Isolation and Characterization of Two Genes That Encode Active Glucoamylase Without a Starch Binding Domain from *Rhizopus oryzae*

Jeffrey A. Mertens,[1] Christopher D. Skory[1,2]

[1]Fermentation Biotechnology Research Unit, National Center for Agricultural Utilization Research, USDA, Agricultural Research Service, 1815 North University Street, Peoria, IL 61604, USA
[2]Bioproducts and Biocatalysis Group Research Unit, National Center for Agricultural Utilization Research, USDA, Agricultural Research Service, 1815 North University Street, Peoria, IL 61604, USA

Abstract. Glucoamylase obtained from *Rhizopus* sp. is frequently preferred for certain applications of starch modification or saccharification. The predominant enzyme, which contains a starch-binding domain on the amino terminus, has been previously characterized from several species. Additionally, the cDNA encoding this protein was cloned and found to show 100% identity to a *R. oryzae* strain 99-880 that has recently been sequenced by the Broad Institute of Harvard and Massachusetts Institute of Technology. Analysis of this genome indicates coding regions for two additional glucoamylase genes, *amyC* and *amyD*, lacking a starch-binding domain. The two genes encode proteins that are approximately 50% identical to the catalytic region of the AmyA protein and 67% identical to each other. The predicted AmyC and AmyD proteins contain the highly conserved signature sequences of Family 15 glycoside hydrolases. The two genes appear to be transcriptionally expressed in cultures grown in fermentable and gluconeogenic carbon sources. The predicted 49.7-kD AmyC and 48.8-kD AmyD proteins were expressed in several different ways using *Pichia pastoris*. When the sequence for putative secretion signal was left intact, glucoamylase activity was detected in the crude cell extracts, but no activity was present in the growth medium. However, replacement of this region with the yeast alpha-secretion signal resulted in secretion of active glucoamylase that was able to degrade soluble starch.

Glucoamylase (α-1,4-glucan glucohydrolase, amyloglucosidase, EC 3.2.1.3) has long been of considerable importance to the fermentation and food industries for saccharification of starch/amylopectin. The enzyme consecutively hydrolyzes $\alpha(1\text{-}4)$ glycosidic bonds from the nonreducing ends of starch, resulting in the production of glucose. It also has a limited ability to produce glucose through the hydrolysis of amylopectin $\alpha(1\text{-}6)$ linkages. Many commercial glucoamylase enzyme preparations are derived from *Aspergillus* and *Rhizopus* sp., due to their near complete conversion of starch to glucose.

Glucoamylase has been isolated and characterized from a number of *Rhizopus* species [12–14]. Takahashi et al. [12, 13] identified three different glucoamylase enzymes with the ability to digest gelatinized starch in a commercial *Rhizopus* extract. However, only the largest of the three enzymes was able to bind raw starch. Partial protein sequencing later demonstrated the two smaller fragments were likely the result of limited proteolytic cleavage [12, 13].

It is now widely accepted that *R. oryzae* (syn. *R. arrhizus*) can be divided into two different taxonomic groups, Type-I and Type-II, based on differences in phenotypes (e.g., acid production), rDNA internal transcribed regions [1], and genomic structure [10]. A cDNA fragment encoding the predominant glucoamylase protein, AmyA, with a starch-binding domain (SBD) on the amino terminus was isolated from *R.*

oryzae (syn. *R. arrhizus*) [2]. It appears that this gene is present in both subgroups of *R. oryzae* [8]. Additionally, *R. oryzae* strain NRRL 395, a Type-I strain, has a second glucoamylase gene highly homologous to *amyA*, except that it lacks the N-terminal SBD [8]. This second glucoamylase gene was found to be transcriptionally regulated in a manner similar to the *amyB* gene of *Aspergillus* [5]. Despite high homology to *amyA* in the catalytic region, no glucoamylase activity could be detected from the recombinant protein, leaving the function of this protein presently unclear.

Recently, the genome of *R. oryzae* strain 99-880 was sequenced as a part of the Fungal Genome Initiative [9]. Analysis of the genome confirms that it has a glucoamylase gene with 100% homology to that described by Ashikari et al. [2]. In addition, a search of the database uncovered two hypothetical genes with high similarity to family 15 glycoside hydrolases, with neither gene containing an N-terminal SBD. To determine whether the two hypothetical proteins are indeed active glucoamylase enzymes, we have amplified and cloned the processed genes from cDNA prepared from *R. oryzae* 99-880. The two genes, *amyC* and *amyD*, while containing the conserved signature sequences of family 15 glycoside hydrolases, are divergent from the glucoamylase genes previously isolated from *R. oryzae*. We demonstrate that both *amyC* and *amyD* encode active enzymes and appear to be transcriptionally expressed in a manner largely different from previously characterized *Rhizopus* glucoamylase genes.

Materials and Methods

Strains and Media. Media components and chemicals were obtained from Sigma Chemical (St. Louis, MO) unless otherwise noted. *R. oryzae* strain 99-880, deposited in the Fungal Genetic Stock Center as FGSC 9543, was the source of RNA for all experiments. Shake flask cultures of 99-880 were grown in 50-mL minimal RZ1 media [11] and the desired carbon source at 1% (wt/vol). Cultures were started by inoculation of approximately 1E+06 spores and then incubated at 30°C, 250 rpm. *Escherichia coli* strain TOP10 (Invitrogen, Carlsbad, CA) was used for construction and propagation of cDNA clones. The wild-type *Pichia pastoris* strain X-33 (Invitrogen) was used for protein expression studies.

Isolation of *amyC* and *amyD* cDNA. The two glucoamylase genes were isolated by polymerase chain reaction (PCR) amplification of cDNA prepared from RNA isolated from *R. oryzae* 99-880 cultured for 24 h, as above, in the presence of soluble potato starch. Primers were designed according to locus RO3G_00127.1, *amyC*, and locus RO3G_06101.1, *amyD*, of the *Rhizopus oryzae* Genome database [9]. The *amyC* cDNA was amplified using Pfu Taq polymerase (Stratagene, LaJolla, CA) with the forward primer, 5′-CGGAATTCATGTTTA TAGGGTATCTGTTACTAAC-3′ and reverse primer 5′-CCTCTAGA CCAGTAGATGTGATCAC-3′. The *amyC* cDNA was also amplified without the putative signal sequence as predicted by the SignalP program [3] using the 5′ primer: 5′-CGGAATTCACCCATATTTTTG

ATATTGTG-3′. The amplification primers contain an engineered *Eco*RI site and *Xba*I site in the forward and reverse primers, respectively. In addition, the reverse primer lacks the native stop codon and is designed to be in frame with the C-terminal 6X His tag found in the *Pichia* pPICZ expression vectors. The *amyD* cDNA was amplified in a similar manner with the forward and reverse primers 5′-CGGAATTCATGTTTATAGG-3′ and 5′-CCTCTAGACCAAAAG CAACCGAGT-3′. The *amyD* cDNA was also amplified without the putative signal sequence as predicted by the SignalP program [3] using the same reverse primer as listed above and the following forward primer: 5′-CGGATTTCAGCAGTATCTTTAAAAAG-3′. The amplified cDNA was TA cloned into pCR2.1 and transformed into *Escherichia coli* TOP10 according to the manufacturer's protocol (Invitrogen; Carlsbad, CA). Individual *amyC* and *amyD* cDNA clones were isolated and sequenced to confirm the fidelity of the sequence.

Expression of *amyC* and *amyD* in *Pichia pastoris*. The *amyC* and *amyD* pCR2.1 constructs were digested with *Eco*RI and *Xba*I and the resulting cDNA fragments were ligated into *Pichia* expression vector pPICZB for internal expression and pPICZalphaA for secreted expression of the enzymes. The *amyC* and *amyD* constructs used for secreted expression using expression vector pPICZalphaA do not contain the putative signal sequence. After sequencing to confirm the fidelity of the constructs, the *amyC* and *amyD* expression constructs were linearized with *Pme*I and transformed into *P. pastoris* strain X-33 (Invitrogen) using electroporation according to the manufacturer's protocol. Transformants were cultured on YPDS + Zeocin plates (1% [wt/vol] yeast extract, 2% [wt/vol] peptone, 2% [wt/vol] dextrose, 1 M sorbitol, and 100 μg/mL Zeocin) at 30°C. After 3 days, 10 isolates of each transformation were chosen at random for expression studies. Transformants were initially grown in 3 mL cultures containing YPD + Zeocin media for 24 h at 250 rpm and 30°C. The cultures were harvested, washed, and transferred to 3 mL of methanol induction media (2.8% [wt/vol] yeast nitrogen base with ammonium sulfate, 100 mM potassium phosphate, pH 6.0, 4×10^{-5}% biotin, and 1% [wt/vol] methanol) to induce expression for 48 h at 30°C and 250 rpm. Individual *Pichia* isolates of *amyC* and *amyD* in the pPICZB and pPICZalphaA were also spotted on plates containing methanol induction media containing 0.25% soluble starch and 1.5% agar. The agar plates were incubated at 30°C for 96 h then stained with KI/I_2 reagent, 0.5 M KI/ 1 mM I_2, then stored at 4°C for 24 h.

Transcriptional Analysis of *amyC* and *amyD*. Total RNA was isolated from mycelia collected from 50-mL shake flask cultures grown for 24 h, 30°C, 250 rpm on RZ1 minimal media containing one of the following energy sources: glucose, glycerol, maltose, fructose arabinose, cellobiose, soluble starch, pullulan, or xylose. The mycelia were disrupted using glass beads [11] and the total RNA purified using the RNeasy Total RNA kit (Qiagen, Valencia, CA) followed by DNase treatment with Turbo DNA-free (Ambion, Austin, TX) for 30 min. RNA quality and quantity was confirmed using the 2100 Bioanalyzer Lab-on-Chip system (Agilent Technologies, Wilmington, DE). Reverse transcription (RT) was performed using the SuperScript III first-strand synthesis system (Invitrogen) and primed using the reverse primers listed above to obtain the *amyC* and *amyD* cDNA. Total RNA, 500 ng, of similar quality isolated from the respective cultures was used as template. The first strand synthesis reaction was carried out at 50°C for 50 min. Upon completion of the RT reaction, 1 μL of RNaseH was added to the reaction and incubated for 20 min at 37°C. The first strand cDNA was then amplified using EconoTaq polymerase (Lucigen, Madison, WI) with 5 μL of the RT reaction as template, and the forward and reverse primers listed above to amplify the full-length *amyC* and *amyD* cDNA. The RT-control reactions were set

```
                           +
1    MFIGYLLLTITTFFFYHAVQQTHIFDIVLKRNNGLDEWISKQMRISFKTILKNIHPPGTTRGFLAASLST
1    MFIGYLLLILAIFGFYYSIQQINGSSIFKKKN-LLDDWISKQNTISFNSILKNINPSGTPRGFVAASLST
                           +                                           *
71   SHPDYFYTWTRDAALVVRVLCDLPETEDNLLKDYVEFQADTQNIPTVCQCLGEPKFNPDGSAFSGPWGRP
70   SYPDYFYTWTRDAALVVRVLSDLPETDDDVLKDYIDFQIHIQNTPTVCHCLGEPKFNPDGTGYTGSWGRP
                                                        ** *
141  QNDGPAERAITFMKIINRFKGKHDDFIRKTIMTALEKDLDYIVRVWEEPCFDLWEEVNGIHFYTLMVMRI
140  QNDGPAERAITFIKIANRLK--NSAYVSQKIVPALEKDLDYILQVWEQPCFDLWEEVDGVHFYTLMAMQR

211  ALMDGYHFFNNKNYQSTANDIQKRLETFWSPEENYIRVTQDEQSGVNKTSGLDVSVLLAASTFSSISEGF
208  ALLDAFDFFNLSKYRSTAYRIQEKIERFWSSEENYIRVTHEVREGANKPSGLDVSVLMAANMFATTREGF

281  FTPASDKILATAAAIESAFLKVYSINKDLKTDLGTAIGRYPEDVYDGYEKSEGNPWFLATAAYAELYYLA
278  FTPGSDKVLATAAAIEKSFSSIYPINKNLQPDLGVAIGRYPEDVYDGYGSSKGNPWFLTTAAYVELYYLA

351  IEEWKQKGIEINKVNKPFFERIAPFDYG--TYAPNSDQWQQIISKVSTEADKFLATVQFHQQANGSMSEQ
348  MKEWKQTGLQINAVNKPFFERFLPIDDHCIYYSPDSQQLKEIINLVSLEADKFLATIQYHQQRNGSMSEQ

419  FNRNTGRMQGARDLTWSHAAFISAVKAKEGLPITPVITST.
418  YSRYDGYMQGARDLTWSHAAFISAIKAREGNSVAF.
```

Fig. 1. Alignment of *Rhizopus oryzae* 99-880 deduced glucoamylase proteins. The upper sequence is the deduced *amyC* amino acid sequence and the lower is the *amyD* deduced amino acid sequence. The outlined residues denote conserved regions of family 15 glycoside hydrolase family members. The asterisks highlight residues required for enzymatic activity. The + symbol indicates the first amino acid of the expected mature peptide.
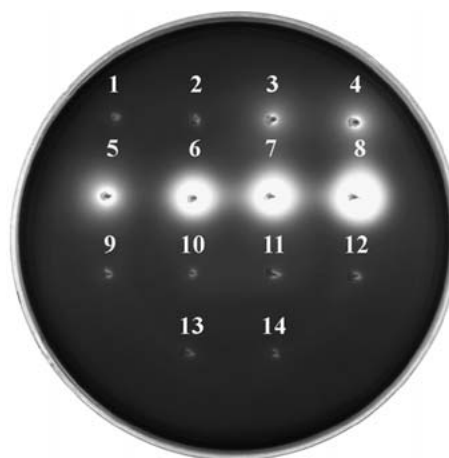
up in a similar manner, except no reverse transcriptase was added to the RT reaction. PCR reactions were performed using the following cycling conditions: 95°C; 30 s, 45°C annealing *amyC*; 46°C, *amyD*; 30 s, and 72°C extension; 1 min 45 s for 30 cycles.

Results and Discussion

Characterization of the Putative Glucoamylase Genes. Two hypothetical proteins with sequence similarity to Family 15 glycoside hydrolases were detected in a search of the *Rhizopus* genome database [9]. The two genes, *amyC* and *amyD*, encoding the hypothetical glucoamylases are 1495 bp and 1471 bp, respectively, with each gene predicted to have two introns located in the same position and spliced in an identical manner [9]. Further sequencing of *amyC* and *amyD* cDNA clones constructed and isolated from RNA of *R. oryzae* 99-880 cultured with soluble starch as the energy source, confirmed the presence and splicing of the two introns. The introns in *amyC* and *amyD* are located in different positions in comparison to the previously isolated *Rhizopus amy* genes. The putative genes do not possess an N-terminal SBD found in the previously isolated *amyA* gene and in this regard, are similar to the recently isolated *amyB* gene from *R. oryzae* NRRL 395 [8]. The *amyC* gene encodes a 459 amino acid protein with a putative 21 amino acid signal sequence for a mature peptide of 438 amino acids and a predicted molecular weight of 49.7 kD (Fig. 1). The *amyD* gene encodes a 453 amino acid protein, but is predicted to have a 24 amino acid signal peptide by the SignalP 3.0 program [3], resulting in a mature peptide of 429 amino acids and a molecular weight of 48.8 kD (Fig. 1). The *amyC* and *amyD* predicted proteins share 67% identity by Lippman Pearson methods [6], but are only 46%

Fig. 2. *amyC* and *amyD* transformants spotted on a starch plate and stained with KI/I₂ as described in Materials and Methods. 1. Untransformed Pichia X-33; 2. pPICZalpha-A vector only; 3–5. *amyC*-pPICZalpha-A isolates; 6–8. *amyD*-pPICZalpha-A isolates. 9–11. *amyC*-pPICZB isolates; 12–14. *amyD*-pPICZB isolates.

and 48% identical, respectively, to the *amyB* gene and the catalytic region of *amyA* gene from *R. oryzae* NRRL 395 (data not shown). While the previously identified *amy* genes are highly homologous, it is not unprecedented for fungal species to have divergent sequences as the *gla* genes of *Aspergillus oryzae* share only 52% identity [4]. Despite the divergence among the three glucoamylase genes lacking the N-terminal starch binding domain and *amyA* from *Rhizopus*, they all share extensive homology with respect to the signature sequences of Family 15 glycoside hydrolases (Fig. 1). The putative AmyC protein contains the conserved residues reportedly required for enzymatic activity. However, the AmyD protein, while containing the conserved tryptophan and consecutive glutamic acid residues, contains an aspartic acid
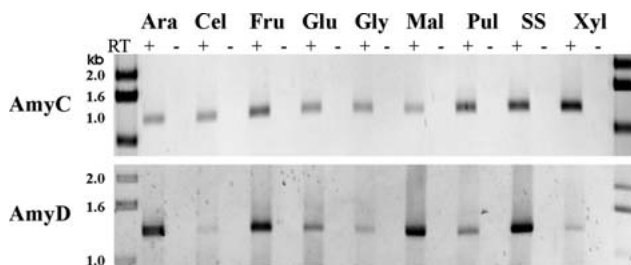
Fig. 3. *amyC* and *amyD* Reverse transcription–polymerase chain reaction assays using total RNA isolated from cultures grown on multiple carbon sources. Assays were performed as described in Materials and Methods and run on a 1% agarose gel. The abbreviations for the carbon sources used in the different cultures are as follows: Ara, arabinose; Cel, cellobiose; Fru, fructose; Glu, glucose; Gly, glycerol; Mal, maltose; Pul, pullulan; SS, soluble starch; and Xyl, xylose.

residue in the active site rather than the conserved asparagine residue (Fig. 1).

Expression of *amy* Genes in *Pichia*. To determine whether the *amyC* and *amyD* genes coded for active enzymes, we cloned the cDNA of the two genes using *amyC* and *amyD* specific primer sets. The PCR amplified 1377 bp *amyC* cDNA and 1359 bp *amyD* cDNA were cloned into *Pichia* expression vector PICZB for internal expression of the enzymes and transformed into *Pichia* strain X-33. After 72 h of methanol induction, cells were harvested and crude cell extracts of *amyC* and *amyD* as well as the culture media of *amyD* were tested for glucoamylase activity. Although activity was detected in the crude cell extracts of *amyC* and *amyD* transformants, no glucoamylase activity was detectable in the culture media of *amyC* or *amyD* transformants (data not shown). Although *Pichia* does recognize the secretion signal of the previously isolated *amy* genes from *Rhizopus* [7, 8], the significantly different signal sequence of *amyC* and *amyD* is apparently not recognized by *Pichia*. Alternatively, it is possible the signal sequence acts to direct these two enzymes to an internal organelle in *Rhizopus*, rather than outside the cell.

In an effort to increase enzyme production and to possibly overcome any potential toxic effects of the AmyC and AmyD proteins toward *Pichia*, the *amyC* and *amyD* cDNA was inserted into *Pichia* expression vector pPICZalpha A for secreted expression. The constructs did not contain the putative signal sequence to avoid potential negative interaction with the alpha mating factor signal sequence, which is in frame with the coding sequence of the two genes. As shown in Fig. 2, secreted *Pichia*-expressed AmyC (Fig. 2, 3–5) and AmyD (Fig. 2, 6–8) were able to degrade starch. The untransformed, as well as the vector-only transformant do not show any evidence of a starch activity (Fig. 2, 1–2). In addition, no starch-degrading activity is demonstrated with the *amyC* and *amyD* constructs designed for internal expression, confirming our earlier results that the *amyC* and *amyD* signal sequence does not result in secreted enzyme in the *Pichia* system (Fig. 2, 9–14).

The *amyD* constructs appear to either produce more enzyme or a more active enzyme based on the radius of the halos produced on the starch plate (Fig. 2). However, based on the work presented here, we cannot make a definitive conclusion.

Transcriptional Analysis of *amyC* and *amyD*. To gain a clearer understanding of the regulation and potential clues to the purpose of the glucoamylase genes, RT-PCR was performed with total RNA isolated from *R. oryzae* 99-880 mycelia grown in shake flask cultures using various energy sources. Amplification product from both genes was detected when 99-880 was cultured in the presence of all nine carbon sources (Fig. 3). The product obtained in each reaction was of the anticipated size for processed transcript and is presumed to be from expression of *amyC* and *amyD*, because no amplification product was detectable in reactions without performing the RT reaction (Fig. 3).

We have demonstrated that the hypothetical Family 15 glycoside hydrolase genes identified in the sequencing of the *R. oryzae* 99-880 do indeed encode for active glucoamylase genes. The *amyC* and *amyD* genes encode enzymes without an N-terminal SBD and may explain differences found in raw starch binding with commercial preparations of *Rhizopus* glucoamylase.

Literature Cited

1. Abe A, Sone T, Sujaya IN, et al. (2003) rDNA ITS sequence of *Rhizopus oryzae*: its application to classification and identification of lactic acid producers. Biosci Biotechnol Biochem 67:1725–1731

2. Ashikari T, Nakamura N, Tanaka Y, et al. (1986) *Rhizopus* raw starch degrading glucoamylase: its cloning and expression in yeast. Agric Biol Chem 50:957–964

3. Bendtsen JD, Nielsen H, vonHeijne G, et al. (2004) Improved prediction signal peptides: SignalP 3.0. J Mol Biol 340:783–795

4. Hata Y, Ishida H, Ichikawa E, et al. (1998) Nucleotide sequence of an alternative glucoamylase-encoding gene (*glaB*) expressed in solid-state culture of *Aspergillus oryzae*. Gene 207:127–134

5. Ishida H, Hata Y, Ichikawa E, et al. (1998) Regulation of the glucoamylase-encoding gene (*glaB*), expressed in solid-state culture (Koji) of *Aspergillus oryzae*. J Ferm Bioeng 86:301–307

6. Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441

7. Liu SH, Chou WI, Sheu CC, Chang MDT (2005) Improved secretory production of glucoamylase in *Pichia pastoris* by combination of genetic manipulations. Biochem Biophys Res Commun 326:817–824

8. Mertens JA, Skory CD (2007) Isolation and characterization of a second glucoamylase gene without a starch binding domain from *Rhizopus oryzae*. Enzyme Microbiol Tech 40:874–880

9. *Rhizopus oryzae* Sequencing Project. Broad Institute of Harvard and MIT (http://www.broad.mit.edu). Accessed August 2006

10. Saito K, Saito A, Ohnishi M, Oda Y (2004) Genetic diversity in *Rhizopus oryzae* strains as revealed by the sequence of lactate dehydrogenase genes. Arch Microbiol 182:30–36

11. Skory CD (2002) Homologous recombination and double strand break repair in the transformation of *Rhizopus oryzae*. Mol Genet Genomics 268:397–406

12. Takahashi T, Tsuchida Y, Irie M (1978) Purification and some properties of three forms of glucoamylase from a *Rhizopus* species. J Biochem 84:1183–1194

13. Takahashi T, Tsuchida Y, Irie M (1982) Isolation of two inactive fragments of *Rhizopus* sp. glucoamylase: relationship among three forms of the enzyme and the isolated fragments. J Biochem 92:1623–1633

14. Yu R, Hang YD (1991) Purification and characterization of a glucoamylase from *Rhizopus oryzae*. Food Chem 40:301–308