Cloning and Characterization of a Gene Encoding a Cell-Bound, Extracellular β-Glucosidase in the Yeast Candida wickerhamii

CHRISTOPHER D. SKORY* AND SHELBY N. FREER

Fermentation Biochemistry Research Unit, National Center for Agricultural Utilization Research, USDA Agricultural Research Service, Peoria, Illinois 61604

Received 23 September 1994/Accepted 9 December 1994

The ability of yeasts to ferment celloextrins is rare. Candida wickerhamii is able to use these sugars for alcohol production because of a cell-bound, extracellular, β-glucosidase that is unusual by not being inhibited by glucose. A cDNA expression library in lambda phage was prepared with mRNA isolated from cellobiose-grown C. wickerhamii. Immunological screening of the library with polyclonal antibodies against purified C. wickerhamii cell-bound, extracellular β-glucosidase yielded 12 positive clones. Restriction endonuclease analysis and sequence data revealed that the clones could be divided into two groups, bgLA and bgLB, which were shown to be genetically distinct by Southern hybridization analyses. Efforts were directed at the study of bgLB since it appeared to code for the cell-bound β-glucosidase. Sequence data from both cDNA and genomic clones showed the absence of introns in bgLB. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis and immunoblotting of cell lysates from Escherichia coli bgLB clones confirmed the presence of an expressed protein with an apparent molecular mass of 72 kDa, which is consistent with that expected for an unglycosylated form of the enzyme. Amino acid comparisons of BglB with other β-glucosidase sequences suggest that it is a member of family 1 glycosyl hydrolases but is unusual in that it contains an additional 100 to 130 amino acids at the N terminus. This sequence did not have homologies to other known protein sequences and may impart unique properties to this β-glucosidase.

The production of fuel ethanol from cellulosic biomass, such as wood, newspaper, and agricultural residues, has been extensively studied from both an economic and a microbiological standpoint. While it is currently possible to produce ethanol with cellulose as the carbon source, economics dictate that the efficiency of production must be improved if this method is to be competitive with starch-based processes. In the current fermentation scheme, the cellulosic feedstock is typically chemically or physically pretreated and then subjected to enzymatic degradation involving a mixture of glucanases that hydrolyze cellulose to cellobiose and celloextrins. These products are subsequently hydrolyzed by β-glucosidase (1,4-β-D-glucoside glucohydrolase, EC 3.2.1.21) to glucose, which can ultimately be converted to ethanol. difficulties in regulating such a system often arise because of accumulation of glucose, which typically leads to end product inhibition of most β-glucosidases (33). This results in an accumulation of cellobiose, which in turn is a potent inhibitor of endoglucanases and exoglucanases (22). Supplementation with additional amounts of β-glucosidase can be used to overcome this rate-limiting step (5, 33). An alternative is to use a cellulase system that has a β-glucosidase which is naturally resistant to end product inhibition (glucose) production. This is a potential strategy that this laboratory is actively exploring.

The yeast Candida wickerhamii ferments soluble celloextrins, with degrees of polymerization of 2 to 6, to ethanol in high yields (8, 10, 11). This organism appears to contain an intracellular β-glucosidase (23), a secreted β-glucosidase (19, 23), and a cell-bound, extracellular β-glucosidase that is released only upon enzymatic digestion of the cell wall (7, 13). The secreted and cell-bound β-glucosidases are highly resistant to glucose inhibition (7, 19). The cell-bound, extracellular β-glucosidase accounts for approximately 90% of the total β-glucosidase activity (7), and this enzyme is primarily responsible for the efficient conversion of celloextrins to glucose (9).

We have isolated a gene (bgLB) from C. wickerhamii that codes for the cell-bound, extracellular β-glucosidase. In this paper we describe the cloning and nucleotide and amino acid sequence analyses of the bgLB gene and compare it with other β-glucosidases.

MATERIALS AND METHODS

Strains and culture conditions. C. wickerhamii (Capriotti) Meyer and Yarrow NRRL Y-2563 (syn. Torulopsis wickerhamii Capriotti) was acquired from the Agricultural Research Service Culture Collection, National Center for Agricultural Utilization Research, Peoria, Ill. Growth of this yeast was carried out at 28°C.

Chemicals. Yeast extract, malt extract, and peptone were bought from Difco Laboratories (Detroit, Mich.). Chemicals, unless specifically referenced otherwise, were purchased from Sigma Chemical Co. (St. Louis, Mo.).

Enzyme assay. β-Glucosidase activity was measured as previously described (7) by incubating appropriately diluted cells or other samples in 2.5 mM p-nitrophenyl-β-D-glucopyranoside in either 50 mM sodium acetate buffer (pH 4.75) or 100 mM phosphate buffer (pH 7.00). One unit of enzyme activity is defined as 1 μmol of product, p-nitrophenol, formed per min.

Isolation and analysis of yeast DNA. DNA was isolated from C. wickerhamii protoplasts by the method of Philippsen et al. (29). Zymolyase 60000 was purchased from Miles Labs (Elkhart, Ind.). Restriction endonuclease (RE) enzymes were purchased from Gibco-BRL (Gaithersburg, Md.) and used according to the supplier’s instructions. Southern hybridization analyses were performed with the Genius System (Boehringer Mannheim, Indianapolis, Ind.) using the manufacturer’s recommendations.

Construction of the C. wickerhamii cDNA library. C. wickerhamii NRRL Y-2563 was grown in yeast extract-peptone-malt extract medium (10% cellbiose) with shaking under anaerobic conditions to promote production of β-glucosidase (12). The cells were harvested by centrifugation at 11,000 × g, washed once with 8% NaCl, frozen in liquid N2, and then ground with glass beads (0.45-mm diameter) with a precooled mortar and pestle. Total RNA was isolated from the disrupted cells by a hot-pheno1 extraction protocol (1) and used to obtain mRNA by further purification with oligo(dT) cellulose purification columns (Gibco-BRL) according to the manufacturer’s recommendations. A cDNA library was prepared from the mRNA with the ZapII cDNA synthesis kit (Strat-
The cDNA library was amplified with Escherichia coli SURE (Stratagene), while in vivo excision of the pBluescript phagemid, containing the cDNA inserts, from the Uni-ZapII XR vector was performed according to the manufacturer's protocol using E. coli XL1-Blue and SOLR (Stratagene). Plasmid pBluescript II (KS-) was transformed by electroporation into E. coli SOLR and served as a control strain containing only vector sequences.

Construction of the C. wickerhamii lambda genomic library. ADashII DNA (Stratagene) was digested with RE XhoI, and then the 5’ overhangs were partially filled in with dTTP and dCTP. Insert DNA was prepared by partially digesting C. wickerhamii DNA with RE Sau3A, and the 5’ overhangs were partially filled in with dGTP and dCTP to create compatible ends with the digested ADashII vector. Ligation, packaging, and amplification were performed according to the manufacturer's recommendations and were followed by packaging with the Gigapack-II XL Lambda packaging extract (Stratagene). E. coli LE392 was used for all manipulations and screening of the lambda genomic library.

Screening colonies and plaques for β-glucosidase activity. Phage were used to infect E. coli SURE and XL1-Blue, which were then inoculated into a top agar medium containing 3 mM isopropyl-thiol-galactoside (IPTG) and 0.5 mM meth­ylumilbellifyl-β-D-glucoside (MUG) or 0.5 mM 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-Glu) as an indicator. The lambda genomic library was screened as described above, except E. coli LE392 was used as the host and IPTG was omitted from the medium. Luria-Bertani (LB) medium was used for both plaque and colony growth. The β-glucosidase activity resulted in a dark-blue formation with X-Glu and a blue fluorescence, visible under long-wave UV (366-nm) light, with MUG.

Production of polyclonal antibodies against purified β-glucosidase protein. Purified β-glucosidase and an antiserum was prepared from New Zealand White rabbits by standard protocols (1) and purified C. wickerhamii cell-bound β-glucosidase (7) as an antigen. Antiserum was purified by precipitation of immunoglobulin G molecules with ammonium sulfate followed by column chromatography using DEAE-AE-Gel Blue resin (Bio-Rad Laboratories, Hercules, Calif.).

Western analysis (immunoblotting) of β-glucosidase protein. The protocol used for detection of phage isolates expressing β-glucosidase was that of Sam­brook et al. (30) with the modification of employing the Genius System (Boehr­inger Mannheim) with Nitro membranes (Schleicher & Schuell, Keene, N.H.) for detection of bound antibodies made against β-glucosidase. Immunological detection was also used to detect the presence of the β-glucosidase protein in E. coli SOLR isolates containing the recombinant phagemids. Isolates were grown to the mid-log phase of growth in LB containing 40 μg of ampicillin per ml washed, resuspended in sodium dodecyl sulfate (SDS)-polyacrylamide gel loading buffer, and lysed with a boiling water bath according to the method of Sambrook et al. (30). Total protein homogenates were separated by SDS-poly­acrylamide gel electrophoresis (PAGE; 10% polyacrylamide) and electroblotted to polyvinylidene difluoride membrane (Bio-Rad) by the methods of Matsudaira (27). The membrane was cut to remove the lanes containing molecular weight markers so that they could be stained with Coomassie R-250. The remaining mem­brane was then used for β-glucosidase detection as discussed above.

Nucleotide sequence analysis. Sequence analysis was performed using the Tag DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.). Reaction mixtures were purified with Quick Spin columns (Boehr­inger Mannheim) and analyzed with an Applied Biosystems 373A DNA sequencer. Oligonucleotide primers M13 and P13 were used for sequencing inserts con­tained in phagemid vector pBluescript II. Primer P13 was chosen as a substitute for the primers reverse M13 and T3, because of its higher thermal denaturation point and ΔC⁰ (5).

Computer analyses of the nucleotide data were performed using the University of Wisconsin Genetics Computer Group package. Protein sequences were com­pared and overlapped in a multiple-sequence format with Pileup. Conservative amino acid substitutions in the multiple-sequence format were defined as those amino acids having a similarity of >0.1 by using the symbol comparison table. PileupPep-cmp, provided with the University of Wisconsin Genetics Computer Group package.

Amino acid sequence analysis. Purified cell-bound, extracellular β-glucosidase from C. wickerhamii (7) was analyzed by the Macromolecular Structure Facility, Michigan State University, East Lansing. The amino-terminal end and several internal fragments obtained by digestion with trypsin were sequenced by the Edman degradation procedure.

Nucleotide sequence accession number. The nucleotide sequence of bglB has been submitted to GenBank under accession number U13672.

RESULTS

Library screening. Approximately 24,000 PFU from the lambda genomic library and over 100,000 PFU from the cDNA library were screened for activity (with and without IPTG included in the growth medium) with MUG as a fluorescent indicator of β-glucosidase activity. This was again repeated with the chromogenic indicator X-Glu. In all cases, no β-glucosidase activity could be detected in any PFU from either the ADashII genomic library or the XZapII cDNA library.

Because of the absence of detectable enzymatic activity, immunological screening was implemented to ascertain whether a recombinant β-glucosidase was present. Conditions for Western analyses using the anti-β-glucosidase antibody were optimized to obtain a detection limit of 1 pg of purified β-glucosidase protein. Screening the XZapII cDNA library using E. coli XL1-Blue cells revealed that 1 of 275 PFU resulted in a strong antibody binding signal over the XZapII control. Twelve of these plaques were isolated and purified for further analysis.

Plasmid analysis of β-glucosidase clones. Plasmids were recovered by in vivo excision, with E. coli SOLR as the recipient, from all 12 XZapII cDNA clones that yielded a positive Western analysis hybridization signal to the anti-β-glucosidase immunoglobulin. RE digestion of the plasmid isolates showed that all but two clones, plasmids pBG1 and pBG28, had inserts of approximately 2 kb in length and virtually the same digest patterns with BamHI, SalI, PstI, and EcoRI. This predom­inant group of clones was designated group B, or bglB (β-glucosidase B). Three of the bglB cDNA clones that appeared to be full length were sequenced in the region of the 5′ end of the insert, and it was found that they differed in length by 24 bp. The largest of the three inserts was from plasmid PBG2. Sequence analysis of the 5′ and 3′ ends of Pbg28 showed that it was an incomplete 0.9-kb bglB cDNA fragment whose sequence matched up identically with analogous regions of Pbg2. The 1.1-kb cDNA insert from plasmid Pbg1 was sequenced and found to be different from any of the bglB clones (data not shown). This clone was designated bglA. Southern analysis confirmed that both bglA and bglB were distinct genes that hybridized to different loci (Fig. 1).

Isolation of genomic bglB clones. The XDashII C. wicker­hamii genomic library was probed with Xhol-linearized Pbg2 which contains the nearly full-length bglB cDNA fragment. Approximately 1 of 320 plaques yielded a strong hybridization signal. DNA was purified from eight of these isolated clones and used for RE digestion and Southern analysis using the

FIG. 1. Southern hybridization analysis of total DNA isolated from C. wick­erhamii, hybridized with XhoI-linearized Pbg2 (bglB [A]) or Pbg1 (bglA [B]) as a probe. Lanes: 1, BamHI digested; 2, EcoRI digested; 3, HindIII digested. Molecular sizes of XHindIII standards are given in kilobases.
FIG. 2. Nucleotide sequence of bglB and deduced amino acid sequence (accession number U13672). Amino acid sequences confirmed by Edman degradation analyses are shown with double underlines and numbered (shown in parentheses) for reference.  , processing site for removal of signal sequence; & , polyadenylation site; /, glutamic acid 514 and aspartic acid 529 proposed to be involved as a nucleophile and acid-base catalyst, respectively.

same probe as above. All clones appeared to have overlapping regions and contained fragments that hybridized strongly to the probe. One of these isolates, LBG2-4, was shown by sequence analysis to possess a 6.5-kb HindIII fragment that contained the entire structural portion of the bglB gene.

Amino acid and nucleotide sequence analyses. Primers P13 and M13 were utilized to sequence the 5' and 3' ends of the cDNA insert contained in plasmid pBG2. Synthetic oligonucleotides specific to the cDNA were then used as primers to continue sequencing inward until overlapping sequences provided an unambiguous nucleotide sequence of the entire cDNA insert. These same primers were again used to sequence

\[ \text{FIG. 2. Nucleotide sequence of bglB and deduced amino acid sequence (accession number U13672). Amino acid sequences confirmed by Edman degradation analyses are shown with double underlines and numbered (shown in parentheses) for reference. , processing site for removal of signal sequence; & , polyadenylation site; /, glutamic acid 514 and aspartic acid 529 proposed to be involved as a nucleophile and acid-base catalyst, respectively.} \]
Further investigation revealed that BglB had approximately the same degree of similarity (50 to 55%) and identity (25 to 30%) to all of the proteins in this family (Fig. 3). They included bacterial \( \beta \)-glucosidases (14, 15, 25, 35, 38), phosphate-\( \beta \)-glucosidases (6, 31), \( \beta \)-galactosidases (24), phosphate-\( \beta \)-galactosidases (3), plant cyanogenic and noncyanogenic \( \beta \)-glucosidases (20, 28), and mammalian lactase phosphoril hydrdrolases (26).

**BglB enzyme characterization and analysis.** None of the 12 \( \lambda \)ZapII cDNA isolates that reacted with anti-\( \beta \)-glucosidase antibody had any detectable \( \beta \)-glucosidase activity associated with the plaques by using MUG-containing LB medium with or without IPTG. However, whole cells of *E. coli* SOLR carrying *bglB* cDNA clones, grown in LB-ampicillin medium, were found to have optimal activity at approximately pH 7 and temperatures between 37 and 40°C. These activities usually were in the range of 5 to 30 mU/ml for the substrate p-nitrophenyl-\( \beta \)-D-glucopyranoside. X-Glu and esculin were also cleaved by the recombinant *bglB* isolates, but no activity against cellobiose was detected. Examination by phase-contrast microscopy of the *bglB* clones grown overnight in LB-ampicillin medium revealed that most cells contained cytoplasmic inclusion granules that occupied approximately 10 to 20% of the cell volume. In attempts to obtain a cell-free enzyme preparation from the *E. coli* SOLR or XLI-Blue *bglB* clones, all detectable \( \beta \)-glucosidase activity was lost when cells were lysed. Methods of breakage included the use of a French press, sonication, and enzymatic treatment with lysozyme followed by deoxycholate or Triton X-100. Various methods that used either 100 mM phosphate (pH 7.00) or 50 mM acetate (pH 4.75) buffer, with or without several different combinations of protein stabilizers (e.g., glycerol, diithiothreitol, and \( \beta \)-mercaptoethanol) and proteinase inhibitors (e.g., phenylmethylsulfonyl fluoride, pepstatin A, and leupeptin), were employed. In all cases, \( \beta \)-glucosidase activity was not detected after cell breakage.

*E. coli* SOLR clones containing plasmids pBG2 and pBG6 expressed a novel 72-kDa protein that was not only partially visible in an SDS-polyacrylamide gel stained with Coomassie R-250 but also readily evident by Western analysis (Fig. 4). The peptide sequence obtained from *E. coli* SOLR or XLI-Blue *bglB* clones, all detectable \( \beta \)-glucosidase activity was lost when cells were lysed. Methods of breakage included the use of a French press, sonication, and enzymatic treatment with lysozyme followed by deoxycholate or Triton X-100. Various methods that used either 100 mM phosphate (pH 7.00) or 50 mM acetate (pH 4.75) buffer, with or without several different combinations of protein stabilizers (e.g., glycerol, diithiothreitol, and \( \beta \)-mercaptoethanol) and proteinase inhibitors (e.g., phenylmethylsulfonyl fluoride, pepstatin A, and leupeptin), were employed. In all cases, \( \beta \)-glucosidase activity was not detected after cell breakage.

The goal of this study was to clone the gene from *C. wickerhamii* responsible for expression of the cell-bound, extracellular \( \beta \)-glucosidase. Previous attempts by this laboratory to clone this gene by using a pBR322 plasmid library and screening *E. coli* transformants for \( \beta \)-glucosidase activity were unsuccessful. It is not known whether difficulties were encountered because of instability of the protein or nonrecognition of the promoter in a gram-negative host. By using a cDNA expression library and immunological screening, two different genes, *bglB* and *bglA*, were isolated. Several convincing lines of evidence confirm that the *bglB* gene encodes the previously purified cell-bound, extracellular enzyme (7). First, the recombinant BglB protein not only reacts with antibodies against the cell-bound, extracellular \( \beta \)-glucosidase but also shows activity...
FIG. 3. Amino acid sequence similarities of several β-glucosidase enzymes found in family 1 of glycosyl hydrolases. Conserved amino acids are shown with a gray background, and identical amino acids have a black background. Bpolyma, Bacillus polymyxa BglA; Ctherma, Clostridium thermocellum BglA; Csacch, Caldocellum saccharolyticum BglA; Bpolymb, B. polymyxa BglB; Agrobact, Agrobacterium sp. Abg; Mbispor, Microbispora bispora BglB; Ecoli, E. coli Bgl; Cwicker, C. wickerhamii processed BglB.

against synthetic and natural β-glucosides. Second, the molecular masses of the deduced peptide sequence and the recombinant BglB protein correspond to that calculated for the native cell-bound, extracellular enzyme by Western analysis. Third, the amino acid composition of the putative peptide sequence is nearly identical to that obtained from the purified native cell-bound protein. Finally, the amino acid sequence from Edman degradation analysis of five different regions, totalling 54 amino acids, of the purified β-glucosidase matched identically the peptide sequence deduced from the isolated bgIB gene. The only exception was the inconsistency of differentiating threonines and glycines by Edman analysis. However, two different examinations of the same region (fragments 1 and 2, Fig. 2) yielded conflicting results concerning the presence of threonines, which strongly suggests faultiness of the Edman analysis procedure.

There were obvious differences between the recombinant BglB and the native cell-bound C. wickerhamii β-glucosidase. The native enzyme cleaves cellobiose and has a pH optimum of 4.75, while the recombinant enzyme could not cleave cellobi-
ose and had a pH optimum of approximately 7. Additionally, the recombinant enzyme was stable only when associated with whole cells, and even then, degradation was often severe. These anomalies may be a consequence of improper folding in the *E. coli* host or a manifestation resulting from the lack of glycosylation. The instability in a cell-free state may explain why we were unable to detect any β-glucosidase activity in the recombinant *A. phage* plaques, whereas activity was observed in *E. coli* clones that contained plasmids with the same cDNA inserts.

The deduced *C. wickerhamii* BglB protein has been found to be a member of family 1 glycosyl hydrolase (17), previously called family BGA (2), which is composed of enzymes from archaea, bacteria, plants, and mammals. This is the first description of a yeast protein belonging to this family.

Closer examination of the amino acid sequence with other family 1 glycosyl hydrolases reveals that Glu-514 in *C. wickerhamii* BglB is positioned in the same relative location as Glu-358 in an *Agrobacterium* sp. This amino acid in the *Agrobacterium* sp. has been determined by site-directed mutagenesis (34, 36) and inhibitor studies (37) to be the nucleophilic amino acid involved in catalysis. It has also been proposed that Asp-374, in the *Agrobacterium* sp., serves as an acid-base catalyst that protonates the leaving group and subsequently deprotonates the water as it hydrolyzes the glycosyl enzyme intermediate (21, 32, 37). This highly conserved amino acid residue occurs in the same location as Asp-529 in *C. wickerhamii* (Fig. 3). The only glycosyl hydrolase from family 1 that differed in this conservation was from creeping white clover (28), which contained an asparagine at this position.

It is interesting to note that most members of family 1 have the conserved sequence Ile-Thr-Glu-Asn-Gly surrounding the glutamic acid nucleophile. Furthermore, Asn and Gly follow immediately after Glu in all of these enzymes described to date, except for *C. wickerhamii* BglB, which contains a Phe in place of Asn. In the *Agrobacterium* sp., it has been demonstrated by site-directed mutagenesis that replacement of Asn-359 with Ser results in only a threefold reduction in β-glucosidase activity (34). However, Ser is a very conservative substitution for Asn, while Phe might be expected to alter the characteristics of this protein region.

Perhaps the most striking difference between BglB and other members of family 1 glycosyl hydrolases is the presence of additional amino acids at the amino terminus of the protein. If the proteins in this family are compared in a multiple-sequence alignment, both the amino and carboxyl termini differ in position by only a few amino acids (Fig. 3). However, the processed BglB contains between 100 and 130 additional amino acids at the amino end. This sequence could be involved in linking the protein to the cell wall or perhaps aids in substrate binding. Similarities to this region could not be found with implied protein sequences in the GenEMBL database.

Preliminary results suggest that *bgIA* codes for the intracellular β-glucosidase. Efforts are currently being aimed at confirming this, as well as improving the stability and expression of BglB in other hosts. High expression of BglB in cellulolytic fungi may yield a non-glucose-inhibited cellulase system. Such a system could be used in cellulose degradation and ultimately improve methods of fuel ethanol production.
ACKNOWLEDGMENTS

We thank Ellen Ball at University of Nebraska, Lincoln, for her efforts with production of the antisera against the purified cell-bound β-glucosidase. The critical review of the manuscript by Robert Hespell is gratefully acknowledged.

REFERENCES


and derived amino acid sequence of the cyanogenic \(\beta\)-glucosidase (linamarase) from white clover (Trifolium repens L.). Plant Mol. Biol. 17:209–219.


