APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 1988, p. 3147–3155
0099-2240/88/123147-09$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 54, No. 12

Nucleotide Sequences of *Saccharomycopsis fibuligera* Genes for Extracellular β-Glucosidases as Expressed in *Saccharomyces cerevisiae*

MAKOTO MACHIDA, ISAO OHTSUKI, SAKUZO FUKUI, AND ICHIRO YAMASHITA*

Center for Gene Science, Hiroshima University, Shitami, Higashi-Hiroshima 724, Japan

Received 20 June 1988/Accepted 26 August 1988

We isolated two genes for extracellular β-glucosidase, *BGL1* and *BGL2*, from the genomic library of the yeast *Saccharomycopsis fibuligera*. Gene products (BGLI and BGLII) were purified from the culture fluids of *Saccharomyces cerevisiae* transformed with *BGL1* and *BGL2*, respectively. Molecular weights of BGLI and BGLII were estimated to be 220,000 and 200,000 by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. The two β-glucosidases showed the same enzymatic characteristics, such as thermodenaturation kinetics and dependencies on pH and temperature, but quite different substrate specificities: BGLI hydrolyzed cellobiose efficiently, but BGLII did not. This result is consistent with the observation that the *S. cerevisiae* transformant carrying *BGL1* fermented cellobiose to ethanol but the transformant carrying *BGL2* did not. Southern blot analysis revealed that the two β-glucosidase genes were derived from *Saccharomycopsis fibuligera* and that the nucleotide sequences of the two genes are closely related. The complete nucleotide sequences of the two genes were determined. *BGL1* and *BGL2* encode 876- and 880-amino-acid proteins which were shown to be highly similar to each other. The putative precursors begin with hydrophobic segments that presumably act as signal sequences for secretion. Amino acid analysis of the purified proteins confirmed that *BGL1* and *BGL2* encode BGLI and BGLII, respectively.

The enzyme β-D-glucosidase catalyzes the hydrolysis of alkyl and aryl β-D-glucosides (e.g., methyl-β-D-glucoside and *p*-nitrophenyl-β-D-glucoside) as well as glycosides containing only carbohydrate residues (e.g., cellobiose). The enzyme is widely distributed, from microorganisms to vertebrates (13). Microbial β-glucosidases play a part in assimilation of cellulose.

Recently, ethanol production by fermentation of cellulose has received much attention as an alternative energy source, since cellulose is available in abundant biomass. During the enzymatic hydrolysis of cellulose, cellobiose is usually accumulated because of the weak β-glucosidase (cellobiase) activity of most cellulolytic microorganisms (13). Although the yeast *Saccharomyces cerevisiae* is most widely used in the ethanol fermentation industry, this organism cannot utilize cellobiose because it does not possess a permease for cellobiose and, with the exception of strain C (6), does not produce an extracellular cellobiase (3, 10).

In recent years, a great deal of interest has developed in genetic manipulation of industrial ethanol-fermenting yeast organisms. We have been very interested in the yeast genes encoding secretable proteins and have developed the ability of *S. cerevisiae* cells to secrete proteins extracellularly. Previously, we constructed *S. cerevisiae* strains capable of fermenting starch to ethanol by introducing either glucoamylase genes or an α-amylase gene (4, 5, 14–17, 19).

It would be interesting to construct a cellobiose-fermenting *S. cerevisiae* strain by introducing a secretable cellobiase gene. We searched for the source of the gene from yeast cells, since we anticipated that gene products derived from closely related organisms would be secreted more efficiently from *S. cerevisiae*. The cellobiose-assimilating yeast *Saccharomycopsis fibuligera* was found to produce an extracellular cellobiase (unpublished data) and is therefore a potential donor of the cellobiase gene.

Although many β-glucosidases were purified from diverse organisms, little is known about the amino acid sequence of the enzyme. However, Kohchi and Toh-e reported the amino acid sequence of *Candida pelliculosa* β-glucosidase, deduced from the nucleotide sequence (7). It has been proposed that an aspartic acid residue plays important roles in the action of β-glucosidase (1), but the overall structure-function relationship of the enzyme is not well understood.

In this paper, we report that *S. cerevisiae* was engineered to ferment cellobiose to ethanol by introduction of a gene coding for an extracellular cellobiase from *Saccharomycopsis fibuligera*. We also describe the nucleotide sequences of *BGL1* and *BGL2* and amino acid analyses of their gene products. These studies led to the conclusion that *BGL1* and *BGL2* are structural genes for two β-glucosidases of different substrate specificities. Comparative studies reveal conserved amino acid sequences in yeast β-glucosidases which may be essential to enzyme function.

MATERIALS AND METHODS

Strains and media. *S. cerevisiae* YIY345 (**a** *leu2 ura3 his4*) was used as a recipient. *Escherichia coli* JA221 (15) was used for plasmid propagation. YEPD, YEPC, and YPGL (18) are rich yeast media containing 1% yeast extract, 2% polypeptone, and, as carbon sources, glucose, cellobiose, and glycerol and lactic acid, respectively. SD minimal medium (12) supplemented with histidine (20 μg/ml) was used for selective growth of yeast cells. The indicator plate for β-glucosidase activity was SD supplemented with 1 mM *p*-nitro-

* Corresponding author.

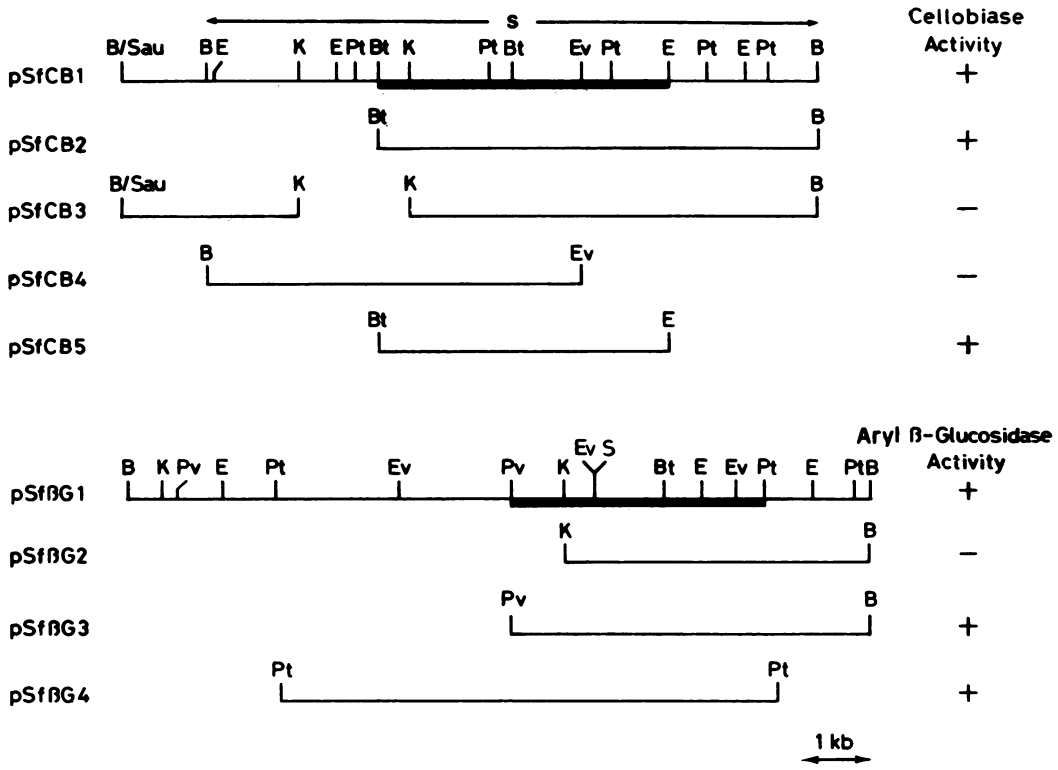


FIG. 1. Restriction maps of cloned segments and localization of essential regions. Plasmids pSfCB1 and pSfβG1 are the original plasmids; the others are subcloned plasmids. The fragment designated S was used as a probe for Southern blot analysis of genomic DNA. The restriction sites for *Bam*HI (B), *Eco*RI (E), *Kpn*I (K), *Pst*I (Pt), *Bst*EII (Bt), *Eco*RV (Ev), *Pvu*II (Pv), *Sal*I (S), and the *Bam*HI-*Sau*3A boundary (B/Sau) are indicated. The yeast transformants carrying each of the plasmids shown were cultured in YEPD. Culture fluids were assayed for β-glucosidase activities with cellobiose and PNPG as substrates. Symbols: +, activities were detected; −, no detectable activities were found. kb, Kilobases.

phenyl-β-D-glucoside (PNPG) and histidine. Bacterial media were described previously (2).

Preparation of DNA and other genetic methods. Plasmid and chromosomal DNAs were prepared as described previously (15). Transformation of yeast and *E. coli* cells and Southern blot analyses were performed as described previously (15). Recombinant plasmid DNA from the *Saccharomycopsis fibuligera* genomic library, which had been constructed by using pYI1 as a vector DNA (15), was used. Plasmid pYI1 carries ampicillin and tetracycline resistance genes for *E. coli* and also *LEU2*, *URA3*, and a replication origin of 2 μm of DNA for *S. cerevisiae*.

TABLE 1. β-Glucosidase activities and fermentation of cellobiose in *S. cerevisiae* transformants[a]

Plasmid	β-Glucosidase activity						Fermentation
	YEPD		YPGL		YEPC		
	Aryl	Cello	Aryl	Cello	Aryl	Cello	
pSfCB1	+	+	+	+	+	+	+
pSfβG1	+	−	+	−	+	−	−
pYI1	−	−	−	−	−	−	−

[a] *S. cerevisiae* transformants carrying pSfCB1, pSfβG1, or pYI1 (the vector plasmid) were cultured with shaking in three media (YEPD, YPGL, and YEPC) for 3 days. Culture fluids were dialyzed extensively against deionized water and assayed for β-glucosidase activity. Aryl β-glucosidase (Aryl) and cellobiase (Cello) activities were determined with PNPG and cellobiose, respectively, as substrates. +, Activities detected; −, no significant activities found. Cell densities, determined at an optical density of 660 nm, were 13 for all cultures in YEPD and YPGL; in YEPC, cell densities were 18.5 for pSfCβ1-carrying transformants, 2.1 for pSfβG1-carrying transformants, and 2.3 for pYI1-carrying transformants. Fermentation tests were carried out by the classic Durham tube method as follows: a loopful amount of cells was inoculated into 5 ml of YEPC with a Durham tube and cultured statically at 28°C for up to 2 weeks. +, Fermenting; −, nonfermenting.

Assay for β-glucosidase activity. The reaction mixture (final volume, 50 μl) contained 4 mM substrate (as described for each experiment), 60 mM McIlvaine buffer (pH 5.0), and the enzyme solution. The reaction mixture was incubated at 30°C. The reaction was stopped, and optical density was determined by one of two methods: (i) for measurement of *p*-nitrophenol, 0.5 ml of 0.25 M Na_2CO_3 was added to the reaction mixture, and optical density at 400 nm was determined; (ii) for measurement of glucose liberated from the substrate, the reaction mixture was boiled for 1 min and then cooled on ice. Glucose was determined with a Gluo-statt assay kit as recommended by the supplier (Fujisawa Yakuhin Kogyo Co., Osaka, Japan). One unit of enzyme activity is defined as the amount of enzyme that hydrolyzed 1 μmol of substrate per min.

Yeast colonies secreting β-glucosidase were easily identified by the yellow halo observed after a solution of 0.25 M Na_2CO_3 was poured onto colonies grown on the indicator plate.

Protein assay. Protein concentration was measured with a protein assay kit (Bio-Rad Laboratories, Richmond, Calif.), with bovine serum albumin as the standard.

TABLE 2. Purification of BGLI and BGLII

Purification step	Total activity[a] (U)		Protein (mg)		Sp act[a] (U/mg of protein)		Yield (%)	
	BGLI	BGLII	BGLI	BGLII	BGLI	BGLII	BGLI	BGLII
Culture fluid	438	3,190	192	150	2.28	21.3	100	100
Ammonium sulfate precipitation	406	3,020	107	95.2	3.81	31.7	93	95
Acetone precipitation	326	2,770	28.4	39.6	11.5	69.9	74	87
DEAE–Sephadex A-50	302	1,150	18.1	17.4	16.7	66.1	69	36
Sepharose 6B	74.3	523	1.69	3.12	44.0	168	17	16

[a] β-Glucosidase activities were determined with PNPG as a substrate.

Determination of molecular weight. Molecular weights were determined by 10% polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE) by the method of Laemmli (8). The gel was stained for protein with Coomassie brilliant blue R-250. A calibration curve was made with erythrocyte membrane proteins.

Purification of β-glucosidase. *S. cerevisiae* transformants carrying either *BGL1* or *BGL2* were cultured in YEPC or YEPD, respectively, at 28°C for 4 days with shaking at 250 rpm on a G10 Gyrotory shaker (New Brunswick Scientific Co., Inc., Edison, N.J.) in four 5-liter Erlenmeyer flasks, each containing 1 liter of medium.

Culture fluid was obtained by centrifugation at 5,000 rpm for 10 min and concentrated fivefold by using a rotary evaporator at room temperature. The solution was brought to 80% saturation by addition of solid ammonium sulfate and left overnight at 4°C. Precipitates were collected by centrifugation. Precipitates remaining in the supernatant were collected by vacuum filtration through a thick bed of Standard Super-Cel (Nakarai Chemicals, Kyoto, Japan) over Toyo no. 2 filter paper (Toyo Roshi Co., Tokyo, Japan) on a Büchner funnel. Both precipitates were dissolved in 10 mM Tris hydrochloride buffer (T buffer), pH 7.0, and dialyzed for 2 days against the same buffer.

Acetone at −20°C was added to the dialyzed solution to a final concentration of 50% (vol/vol). The precipitate was collected by centrifugation, dissolved in T buffer, and dialyzed overnight against the same buffer.

The dialyzed solution was applied to a DEAE–Sephadex A-50 column (1.5 by 11.4 cm) equilibrated with T buffer. After the column was washed with T buffer, the enzyme was eluted with a linear gradient of NaCl (0.0 to 1.0 M) in T buffer. Active fractions were collected, concentrated in a dialysis tube with powder of polyvinylpyrrolidone K-90, and dialyzed overnight against T buffer.

NaCl and glycerol were added to the dialyzed solution to final concentrations of 0.1 M and 10% (vol/vol), respectively. The solution was applied to a Sepharose 6B column (1.5 by 71.6 cm) equilibrated with T buffer containing 0.1 M NaCl. The enzyme was eluted with the same buffer, and the active fractions were collected and concentrated with polyvinylpyrrolidone K-90. The purified enzyme was dialyzed overnight against T buffer.

Amino acid analysis. Amino-terminal sequences of the purified β-glucosidases were determined with an automated protein sequencer (Applied Biosystems, Foster City, Calif.). Amino acid compositions were determined with an amino acid analyzer (Japan Electronic Co.) after hydrolysis of proteins at 110°C for 22 h in 6 N HCl containing 1% thioglycolic acid.

DNA sequence analysis. DNA was sequenced from M13 subclones by the dideoxy chain termination method of Sanger et al. (11).

RESULTS

Cloning of the *Saccharomycopsis fibuligera* β-glucosidase genes. Recombinant plasmid DNA from the *Saccharomycopsis fibuligera* genomic library was used to transform *S. cerevisiae* YIY345 (*leu2 ura3*) to leucine and uracil prototrophy. Transformants carrying a plasmid capable of producing β-glucosidase activity were then selected by their ability to form yellow halos around colonies on the selection plates containing PNPG (the substrate for β-glucosidase). Several halo-forming transformants thus obtained were subcultured to single colonies on YEPD (rich) agar to allow other plasmids present in the transformant to segregate out. From the resulting halo-forming clones, the plasmids were recovered and then transformed again to bacteria by transformation to ampicillin resistance. Recombinant plasmids were purified from bacterial cultures and used to transform the yeast recipient strain (YIY345) to Leu^+ Ura^+. The plasmids transforming the yeast cells to halo forming were selected and identified as plasmids carrying putative β-glucosidase genes. We thus obtained two plasmid DNAs (pSfCB1 and pSfβG1). Restriction maps of the cloned DNAs are shown in Fig. 1.

Fermentation of cellobiose to ethanol by *S. cerevisiae* transformants. *S. cerevisiae* transformants carrying pSfCB1, pSfβG1, or pYI1 (the vector plasmid) were cultured in three media (YEPD, YPGL, and YEPC) (Table 1). The transformant carrying pSfβG1 secreted only aryl β-glucosidase activity, whereas the transformant carrying pSfCB1 secreted both cellobiase and aryl β-glucosidase activities (the latter activ-

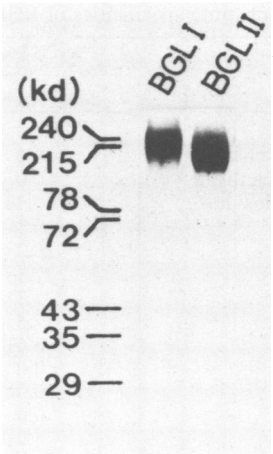


FIG. 2. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of purified β-glucosidases (BGLI and BGLII). Positions of marker proteins are indicated; kd, kilodaltons.

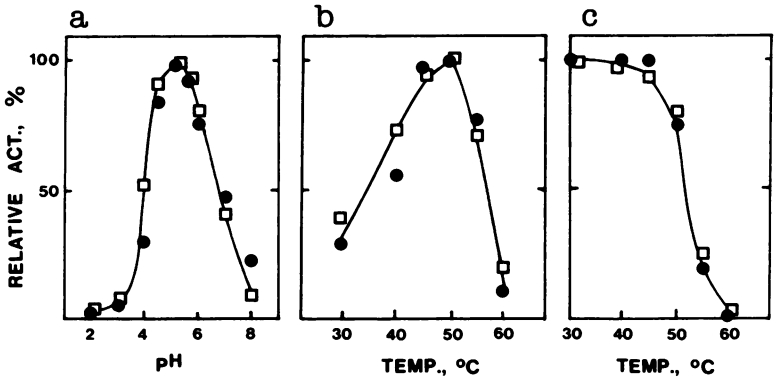


FIG. 3. Dependencies on pH (a) and temperature (b) and thermodenaturation kinetics (c) of BGLI (●) and BGLII (□). β-Glucosidase activity was determined with PNPG as a substrate. pH dependencies were examined in McIlvaine buffer (pH 2.0 to 8.0; final ionic strength was adjusted to 0.1 with KCl) at 30°C. Temperature dependencies were examined at 30 to 60°C in 60 mM McIlvaine buffer (pH 5.0). Activities are presented as percentages of the maximum activity. Thermodenaturation kinetics were examined as follows. Samples were incubated at the indicated temperatures for 30 min and then immediately cooled on ice. The remaining activities were assayed at 30°C. Residual activity after the heat treatment is presented as a percentage of the original activity.

ity was apparently contributed by an aryl β-glucosidase activity of cellobiase, as demonstrated by further investigations). The transformed cells with pSfCB1 grew fully in YEPC and fermented cellobiose in static culture, in which ethanol was accumulated in the medium at a concentration of about 1%, as detected by high-performance liquid chromatography. It is likely that the β-glucosidase activities in the culture fluids were due to secretion but not to leakage from cells, since the activities increased in proportion to cell growth (data not shown). The transformed cells carrying either pSfβG1 or pYI1 grew very little in YEPC or did not ferment cellobiose. These results suggest that pSfCB1 and pSfβG1 encode different types of β-glucosidase: cellobiase and aryl β-glucosidase, respectively.

Analysis of gene products. One of the two types of β-glucosidases (BGLI or BGLII) was purified from the culture fluids of the transformants carrying either pSfCB1 or pSfβG1 (Table 2). Molecular weights of BGLI and BGLII were estimated to be 220,000 and 200,000 by SDS-PAGE (Fig. 2). The two β-glucosidases showed the same enzymatic characteristics, such as dependencies on pH (Fig. 3a) and temperature (Fig. 3b) and thermodenaturation kinetics (Fig. 3c). However, substrate specificities of BGLI and BGLII were quite different. In particular, BGLI hydrolyzed cellobiose and cellooligosaccharides efficiently, but BGLII did not (Table 3). This result is consistent with the observation that the transformed cells carrying pSfCB1 (encoding BGLI) fermented cellobiose but that the transformed cells carrying pSfβG1 (encoding BGLII) did not (Table 1).

TABLE 3. Substrate specificities of BGLI and BGLII

Substrate	Sp act[a] (U/mg of protein)	
	BGLI	BGLII
PNPG	43.3	168
p-Nitrophenyl α-glucoside	0.0	0.0
p-Nitrophenyl β-galactoside	0.0	0.0
p-Nitrophenyl β-xyloside	0.54	2.08
Cellobiose	20.1	0.84
Cellotriose	26.2	1.68
Cellotetraose	27.1	1.46
Gentiobiose	25.7	67.5
Salicin	17.5	14.3
Methyl-β-glucoside	3.61	52.4
Methyl-α-glucoside	0.0	0.0
Maltose	0.0	0.0
Lactose	0.0	0.0
Sucrose	5.01	0.34

[a] β-Glucosidase activities were determined with each of the substrates listed. Average specific activities from at least two independent experiments are presented (deviation was less than 10%).

Southern blot analyses. To localize the region responsible for cellobiase (BGLI) or aryl β-glucosidase (BGLII) activity in plasmid pSfCB1 or pSfβG1, we subcloned restriction endonuclease-digested fragments of the original inserts. The smallest sequences essential for the activities (indicated by

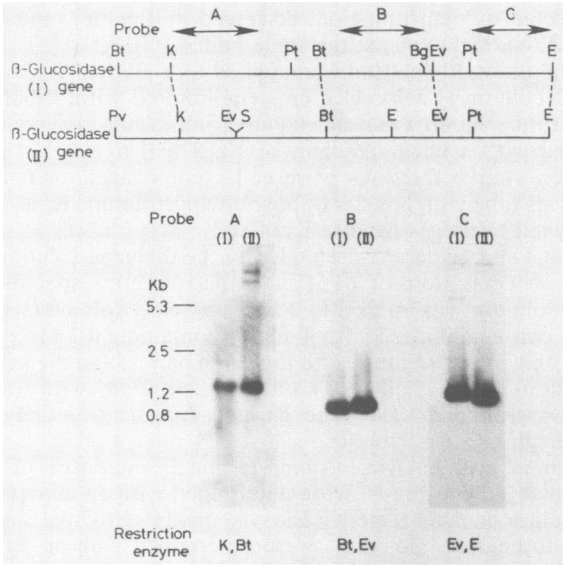


FIG. 4. Southern blot analyses of the plasmids carrying *BGL1* and *BGL2*. Plasmids pSfCB1 (I) and pSfβG1 (II) were digested with the three combinations of restriction enzymes indicated (abbreviations are as in the legend to Fig. 1), electrophoresed, transferred to nitrocellulose papers, and hybridized with the fragments (A, B, and C) of the BGLI-coding sequence. Kb, Kilobases.

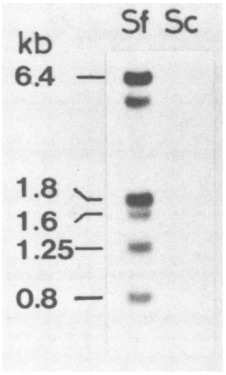


FIG. 5. Southern blot analyses of genomic DNAs from *Saccharomycopsis fibuligera* (Sf) and *S. cerevisiae* (Sc). Genomic DNAs were prepared, digested with *Pst*I, and processed for hybridization with probe S (Fig. 1). kb, Kilobases.

thick lines in Fig. 1) showed similar restriction patterns, such as the sites for *Kpn*I, *Bst*EII, and *Eco*RV. Southern blot analyses were performed to examine whether DNA sequences of the two essential regions showed homology to each other (Fig. 4). Three fragments (designated A, B, and C) derived from the BGLI-encoding sequence were subcloned and used as hybridization probes. Plasmids pSfCB1 and pSfβG1 were digested with three combinations of restriction enzymes (*Kpn*I plus *Bst*EII, *Bst*EII plus *Eco*RV, and *Eco*RV plus *Eco*RI) and probed with the subcloned A, B, and C fragments, respectively. Analyses revealed that the corresponding DNA fragments derived from BGLI- and BGLII-encoding sequences were highly homologous with each other.

We also performed Southern blot analyses to determine whether the two β-glucosidase genes were derived from the genomic DNA of *Saccharomycopsis fibuligera* (Fig. 5). Chromosomal DNA of *Saccharomycopsis fibuligera* was digested with *Pst*I and probed with a subcloned fragment (S in Fig. 1) of pSfCB1. The data showed hybridizations of 1.8-, 1.6-, 1.25-, and 0.8-kilobase fragments derived from the BGLI-encoding sequence and of 6.4- and 1.25-kilobase fragments derived from the BGLII-encoding sequence, confirming that the two β-glucosidase genes were derived from *Saccharomycopsis fibuligera*. A 5-kilobase band could be derived from a 5′ portion of *BGL1*. Chromosomal DNA of *S. cerevisiae* was also processed for hybridization, and no hybridized fragments were detected.

Nucleotide and deduced amino acid sequences of *BGL1* and *BGL2*. Nucleotide and deduced protein sequences of *BGL1* and *BGL2* are shown in Fig. 6. *BGL1* and *BGL2* encode proteins of 876 and 880 amino acid residues with molecular weights of 96,200 and 96,800, respectively. The molecular weights of mature and secreted β-glucosidases (BGLI and BGLII) were found to be 220,000 and 200,000, respectively (Fig. 2), which suggests that a large part of the molecular weights may be contributed by carbohydrate; the deduced proteins encoded by *BGL1* and *BGL2* contain 16 and 12 potential N-glycosylation sites (Asn-X-Thr or Ser; asparagine residues are circled in Fig. 7). The deduced protein sequences were found to be highly similar to each other; about 83% of amino acid residues were identical (Fig. 7).

Amino acid analyses of BGLI and BGLII. The amino-terminal sequences of BGLI and BGLII, which were purified from culture fluids of the *S. cerevisiae* transformants, were determined by Edman degradation. The sequences NH_2-Val-Pro-Ile-Gln-X-Tyr-X-Gln-Ser-Pro-Ser-Gln-X-Asp-Glu-Ser-Ser and NH_2-Leu-Pro-Val-Gln-Thr-His-Asn-Leu-Thr-Asp-Asn-Gln-Gly-Phe-Asp-Glu-Glu-Ser-Ser were obtained for BGLI and BGLII, respectively. These peptides were identical to those predicted from the DNA sequences (underlined in Fig. 6). Amino acid compositions of the purified proteins matched those predicted (data not shown). The correlation between DNA and protein sequences confirmed that *BGL1* and *BGL2* are the structural genes for BGLI and BGLII and showed that both precursors contain amino-terminal extensions of 17 amino acids that are probably cleaved during export of the proteins. The leader sequences resemble signal sequences found in a wide variety of secretory protein precursors.

DISCUSSION

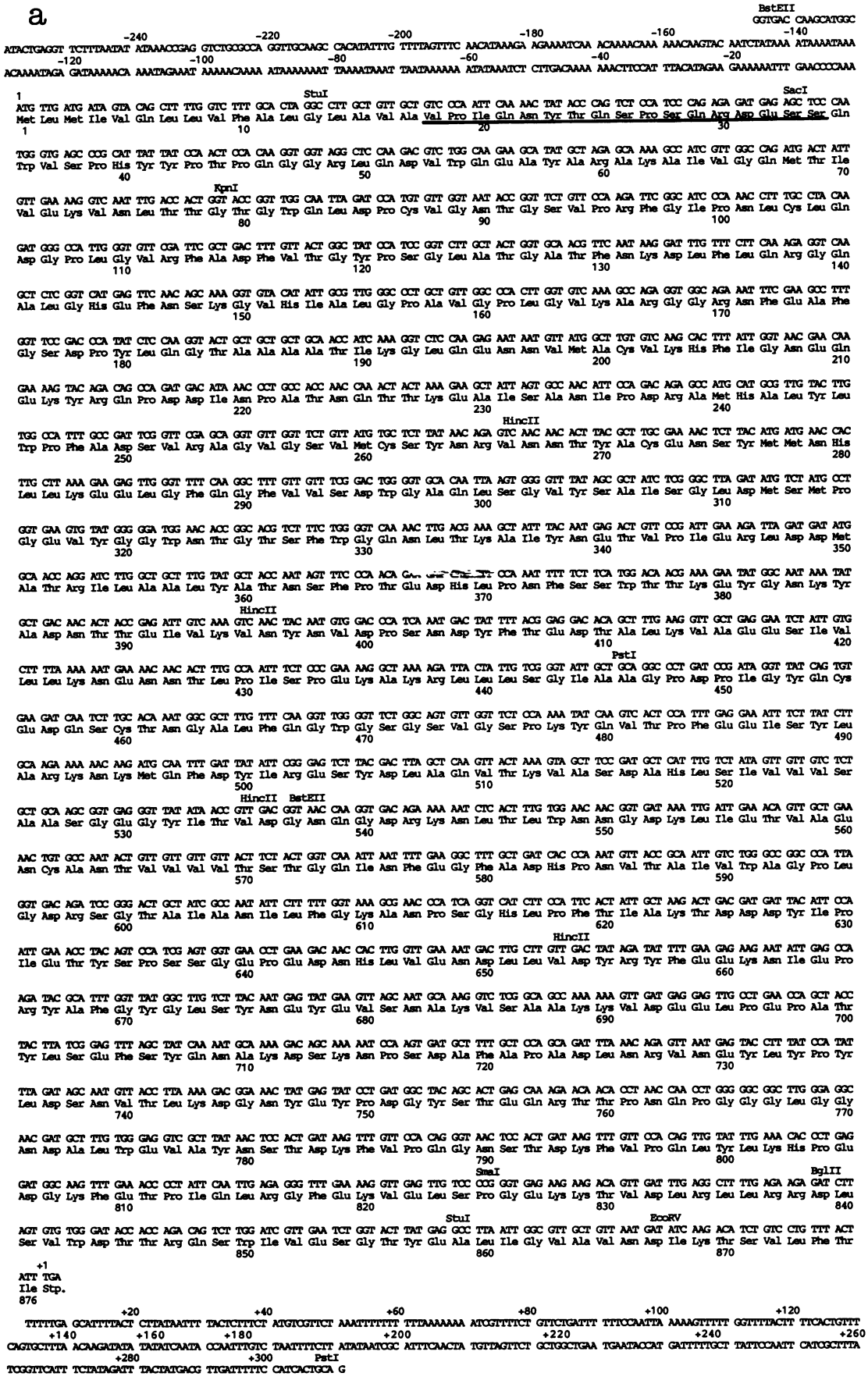
Construction of cellobiose-fermenting *S. cerevisiae*. The structural gene for β-glucosidase is present in *S. cerevisiae* (3), but it is very poorly expressed (10). Furthermore, this organism does not have a permease for cellobiose (10). These properties could explain why *S. cerevisiae* is unable to ferment cellobiose. The simplest strategy for constructing cellobiose-fermenting *S. cerevisiae* is to clone and express a gene for a secretable cellobiase. Several of the β-glucosidase genes thus far cloned failed to confer cellobiose-fermenting ability on *S. cerevisiae*, since their gene products were either not secreted or unable to hydrolyze cellobiose (7, 9, 10).

We isolated the gene for extracellular cellobiase and constructed, by introducing the gene, an *S. cerevisiae* strain capable of fermenting cellobiose to ethanol. However, ethanol production by this transformant was very low, mainly because the recipient used was a laboratory strain. Further studies should be done on introduction of the gene into industrial strains.

Polymorphic β-glucosidases with different substrate specificities. It was shown by Southern blot analysis that the two genes for extracellular β-glucosidases cloned in this study exist in the genome of *Saccharomycopsis fibuligera*. One, as described above, encodes the β-glucosidase which is capable of hydrolyzing cellobiose, and another apparently encodes aryl or alkyl β-glucosidase. To our knowledge, this is the first report showing that polymorphic β-glucosidases with highly similar sequences possess quite different substrate specificities.

Figure 7 shows the best-fit alignment of three amino acid sequences from *Saccharomycopsis fibuligera* (this work) and *C. pelliculosa*. These proteins share several homologous peptides (boxed in Fig. 7) which are likely to be essential for enzymatic activities. Although BGLI and BGLII are highly homologous with each other, they also contain several nonhomologous peptides (underlined in Fig. 7), some of which may function in specific enzymatic actions so as to determine substrate specificities.

On the basis of studies of the chemical modification of the fungal β-glucosidase (1), the aspartic acid residue in the peptide NH_2-Val-Met-Ser-Asp-Trp-Ala-Ala-His-His-Ala-Gly-Val-Ser-Gly-Ala-Leu is proposed to be essential for enzymatic activities. We found similar peptides in both BGLI and BGLII (asterisks in Fig. 7). To elucidate further the structure-function relationship of β-glucosidase, we should analyze mutant enzymes whose amino acid residues are substituted by protein engineering.


a
BstEII
GGTGAC CAAGCATGGC
-240 -220 -200 -180 -160 -140
ATACTGAGGT TCTTTAATAT ATAAACCGAG GTCTGCGCCA GGTTGCAAGC CACATATTTG TTTTAGTTTC AACATAAAGA AGAAAATCAA ACAAAACAAA AAACAAGTAC AATCTATAAA ATAAAATAAA
-120 -100 -80 -60 -40 -20
ACAAAATAGA GATAAAAACA AAATAGAAAT AAAAACAAAA ATAAAAAAAT TAAAATAAAT TAATAAAAAA ATATAAATCT CTTGACAAAA AAACTTCCAT TTACATAGAA GAAAAAATTT GAACCCCAAA
StuI
SacI
1
ATG TTG ATG ATA GTA CAG CTT TTG GTC TTT GCA CTA GGC CTT GCT GTT GCT GTC CCA ATT CAA AAC TAT ACC CAG TCT CCA TCC CAG AGA GAT GAG AGC TCC CAA
Met Leu Met Ile Val Gln Leu Leu Val Phe Ala Leu Gly Leu Ala Val Ala Val Pro Ile Gln Asn Tyr Thr Gln Ser Pro Ser Gln Arg Asp Glu Ser Ser Gln
1 10 20 30
TGG GTG AGC CCG CAT TAT TAT CCA ACT CCA CAA GGT GGT AGG CTC CAA GAC GTC TGG CAA GAA GCA TAT GCT AGA GCA AAA GCC ATC GTT GGC CAG ATG ACT ATT
Trp Val Ser Pro His Tyr Tyr Pro Thr Pro Gln Gly Gly Arg Leu Gln Asp Val Trp Gln Glu Ala Tyr Ala Arg Ala Lys Ala Ile Val Gly Gln Met Thr Ile
40 50 60 70
KpnI
GTT GAA AAG GTC AAT TTG ACC ACT GGT ACC GGT TGG CAA TTA GAT CCA TGT GTT GGT AAT ACC GGT TCT GTT CCA AGA TTC GGC ATC CCA AAC CTT TGC CTA CAA
Val Glu Lys Val Asn Leu Thr Thr Gly Thr Gly Trp Gln Leu Asp Pro Cys Val Gly Asn Thr Gly Ser Val Pro Arg Phe Gly Ile Pro Asn Leu Cys Leu Gln
80 90 100
GAT GGG CCA TTG GGT GTT CGA TTC GCT GAC TTT GTT ACT GGC TAT CCA TCC GGT CTT GCT ACT GGT GCA ACG TTC AAT AAG GAT TTG TTT CTT CAA AGA GGT CAA
Asp Gly Pro Leu Gly Val Arg Phe Ala Asp Phe Val Thr Gly Tyr Pro Ser Gly Leu Ala Thr Gly Ala Thr Phe Asn Lys Asp Leu Phe Leu Gln Arg Gly Gln
110 120 130 140
GCT CTC GGT CAT GAG TTC AAC AGC AAA GGT GTA CAT ATT GGG TTG GGC CCT GCT GTT GGC CCA CTT GGT GTC AAA GCC AGA GGT GGC AGA AAT TTC GAA GCC TTT
Ala Leu Gly His Glu Phe Asn Ser Lys Gly Val His Ile Ala Leu Gly Pro Ala Val Gly Pro Leu Gly Val Lys Ala Arg Gly Gly Arg Asn Phe Glu Ala Phe
150 160 170
GGT TCC GAC CCA TAT CTC CAA GGT ACT GCT GCT GCT GCA ACC ATC AAA GGT CTC CAA GAG AAT AAT GTT ATG GCT TGT GTC AAG CAC TTT ATT GGT AAC GAA CAA
Gly Ser Asp Pro Tyr Leu Gln Gly Thr Ala Ala Ala Ala Thr Ile Lys Gly Leu Gln Glu Asn Asn Val Met Ala Cys Val Lys His Phe Ile Gly Asn Glu Gln
180 190 200 210
GAA AAG TAC AGA CAG CCA GAT GAC ATA AAC CCT GCC ACC AAC CAA ACT ACT AAA GAA GCT ATT AGT GCC AAC ATT CCA GAC AGA GCC ATG CAT GCG TTG TAC TTG
Glu Lys Tyr Arg Gln Pro Asp Asp Ile Asn Pro Ala Thr Asn Gln Thr Thr Lys Glu Ala Ile Ser Ala Asn Ile Pro Asp Arg Ala Met His Ala Leu Tyr Leu
220 230 240
HincII
TGG CCA TTT GCC GAT TCG GTT CGA GCA GGT GTT GGT TCT GTT ATG TGC TCT TAT AAC AGA GTC AAC AAC ACT TAC GCT TGC GAA AAC TCT TAC ATG ATG AAC CAC
Trp Pro Phe Ala Asp Ser Val Arg Ala Gly Val Gly Ser Val Met Cys Ser Tyr Asn Arg Val Asn Asn Thr Tyr Ala Cys Glu Asn Ser Tyr Met Met Asn His
250 260 270 280
TTG CTT AAA GAA GAG TTG GGT TTT CAA GGC TTT GTT GTT TCG GAC TGG GGT GCA CAA TTA AGT GGG GTT TAT AGC GCT ATC TCG GGC TTA GAT ATG TCT ATG CCT
Leu Leu Lys Glu Glu Leu Gly Phe Gln Gly Phe Val Val Ser Asp Trp Gly Ala Gln Leu Ser Gly Val Tyr Ser Ala Ile Ser Gly Leu Asp Met Ser Met Pro
290 300 310
GGT GAA GTG TAT GGG GGA TGG AAC ACC GGC ACG TCT TTC TGG GGT CAA AAC TTG ACG AAA GCT ATT TAC AAT GAG ACT GTT CCG ATT GAA AGA TTA GAT GAT ATG
Gly Glu Val Tyr Gly Gly Trp Asn Thr Gly Thr Ser Phe Trp Gly Gln Asn Leu Thr Lys Ala Ile Tyr Asn Glu Thr Val Pro Ile Glu Arg Leu Asp Asp Met
320 330 340 350
GCA ACC AGG ATC TTG GCT GCT TTG TAT GCT ACC AAT AGT TTC CCA ACA GA[illegible] CCA AAT TTT TCT TCA TGG ACA ACG AAA GAA TAT GGC AAT AAA TAT
Ala Thr Arg Ile Leu Ala Ala Leu Tyr Ala Thr Asn Ser Phe Pro Thr Glu Asp His Leu Pro Asn Phe Ser Ser Trp Thr Thr Lys Glu Tyr Gly Asn Lys Tyr
360 370 380
HincII
GCT GAC AAC ACT ACC GAG ATT GTC AAA GTC AAC TAC AAT GTG GAC CCA TCA AAT GAC TAT TTT ACG GAG GAC ACA GCT TTG AAG GTT GCT GAG GAA TCT ATT GTG
Ala Asp Asn Thr Thr Glu Ile Val Lys Val Asn Tyr Asn Val Asp Pro Ser Asn Asp Tyr Phe Thr Glu Asp Thr Ala Leu Lys Val Ala Glu Glu Ser Ile Val
390 400 410 420
PstI
CTT TTA AAA AAT GAA AAC AAC ACT TTG CCA ATT TCT CCC GAA AAG GCT AAA AGA TTA CTA TTG TCG GGT ATT GCT GCA GGC CCT GAT CCG ATA GGT TAT CAG TGT
Leu Leu Lys Asn Glu Asn Asn Thr Leu Pro Ile Ser Pro Glu Lys Ala Lys Arg Leu Leu Leu Ser Gly Ile Ala Ala Gly Pro Asp Pro Ile Gly Tyr Gln Cys
430 440 450
GAA GAT CAA TCT TGC ACA AAT GGC GCT TTG TTT CAA GGT TGG GGT TCT GGC AGT GTT GGT TCT CCA AAA TAT CAA GTC ACT CCA TTT GAG GAA ATT TCT TAT CTT
Glu Asp Gln Ser Cys Thr Asn Gly Ala Leu Phe Gln Gly Trp Gly Ser Gly Ser Val Gly Ser Pro Lys Tyr Gln Val Thr Pro Phe Glu Glu Ile Ser Tyr Leu
460 470 480 490
GCA AGA AAA AAC AAG ATG CAA TTT GAT TAT ATT CGG GAG TCT TAC GAC TTA GCT CAA GTT ACT AAA GTA GCT TCC GAT GCT CAT TTG TCT ATA GTT GTT GTC TCT
Ala Arg Lys Asn Lys Met Gln Phe Asp Tyr Ile Arg Glu Ser Tyr Asp Leu Ala Gln Val Thr Lys Val Ala Ser Asp Ala His Leu Ser Ile Val Val Val Ser
500 510 520
HincII BstEII
GCT GCA AGC GGT GAG GGT TAT ATA ACC GTT GAC GGT AAC CAA GGT GAC AGA AAA AAT CTC ACT TTG TGG AAC AAC GGT GAT AAA TTG ATT GAA ACA GTT GCT GAA
Ala Ala Ser Gly Glu Gly Tyr Ile Thr Val Asp Gly Asn Gln Gly Asp Arg Lys Asn Leu Thr Leu Trp Asn Asn Gly Asp Lys Leu Ile Glu Thr Val Ala Glu
530 540 550 560
AAC TGT GCC AAT ACT GTT GTT GTT GTT ACT TCT ACT GGT CAA ATT AAT TTT GAA GGC TTT GCT GAT CAC CCA AAT GTT ACC GCA ATT GTC TGG GCC GGC CCA TTA
Asn Cys Ala Asn Thr Val Val Val Val Thr Ser Thr Gly Gln Ile Asn Phe Glu Gly Phe Ala Asp His Pro Asn Val Thr Ala Ile Val Trp Ala Gly Pro Leu
570 580 590
GGT GAC AGA TCC GGG ACT GCT ATC GCC AAT ATT CTT TTT GGT AAA GCG AAC CCA TCA GGT CAT CTT CCA TTC ACT ATT GCT AAG ACT GAC GAT GAT TAC ATT CCA
Gly Asp Arg Ser Gly Thr Ala Ile Ala Asn Ile Leu Phe Gly Lys Ala Asn Pro Ser Gly His Leu Pro Phe Thr Ile Ala Lys Thr Asp Asp Asp Tyr Ile Pro
600 610 620 630
HincII
ATT GAA ACC TAC AGT CCA TCG AGT GGT GAA CCT GAA GAC AAC CAC TTG GTT GAA AAT GAC TTG CTT GTT GAC TAT AGA TAT TTT GAA GAG AAG AAT ATT GAG CCA
Ile Glu Thr Tyr Ser Pro Ser Ser Gly Glu Pro Glu Asp Asn His Leu Val Glu Asn Asp Leu Leu Val Asp Tyr Arg Tyr Phe Glu Glu Lys Asn Ile Glu Pro
640 650 660
AGA TAC GCA TTT GGT TAT GGC TTG TCT TAC AAT GAG TAT GAA GTT AGC AAT GCA AAG GTC TCG GCA GCC AAA AAA GTT GAT GAG GAG TTG CCT GAA CCA GCT ACC
Arg Tyr Ala Phe Gly Tyr Gly Leu Ser Tyr Asn Glu Tyr Glu Val Ser Asn Ala Lys Val Ser Ala Ala Lys Lys Val Asp Glu Glu Leu Pro Glu Pro Ala Thr
670 680 690 700
TAC TTA TCG GAG TTT AGC TAT CAA AAT GCA AAA GAC AGC AAA AAT CCA AGT GAT GCT TTT GCT CCA GCA GAT TTA AAC AGA GTT AAT GAG TAC CTT TAT CCA TAT
Tyr Leu Ser Glu Phe Ser Tyr Gln Asn Ala Lys Asp Ser Lys Asn Pro Ser Asp Ala Phe Ala Pro Ala Asp Leu Asn Arg Val Asn Glu Tyr Leu Tyr Pro Tyr
710 720 730
TTA GAT AGC AAT GTT ACC TTA AAA GAC GGA AAC TAT GAG TAT CCT GAT GGC TAC AGC ACT GAG CAA AGA ACA ACA CCT AAC CAA CCT GGG GGC GGC TTG GGA GGC
Leu Asp Ser Asn Val Thr Leu Lys Asp Gly Asn Tyr Glu Tyr Pro Asp Gly Tyr Ser Thr Glu Gln Arg Thr Thr Pro Asn Gln Pro Gly Gly Gly Leu Gly Gly
740 750 760 770
AAC GAT GCT TTG TGG GAG GTC GCT TAT AAC TCC ACT GAT AAG TTT GTT CCA CAG GGT AAC TCC ACT GAT AAG TTT GTT CCA CAG TTG TAT TTG AAA CAC CCT GAG
Asn Asp Ala Leu Trp Glu Val Ala Tyr Asn Ser Thr Asp Lys Phe Val Pro Gln Gly Asn Ser Thr Asp Lys Phe Val Pro Gln Leu Tyr Leu Lys His Pro Glu
780 790 800
SmaI BglII
GAT GGC AAG TTT GAA ACC CCT ATT CAA TTG AGA GGG TTT GAA AAG GTT GAG TTG TCC CCG GGT GAG AAG AAG ACA GTT GAT TTG AGG CTT TTG AGA AGA GAT CTT
Asp Gly Lys Phe Glu Thr Pro Ile Gln Leu Arg Gly Phe Glu Lys Val Glu Leu Ser Pro Gly Glu Lys Lys Thr Val Asp Leu Arg Leu Leu Arg Arg Asp Leu
810 820 830 840
StuI EcoRV
AGT GTG TGG GAT ACC ACC AGA CAG TCT TGG ATC GTT GAA TCT GGT ACT TAT GAG GCC TTA ATT GGC GTT GCT GTT AAT GAT ATC AAG ACA TCT GTC CTG TTT ACT
Ser Val Trp Asp Thr Thr Arg Gln Ser Trp Ile Val Glu Ser Gly Thr Tyr Glu Ala Leu Ile Gly Val Ala Val Asn Asp Ile Lys Thr Ser Val Leu Phe Thr
850 860 870
+1
ATT TGA
Ile Stp.
876
+20 +40 +60 +80 +100 +120
TTTTTGA GCATTTTACT CTTATAATTT TACTCTTTCT ATGTCGTTCT AAATTTTTTT TTTAAAAAAA ATCGTTTTCT GTTCTGATTT TTTCCAATTA AAAAGTTTTT GGTTTTACTT TTCACTGTTT
+140 +160 +180 +200 +220 +240 +260
CAGTGCTTTA ACAAGATATA TATATCAATA CCAATTTGTC TAATTTTCTT ATATAATGGC ATTTCAACTA TGTTAGTTCT GCTGGCTGAA TGAATACCAT GATTTTTGCT TATTCCAATT CATGCTTTA
+280 +300 PstI
TCGGTTCATT TCTATAGATT TACTATGACG TTGATTTTTC CATCACTGCA G

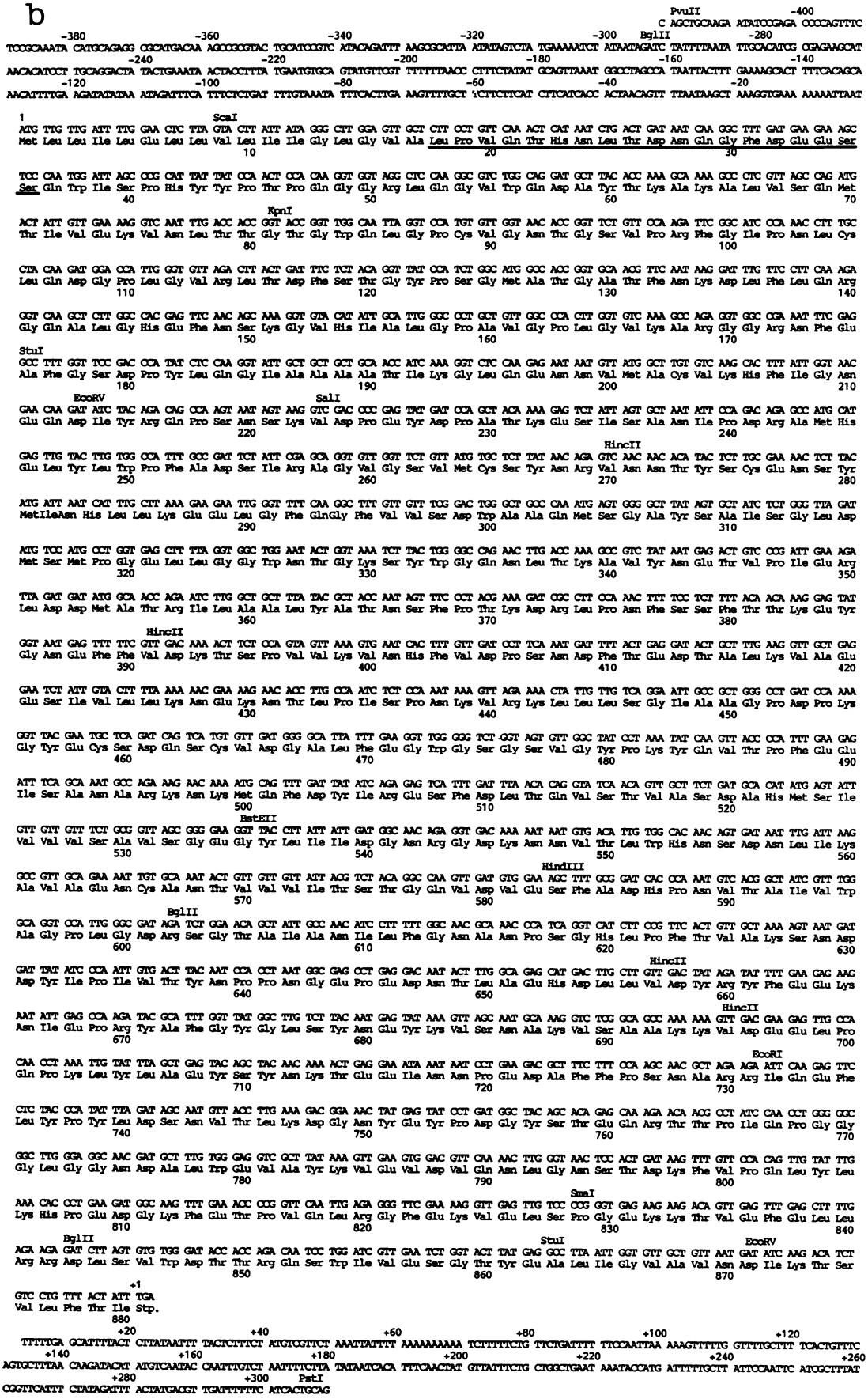


FIG. 6. Nucleotide and deduced protein sequences of *BGL1* (a) and *BGL2* (b). Numbers above a sequence indicate the number of the nucleotide, in each direction, from the A in the translation-initiating ATG or the T in the stop codon TGA. Restriction sites are indicated above the nucleotide sequence. Numbers below the protein sequence denote the amino acid number, starting with 1 at the initiator methionine. Underlined is the amino-terminal peptide of the mature or secreted protein determined by direct sequencing.

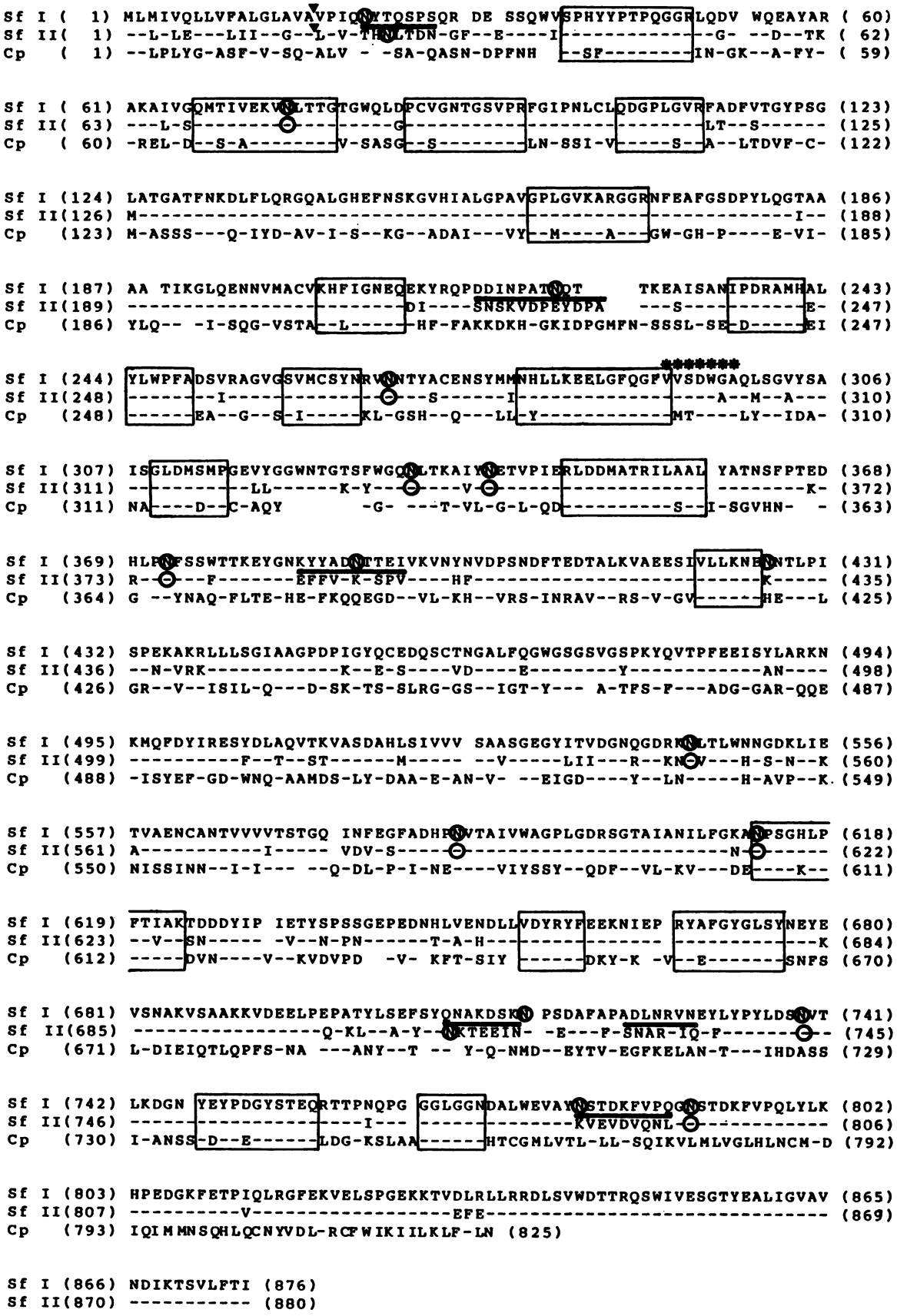


FIG. 7. Sequence homologies among yeast β-glucosidases. Amino acid sequences deduced from the following β-glucosidase genes were aligned to maximize the homologies: *Saccharomycopsis fibuligera BGL1* (SfI); *Saccharomycopsis fibuligera BGL2* (SfII); and *Candida pelliculosa* β-glucosidase gene (Cp) (7). Symbols: -, residue is identical to that shown for *Saccharomycopsis fibuligera* BGLI; □, highly conserved region; ○, potential N-glycosylation site; ▼, possible signal sequence cleavage site; *, conserved peptide including the active-site aspartic acid residue. Underlined are the nonhomologous peptides between BGLI and BGLII. Numbers in parentheses indicate amino acid numbers.

ACKNOWLEDGMENTS

We thank K. Yamaguchi (Kyowa Hakko Co.) for amino acid analyses of β-glucosidases.

This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

1. **Bause, E., and G. Legler.** 1974. Isolation and amino acid sequence of a hexadecapeptide from the active site of β-glucosidase A3 from *Aspergillus wentii*. Hoppe-Seyler's Z. Physiol. Chem. **355**:438–442.
2. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
3. **Duerksen, J. D., and H. Halvorson.** 1958. Purification and properties of an inducible β-glucosidase of yeast. J. Biol. Chem. **233**:1113–1120.
4. **Itoh, T., I. Ohtsuki, I. Yamashita, and S. Fukui.** 1987. Nucleotide sequence of the glucoamylase gene *GLU1* in the yeast *Saccharomycopsis fibuligera*. J. Bacteriol. **169**:4171–4176.
5. **Itoh, T., I. Yamashita, and S. Fukui.** 1987. Nucleotide sequence of the α-amylase gene (*ALP1*) in the yeast *Saccharomycopsis fibuligera*. FEBS Lett. **219**:339–342.
6. **Kaplan, J. G., and W. Tacreiter.** 1966. The β-glucosidase of the yeast cell surface. J. Gen. Physiol. **50**:9–24.
7. **Kohchi, C., and A. Toh-e.** 1985. Nucleotide sequence of *Candida pelliculosa* β-glucosidase gene. Nucleic Acids Res. **13**: 6273–6282.
8. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
9. **Penttilae, M. E., K. M. H. Nevalainen, A. Raynal, and J. K. C. Knowles.** 1984. Cloning of *Aspergillus niger* genes in yeast. Expression of the gene coding *Aspergillus* β-glucosidase. Mol. Gen. Genet. **194**:494–499.
10. **Raynal, A., and M. Guerineau.** 1984. Cloning and expression of the structural gene for β-glucosidase of *Kluyveromyces fragilis* in *Escherichia coli* and *Saccharomyces cerevisiae*. Mol. Gen. Genet. **195**:108–115.
11. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
12. **Sherman, F., G. R. Fink, and C. W. Lawrence.** 1974. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Woodward, J., and A. Wiseman.** 1982. Fungal and other β-D-glucosidases—their properties and applications. Enzyme Microb. Technol. **4**:73–79.
14. **Yamashita, I., and S. Fukui.** 1983. Molecular cloning of a glucoamylase-producing gene in the yeast *Saccharomyces*. Agric. Biol. Chem. **47**:2689–2692.
15. **Yamashita, I., T. Itoh, and S. Fukui.** 1985. Cloning and expression of the *Saccharomycopsis fibuligera* glucoamylase gene in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. **23**: 130–133.
16. **Yamashita, I., T. Itoh, and S. Fukui.** 1985. Cloning and expression of the *Saccharomycopsis fibuligera* α-amylase gene in *Saccharomyces cerevisiae*. Agric. Biol. Chem. **49**:3089–3091.
17. **Yamashita, I., T. Maemura, T. Hatano, and S. Fukui.** 1985. Polymorphic extracellular glucoamylase genes and their evolutionary origin in the yeast *Saccharomyces diastaticus*. J. Bacteriol. **161**:574–582.
18. **Yamashita, I., M. Nakamura, and S. Fukui.** 1987. Gene fusion is a possible mechanism underlying the evolution of *STA1*. J. Bacteriol. **169**:2142–2149.
19. **Yamashita, I., K. Suzuki, and S. Fukui.** 1985. Nucleotide sequence of the extracellular glucoamylase gene *STA1* in the yeast *Saccharomyces diastaticus*. J. Bacteriol. **161**:567–573.