 members, named AxlA and AxlB, are highly similar to AN7505.2 from *A. nidulans*, which was recently characterized as an α-xylosidase (Bauer et al. 2006). Gene *axlA* was highly expressed in the presence of xylose, while no expression was detected in maltose grown cultures (Fig. 2). The high expression of *axlA* on xylose further supports its putative function as a xylanolytic enzyme. The gene encoding the putatively intracellular AxlB was expressed at a very low level independent of the carbon source (Fig. 2).

A BLASTP search in the genomes of *A. fumigatus*, *A. nidulans* and *A. oryzae* for family GH31 enzymes resulted in a similar collection of enzymes as identified in *A. niger* (Fig. S1C). Several clusters of orthologous proteins are distinguishable, but the assignment of an enzymatic activity to these clusters is not yet possible due to a lack of well studied homologues for these enzymes, thus requiring biochemical studies in our future work.

**Genome wide analysis of AmyR dependent maltose induced genes using microarrays**

The expression analysis of genes encoding family GH13, GH15 and GH31 enzymes in the *A. niger* genome revealed that the expression of only a limited number of genes was induced by maltose in an AmyR dependent way. In fact, only four genes (*aamA*, *glaA*, *agdA*, and *agdB*) showed the predicted expression pattern for genes encoding enzymes involved in the breakdown of starch (Fig. 2). To identify additional genes with a possible role in starch
metabolism, the expression of all 14,509 predicted open reading frames on the Affymetrix microarray chips was analyzed. Comparison of the transcriptome of the wild type strain (N402), grown for 2 h after transfer on either xylose or maltose, identified 634 genes that were significantly induced by maltose (One-way ANOVA analysis p<0.08 and >2-fold change in expression level). A set of 12 genes was expressed > 2-fold higher in N402 compared to the ΔamyR strain (One-way ANOVA analysis p<0.08) after the transfer of these strains from xylose to maltose. Combining the two gene sets resulted in a collection of 6 genes that were > 2-fold induced on maltose in an AmyR dependent manner (Fig. 3, Table 3). Five of these were assigned to the category of carbohydrate transport and metabolism according to FunCat (Ruepp et al. 2004), including the three genes encoding known extracellular starch degrading enzymes (aamA, glaA and agdA) and a putative α-glucosidase (agdB) (Table 3A). Gene An15g03940, encoding a protein with high similarity (68%) to a Candida intermedia glucose/xylose symporter (Leandro et al. 2006), was 2.3-fold induced by maltose; this induction was not observed in the ΔamyR strain. The sixth induced gene (An11g02550) encodes a protein highly similar (72%) to Kluyveromyces lactis phosphoenolpyruvate carboxykinase (Kitamoto et al. 1998) which functions in gluconeogenesis by catalyzing the conversion of oxaloacetate into phosphoenolpyruvate.

A similar comparative analysis was performed for samples taken for cultures that had grown for 8 h after the transfer. A total of 28 genes were significantly induced by maltose in comparison to xylose, and 161 genes were 2 fold higher expressed in N402 compared to the ΔamyR strain (One-way ANOVA analysis p<0.08). By combining the two sets, we identified 18 genes which were induced by maltose and whose induction was AmyR dependent (Fig. 3, Table 3B). Nine of these genes encode proteins involved in carbohydrate transport and metabolism, from which 5 were also identified as differentially expressed after 2 h (aamA, glaA, agdA, agdB, and An15g03940). Additionally identified genes included amyC, encoding an extracellular α-amylase, three genes encoding putative sugar transporters, and several other genes belonging to various functional categories (Table 3). The results from the microarray experiments were validated using Northern blot analysis for a selected number of amylolytic genes. These genes include glaA, aamA, agdB and amyC. As shown in supplementary Fig. 3, the results of the Northern hybridizations are in good agreement with the microarray data (Suppl. Table 1). The Northern analysis also revealed an additional mRNA of a larger size for the amyC gene, 8 h after transfer. The detection of two different sized mRNAs suggests two different mRNA start sites or different polyadenylation sites, but the exact sequence of the different mRNAs and possible consequences for the gene model have not been addressed so far.

Interestingly, the transcription factor gene amyR was also induced 2 and 8 h after transfer to maltose compared to the transfer to xylose (2.6 fold, P-value 0.014; 2.8-fold, P-value 0.027, respectively), indicating transcriptional regulation of the amyR gene itself. As the AmyR transcription factor is per se missing in the deletion strain, it is not appropriate to include the amyR gene in the group of maltose induced, AmyR dependent genes. However, the 1-kb promoter region of the amyR gene contains two AmyR binding motifs, allowing for the possibility that AmyR can induce its own expression.
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Fig. 3. Results of microarray analysis for maltose induced and AmyR dependent genes. Venn-diagram showing the number of genes induced on maltose compared to xylose in A. niger N402, and the number of genes induced in N402 compared to the ΔamyR deletion strain. The number of genes both induced by maltose and dependent on AmyR is indicated in bold. The maltose induced and AmyR dependent genes which are present in both 2 h and 8 h after transfer from a preculture are shown in the box.

Discussion

In the present study, we have mined the full genome sequence of A. niger strain CBS 513.88 (Pel et al. 2007) for the presence of glycoside hydrolases belonging to families GH13, GH15 and GH31. Members of these protein families in aspergilli, including A. niger, have been studied extensively, mainly because of their industrial relevance. Nevertheless, our study revealed the presence of several groups of enzymes that had not been identified previously. These novel enzymes are conserved among several Aspergillus species as well as other Ascomycetes, indicating that they may play an important role in fungal metabolism.

The GH13 family in A. niger contains six separate groups of amylase-type enzymes (Fig. 1), of which three groups had not been described thus far. The best described group comprises the extracellular α-amylases, and is now extended with AmyC in addition to the three known α-amylases. Overexpression of the amyC gene in A. niger resulted in increased levels of endo-α-amylase activity in the medium, indicating that this gene indeed encodes an extracellular α-amylase (Van der Kaaij and Yuan, unpublished results). The relatively low expression of amyC compared to other starch degrading enzymes may explain why this protein has not been identified previously. The localization of the amyC gene in the genome is noteworthy, as it is part of a small cluster of amylytic genes (with agdA) and their transcriptional regulator gene amyR. A similar organization is observed in the genomes of other aspergilli such as A. nidulans and A. oryzae RIB40 (Gomi et al. 2000). In A. fumigatus the same cluster is extended with a glucoamylase (similar to GlaA). Gene clusters of transcriptionally co-regulated genes in filamentous fungi are often involved in the same process, e.g. secondary metabolite production (Woloshuk et al. 1995) or catabolism of amino acids (Hull et al. 1989). A possible function of AmyC might be to act as a scouting enzyme for the presence of starch, resulting in the subsequent activation of AmyR by starch derived molecules (maltose or isomaltose). Alternatively, AmyC might be regulated by another system additional to AmyR, e.g. pH regulated expression, or its expression might be upregulated locally, e.g. in hyphal tips.
Identification and transcriptional regulation of starch and \( \alpha \)-glucan acting enzymes

**Table 3A**: Overview of 6 maltose induced and AmyR dependent genes in *A. niger* at 2 h after transfer from preculture.

| ORF no.\(^a\) | N402 maltose\(^b\) | call\(^c\) | N402 xylose | call \(\Delta \text{AmyR}\) maltose | call \(\Delta \text{AmyR}\) xylose | Fold induction maltose vs xylose | N402 vs \(\Delta \text{AmyR}\) | P-value | Putative function |
|---------------|-------------------|----------|-------------|-----------------|----------------|-----------------|----------------|---------|------------------|
| *Carbohydrate transport and metabolism* |
| An11g03340    | 3.15 ± 0.12       | P        | 0.08 ± 0.04 | A               | 0.13 ± 0.09    | A               | 38.6           | 24.2    | 0.080 acid alpha-amylase AamA |
| An04g06920    | 85.96 ± 22.54     | P        | 6.24 ± 0.98 | P               | 5.03 ± 0.03    | P               | 13.8           | 17.1    | 0.064 extracellular alpha-glucosidase AgdA |
| An01g10930    | 96.50 ± 31.50     | P        | 10.47 ± 2.59| P               | 8.35 ± 2.13    | P               | 9.2            | 11.5    | 0.062 extracellular alpha-glucosidase AgdB |
| An03g06550    | 117.80 ± 43.16    | P        | 13.78 ± 4.14| P               | 5.85 ± 0.47    | P               | 8.5            | 20.1    | 0.072 glucoamylase GlaA |
| An15g03940    | 152.30 ± 19.00    | P        | 65.74 ± 3.04| P               | 70.59 ± 2.88   | P               | 2.3            | 2.2     | 0.074 putative monosaccharide transporter |
| *Energy*      |
| An11g02550    | 58.87 ± 3.86      | P        | 10.13 ± 1.41| P               | 25.27 ± 5.35   | P               | 5.8            | 2.3     | 0.033 putative phosphoenolpyruvate carboxykinase |

\(^a\) Gene name in bold indicates that the gene is >2-fold induced on maltose in an AmyR dependant way after both 2 and 8 hours.

\(^b\) The expression level was based on the geometric mean value of the duplicate samples and the deviation values between the duplicate samples are indicated. The P-value was based on 1-way ANOVA analysis.

\(^c\) P, M or A represent detection calls for present, marginal or absent, respectively. Genes were divided over different functional groups according to FunCat (Ruepp et al. 2004).
Identification and transcriptional regulation of starch and α-glucan acting enzymes

Table 3B. Overview of the 18 maltose induced and AmyR dependent genes in *A. niger* at 8 h after transfer from preculture.

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
|         | P | A | P | P | A | A | 113.0 | 808.4 | 0.004 | acid alpha-amylase AamA |
| An1g03340 | 65.26 ± 7.48 | 0.58 ± 0.57 | 0.08 ± 0.01 | 6.7 61.9 0.012 | extracellular alpha-glucosidase AgdB |
| An01g10930 | 49.46 ± 3.42 | 2.37 ± 0.08 | 6.84 ± 1.99 | 20.9 7.2 0.006 | extracellular alpha-glucosidase AgdA |
| An04g06920 | 179.10 ± 13.90 | 9.89 ± 3.63 | 3.36 ± 0.46 | 17.6 51.7 0.012 | extracellular alpha-glucosidase AgdA |
| An02g03540 | 62.79 ± 8.36 | 8.81 ± 0.61 | 0.96 ± 0.82 | 6.7 61.9 0.012 | putative hexose transport protein MstC |
| An15g03940 | 144.40 ± 6.80 | 35.93 ± 2.72 | 3.04 ± 0.03 | 4.0 47.6 0.003 | putative monosaccharide transporter |
| An03g06550 | 239.30 ± 2.05 | 83.03 ± 4.35 | 19.03 ± 1.75 | 2.9 12.6 0.012 | glucoamylase GlaA |
| An04g06930 | 4.36 ± 0.50 | 1.50 ± 0.16 | 1.40 ± 0.03 | 2.9 3.1 0.062 | extracellular alpha-amylase AmyC |
| An09g04810 | 18.47 ± 0.44 | 6.46 ± 0.35 | 2.52 ± 0.02 | 2.9 7.4 0.003 | putative hexose transporter |
| An12g07450 | 137.60 ± 13.00 | 51.57 ± 11.82 | 52.50 ± 6.31 | 2.7 2.6 0.064 | sugar/H+ symporter MstA |

Energy

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
| An03g06270 | 4.72 ± 0.29 | 1.59 ± 0.14 | 0.20 ± 0.25 | 3.3 25.6 0.035 | putative isoamyl alcohol oxidase |
| An16g06010 | 5.18 ± 1.55 | 1.80 ± 0.39 | 0.22 ± 0.02 | 2.9 23.8 0.032 | putative phosphoglycerate mutase |

Amino acid transport and metabolism

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
| An03g00280 | 8.35 ± 0.90 | 3.57 ± 0.74 | 0.12 ± 0.06 | 2.3 68.0 0.050 | similarity to tyrosinase protein |

Cell rescue, defence and virulence

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
| An04g05730 | 3.99 ± 0.59 | 0.23 ± 0.03 | 1.37 ± 0.73 | 18.8 3.2 0.037 | similarity to integral membrane protein |
| An03g00290 | 5.96 ± 0.89 | 1.68 ± 0.47 | 0.64 ± 0.21 | 3.5 9.3 0.068 | similarity to integral membrane protein |
| An06g00490 | 8.14 ± 0.89 | 3.81 ± 0.30 | 2.23 ± 0.44 | 2.1 3.7 0.066 | similarity to integral membrane protein |

Protein fate

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
| An18g04260 | 5.44 ± 0.40 | 2.57 ± 0.07 | 1.77 ± 0.28 | 2.1 3.1 0.056 | similarity to UDP-galactose transporter |

unclassified

| ORF no. | N402 maltose | N402 xylose | ΔamyR maltose | N402 vs xylose | ΔamyR | Fold induction maltose vs xylose | P-value | Putative function |
|---------|--------------|-------------|----------------|----------------|-------|-------------------------------|---------|------------------|
| An16g01290 | 15.04 ± 0.28 | 5.22 ± 1.69 | 1.97 ± 0.26 | 2.6 6.9 0.027 | unknown |
| An09g06130 | 12.62 ± 2.51 | 5.41 ± 0.93 | 2.69 ± 0.30 | 2.3 4.7 0.059 | unknown |

*abc* see for footnotes Table 3A
The lack of detectable expression of \textit{amyA} and \textit{amyB} in \textit{A. niger} N402 is puzzling. Both genes were expressed in \textit{A. niger} strain CBS 513.88 in a fed-batch fermentation using glucose as a carbon source (Pel et al. 2007), but batch cultivation of \textit{A. niger} N402 on glucose did not result in detectable expression in microarray studies (E. Martens, H. Kools, P. Schaap, pers. comm.). These results indicate a difference between both strains with regard to either the presence of \textit{amyA}/B in \textit{A niger} strain N402 or in the transcriptional regulation of \textit{amyA}/B. In the recently released genome sequence of \textit{A. niger} strain ATCC 1015 from the Joint Genome Institute (Baker, 2006) we could not identify orthologues of the AmyA/AmyB proteins suggesting that these genes might be actually absent from the N402 strain. The absence, or lack of expression of \textit{amyA}/amyB, is consistent with the observation that the \textit{A. niger} MGG029\textDeltaaamA strain (N402 background) with gene disruptions for both \textit{aamA} and \textit{glaA} grows very poorly on starch (Weenink et al. 2006), confirming that AmyA and AmyB do not contribute significantly to the amylolytic potential of N402.

In this study we have identified three new groups within the GH13 family that had not been described before in aspergilli: α-amylase-like proteins predicted to be GPI-anchored enzymes (group II) (subfamily GH13_1, as defined by Stam et al. 2006), intracellular α-glucosidases (group III) (unknown subfamily) and intracellular α-amylases (group IV) (subfamily GH13_5). Detailed biochemical studies of two GPI-anchored proteins revealed that they are α-glucanotransferases, acting mainly on maltooligosaccharides (van der Kaaij et al. 2007a). The expression of these genes was not regulated in response to maltose metabolism or AmyR. The Agt enzymes are homologues of a group of four GPI-anchored amylase-like proteins recently identified in \textit{S. pombe} (Aah1-4p), which have similarly aberrant conserved regions (Table 2). A deletion of one of these homologues resulted in aberrant morphology of the cell and an increased sensitivity towards cell wall degrading enzymes (Morita et al. 2006). A similar phenotype was observed for an \textit{A. niger} knockout of \textit{agtA} (van der Kaaij et al. 2007a). Such a phenotype can be explained if α-glucan synthesis is affected. The clustering of \textit{agt} genes with genes encoding α-glucan synthases in all four \textit{Aspergillus} species studied, and the expression pattern of the \textit{A. niger} \textit{agt} genes both support the hypothesis that Agt enzymes play a role in α-glucan synthesis.

\textit{A. niger} contains two putative intracellular α-amylases, AmyD and AmyE, that belong to the recently identified subfamily GH13_5 (Stam et al. 2006). A homologous gene in \textit{H. capsulatum}, \textit{AMY1}, was shown to be essential for the formation of α-(1,3)-glucan in the cell wall (Marion et al. 2006). The possibility of a role for subfamily GH13_5 enzymes in the cell wall of \textit{Aspergillus} is strengthened by their location in the genome: \textit{A. niger}, \textit{A. nidulans} and \textit{A. oryzae} all contain a cluster of genes, containing a homologue of \textit{AMY1} as well as \textit{ags} and \textit{agt} genes (Table S2). Both the GPI-anchored α-glucanotransferases and the family GH13_5 α-amylase like proteins are only present in fungi with cell wall α-glucan and not in the true yeasts, which lack this type of cell wall glucans. However, a study with heterologously expressed AmyD did not confirm a role for this enzyme in α-(1,3)-glucan formation, as the enzyme only showed a (low) hydrolyzing activity on substrates with α-(1,4) glycosidic bonds \textit{in vitro} (van der Kaaij et al. 2007b). The expression levels of \textit{amyD}
and amyE were below detection level, and thus could not confirm nor refute the possible involvement of these proteins in cell wall α-glucan synthesis in A. niger.

At this moment, no function can be assigned to the predicted intracellular α-glucosidases AgdC and AgdD and their homologues identified in other fungi. The enzymes show similarity to a Bacillus α-glucosidase as well as to a maltase from C. albicans, and might therefore be involved in the intracellular hydrolysis of maltose. Alternatively, these enzymes could play a role in trehalose metabolism, as they show some similarity to bacterial trehalose-6-phosphate hydrolases. Trehalose is known to be a common reserve carbohydrate in fungi including A. niger (Wolschek and Kubicek 1997; Arguelles 2000).

A surprising finding from the genome mining of A. niger was the high number of GH enzymes with a predicted intracellular location. This group included not only the starch-acting enzymes described here, but also other predicted intracellular glycoside hydrolases such as β-glucosidase, chitinase, β-mannosidase, β-xyllosidase, rhamnosidase and invertase, present in various Aspergillus genomes (Pel et al. 2007; Goosen et al. 2007). Most of these proteins were predicted to be exo-acting, i.e. releasing terminal mono- or disaccharide residues from the non-reducing end of the substrate. In A. niger, the genes encoding these proteins are expressed at a relatively low level compared to the extracellular starch degrading enzymes, sometimes even below the detection limit. In a recent publication, expression of intracellular chitinases in A. nidulans was found to be induced during the autolysis phase of the culture (Yamazaki et al. 2007). A role for these and other predicted intracellular proteins during the autolysis phase would explain their low expression levels observed in A. niger during vegetative growth, as well as their lack of secretion signals.

The genome wide expression analysis identified only five genes that were >2-fold induced by maltose in an AmyR dependent way, 2 and 8 h after the transfer to maltose. Among them are genes encoding known starch degrading enzymes (aamA, glaA), two (putative) α-glucosidases (agdA, agdB), and a putative sugar transporter protein (An15g03940). The specificity of the putative sugar transporter encoded by the An15g03940 gene has not been examined, but its expression profile suggests that it might function as a glucose or maltose transporter. Three additional sugar transporters were higher expressed on maltose than on xylose after 8 h (Table 3B). One of them, An12g07450/MstA has been functionally characterized as a high affinity glucose transporter (Vankuyk et al. 2004) which is in line with its expression profile. The two additional sugar transporters encoding genes (An02g03540/mstC and An09g04810) showed an interesting expression pattern, with induced expression on maltose but also relatively high expression levels on xylose (Table 3B). In the ΔamyR strain both their expression levels were strongly reduced, also relative to the xylose conditions, indicating that the deletion of AmyR might have had an indirect effect. A possible explanation is that disruption of the AmyR transcription factor resulted in low levels of extracellular enzymes converting maltose to glucose, and consequently stress due to low glucose levels. The low availability of glucose might be a signal to downregulate mstC and An09g04810 if these genes encode low affinity sugar transporters. Interestingly, the amyR gene itself was also induced by the presence of maltose
suggesting that its regulation takes place, at least partly, at the transcriptional level. Furthermore, conserved AmyR binding motifs are present in the promoters of all co-regulated genes, including the 1-kb amyR promoter region, with the exception of aamA (Table 1A). Possibly the aamA promoter contains a currently unknown AmyR binding motif.

The results presented in this paper suggest that A. niger can metabolize maltose by inducing the expression of a limited amount of enzymes and sugar transporters. Other members of the GH13, GH15 and GH31 protein families might function in starch metabolism in different conditions (solid state growth, or influenced by pH, nitrogen metabolism, etc.), or they may be involved in other processes such as the synthesis of cell wall α-glucan, or the glycosylation of proteins. To determine the exact function of these enzymes will require further biochemical characterization in combination with the detailed analysis of gene deletion mutants.

**Acknowledgements**

We thank Harrie Kools (Wageningen University, The Netherlands) for performing the HMM searches, Patricia van Kuyk (Wageningen University, The Netherlands) for the amyR plasmid , Marie-Françoise Coeffet-Le Gal and David Archer for sharing unpublished results. This work was financially supported by SenterNovem in the framework of the Dutch IOP-Genomics program (project IGE1021).
Supplementary Data

Suppl. Fig. 1. Bootstrapped phylogenetic trees of family GH13 (A), GH15 (B), and GH31 (C) enzymes from *A. niger*, *A. fumigatus*, *A. nidulans* and *A. oryzae*. Bootstrap values are indicated on the node of each branch. The tree was created with DNAMAN 4.0 using gap and extension penalties of 10 and 0.5 respectively. The scale bar corresponds to a genetic distance of 0.05 substitution per position. A description of each protein is listed in Supplementary Table 2.
Identification and transcriptional regulation of starch and α-glucan acting enzymes

**Suppl. Fig. 2.** Alignments showing different AmyA homologues from *A. foetidus* (CBS115.52), *A. vanedensis* (CBS113.365), *A. niger* N402 (CBS120.49, ATCC1015), *A. niger* DSM (CBS513.88) and *A. niger* (NRRL3122). (A) Alignment of the 3’ end of *aamA* and homologues. The stop codon is indicated in bold, non-coding DNA sequences are underlined. (B) Alignment of the C-terminal part of AamA and homologous proteins. The C-terminal end is indicated with *.
Suppl. Fig. 3. Northern blot analysis of representative amylolytic genes in wild-type (N402) and the ΔamyR strain from mycelia transferred from xylose (preculture) to maltose. Total RNA was isolated from mycelia growing in maltose 2 h and 8 h after the transfer as indicated.

Suppl. Table 1. The numeric values, Present/Absent (P/A) calls, and calculated fold-changes between different conditions for all genes encoding family GH13, 15 and 31 enzymes identified in A. niger.
### Suppl. Table 2. Inventory of all family GH13, 15 and 31 enzymes in *A. fumigatus, A. nidulans* and *A. oryzae*.

| Aspergillus   | ORF no. | Accession no.     | Closest A. niger homologuea Featuresb | Featuresb |
|---------------|---------|-------------------|--------------------------------------|-----------|
| **GH13 family** |         |                   |                                      |           |
| A. fumigatus  |         |                   |                                      |           |
| Group I       |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. nidulans   |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| **GH15 family** |         |                   |                                      |           |
| A. fumigatus  |         |                   |                                      |           |
| Group I       |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| **GH31 family** |         |                   |                                      |           |
| A. fumigatus  |         |                   |                                      |           |
| Group I       |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |
| A. niger      |         |                   |                                      |           |
| A. oryzae     |         |                   |                                      |           |

*a* Closest homologues from other aspergilli were obtained by using full length sequences of A. niger GH13, 15 and GH31 family proteins in a BlastP search.

*b* SS = predicted N-terminal Signal Sequence; GPI = predicted Glycosylphosphatidylinositol anchor signal; SBD = predicted starch binding domain.

*c* These proteins are clustered with α-glucan synthases in the genome.

*d* Annotation improved manually.

*e* An earlier start codon was identified resulting in a predicted signal sequence for excretion which was lacking in the original annotation.

*f* GPI-anchoring was not predicted by the online tool of Eisenhaber et al. (2004) but according to the consensuses sequence described by De Grooth et al. (2003), a GPI-anchoring signal may be present.