Applied and Environmental Microbiology, July 1998, p. 2748–2754 Vol. 64, No. 7
0099-2240/98/$04.00+0
Copyright © 1998, American Society for Microbiology. All Rights Reserved.

Gene Cloning and Characterization of a Novel Cellulose-Binding β-Glucosidase from *Phanerochaete chrysosporium*

BIN LI and V. RENGANATHAN*

Department of Biochemistry and Molecular Biology, Oregon Graduate Institute of Science and Technology, Portland, Oregon 97291-1000

Received 23 December 1997/Accepted 18 April 1998

Analysis of a 2.4-kb cDNA of the cellulose-binding extracellular β-glucosidase (CBGL) from *Phanerochaete chrysosporium* suggested that CBGL is organized into two domains, an N-terminal cellulose-binding domain and a C-terminal catalytic domain. Genomic sequence analysis suggested that *cbgl* is encoded by 30 exons. Southern analysis of DNA from homokaryotic cultures indicated that CBGL is encoded by two alleles, *cbgl-1* and *cbgl-2*, of a single gene.

Cellulose-degrading cultures of the white rot basidiomycete *Phanerochaete chrysosporium* apparently produce three different β-glucosidases—extracellular, intracellular, and cell wall bound—depending on the carbon source (9, 26). Deshpande et al. (9) reported that cellulose induces intracellular and cell wall-bound enzymes and purified five isozymes of extracellular β-glucosidases from cellulose-degrading cultures of *P. chrysosporium*. Molecular weights of these glucosidases ranged from 165,000 to 182,000. Smith and Gold (26) partially purified an extracellular β-glucosidase from *P. chrysosporium* OGC101 and charaterized it as a monomer with a molecular weight of 90,000. Recently, we purified and characterized a cellulose-binding extracellular β-glucosidase (CBGL) with a molecular mass of 114,000 from cellulose-supplemented cultures of *P. chrysosporium* OGC101. When CBGL was treated with papain, its molecular weight decreased to 95,000; it lost the ability to bind to cellulose, but its catalytic activity was unchanged. This suggested that CBGL is organized into two domains—a cellulose-binding domain (CBD) and a catalytic domain (20). The glucosidase isolated previously by Smith and Gold (26) from this strain might be the non-cellulose-binding form. The kinetic properties of the cellulose-binding and nonbinding forms were similar, indicating that the CBD was not involved in catalysis. Here cloning and characterization of a cDNA clone and a genomic clone of CBGL are reported. Sequence analysis confirmed our prediction that this β-glucosidase consists of a catalytic domain and a CBD.

Organism. *P. chrysosporium* OGC101 (a derivative of BKM-F-1767) was obtained from Michael H. Gold (Oregon Graduate Institute) (1). *Escherichia coli* XL1-Blue MRF′ and SOLR were obtained from Stratagene (La Jolla, Calif.).

Nucleotides. Oligonucleotides were prepared by the Oregon Regional Primate Research Center (Beaverton, Oreg.). The plasmid isolation kit was obtained from Qiagen, Inc. (Chatsworth, Calif.).

Isolation of a cDNA clone of *cbgl*. The cDNA λZAP expression library, prepared as described previously (18), was screened with an anti-CBGL antibody and a secondary antibody labeled with alkaline phosphatase. The pBluescript SK(−) plasmid containing a putative β-glucosidase cDNA insert was rescued by in vivo excision with a helper phage. The plasmid was purified with a commercial plasmid isolation kit (Qiagen, Inc.). The cDNA was sequenced by the dideoxy method with the primer walking strategy (25, 27).

Isolation of a genomic clone of *cbgl*. A λEMBL3 genomic library of *P. chrysosporium* OGC101 was screened at high stringency (4.8× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]–48% formamide at 50°C) with a 550-bp *Apa*I fragment from the 3′ end of the *cbgl* cDNA clone. Based on the restriction mapping of the genomic clones, four overlapping restriction fragments (3.6-kb *Sac*I, 1.7-kb *Sac*I, 4.5-kb *Sal*I, and 1.2-kb *Sal*I) covering the entire region of *cbgl* were subcloned into pBluescript SK (Stratagene) and sequenced by the primer walking method (27). Sequencing was performed with an automatic sequencer (model 377; Applied Biosystems) and Ampli*Taq* DNA polymerase, FS.

Isolation of a full-length cDNA clone of *cbgl*. The cDNA library of *P. chrysosporium* was probed with a restriction fragment (66 to 324 bp) obtained by digesting *cbgl-2* with *Msc*I and *Nde*I. Hybridization was performed at high stringency (4.8× SSC, 48% formamide, 50°C). Positive clones were purified by further screening. The pBluescript II SK plasmid containing the putative *cbgl* cDNA insert was rescued by in vivo excision with a helper phage. The plasmid was purified with a plasmid midi kit (Qiagen, Inc.).

Isolation and analysis of homokaryons. Single homokaryotic basidiospores were isolated as described previously (1, 12). DNAs from homokaryotic cultures were isolated by standard procedures and restriction digested with *Sal*I, size fractionated in a 0.7% agarose gel, blotted onto a Magnagraph nylon transfer membrane (Micron Separations, Westboro, Mass.), and probed with a ^{32}P-labeled 1.4-kb *Sac*I fragment of *cbgl* (nucleotides 1446 to 2866).

Northern (RNA) blot analysis. Total RNA was isolated from 11-day-old mycelia of *P. chrysosporium* cultured with 1% cotton linters, cellobiose, or glucose as the carbon source (2, 8). RNA was electrophoresed in 1.5% agarose gel containing 2.2 M formaldehyde, transferred to Magnagraph nylon membranes (Micron Separations), and probed with cDNA for CBGL at 42°C as described previously (6).

Southern blot analysis of *cbgl*. DNA from *P. chrysosporium* was restriction digested and electrophoresed with a 0.7% agarose gel. The DNA was transferred to Magnagraph nylon mem-

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Oregon Graduate Institute of Science and Technology, P.O. Box 91000, Portland, OR 97291-1000. Phone: (503) 690-1134. Fax: (503) 690-1464. E-mail: vreng@bmb.ogi.edu.

```
TCGAAGTTGCCGCTGTTTGTCCGCGCCTTATCCGCTCGACGCGGCTGCTGTTGGGCAACGACGTGGCCAAGACCATTTCCGATCTACATATCGACTTCTTCAAACATGTTCGAGAAGTAT 120
AAGTGCTTTGACCATGCGTACGCTGTCTACGCAAGGCAGCCCCCTCTCCCTACGCTGACAAAATGGGACTGACACTGGTCGTCTTATTACATCTTGCTCTAGGGTTGCTGACTGGGGTCC 240
                                                               M  G  L  T  L  V  V  L  L  H  L  A  L  G  L  L  T  G  V    19
AAGCTCAGTCCGGCCTGTATCAGCAATGTGGCGGTATCGGATGgttagctggcttgtctctcctgcgtccggagtggatgctcatatgttcctcgcatgcagGACTGGTGCAACTACTTG 360
Q  A  Q  S  G  L  Y  Q  Q  C  G  G  I  G  W                                                             T  G  A  T  T  C 40
                                                                                            A
CGTCAGCGGCGCGACCTGCACAGTGCTGAACCCATgtacgtcgccacacagtacatgtcacagcagccgagctcatgacacgtagACTACTCGCAATGTCTCCCTGGCGCCGCCACCACG 480
  V  S  G  A  T  C  T  V  L  N  P                                                    Y  Y  S  Q  C  L  P  G  A  A  T  T  63
                          C                            A                     C                 C              T
AGTGTTTCGTCTAGTCATTCGTCAAGTAGCAGCGTCAGCAGCCACTCCAGCAGTGCAAGCAGCAGCAGCATAAGTTCTACTAGCACCAGCCCCCCGGCGCCCTCCCAGACCGTGGCGAAC 600
 S  V  S  S  S  H  S  S  S  S  S  V  S  S  H  S  S  S  A  S  S  S  S  I  S  S  T  S  T  S  P  P  A  P  S  Q  T  V  A  N  103
                                                      (E)
     G           T  T  C                                                                              G   T
GTGTCTCCAGAATGGGCCGCCGCTTACGTCAAGgttggaacgacgtcttcagatgtatatgttgatgtcactgaaacacgcgcagGCACAAGCTGCTGTAGCAAAGCTATCGGTGACTGA 720
 V  S  P  E  W  A  A  A  Y  V  K                                                      A  Q  A  A  V  A  K  L  S  V  T  D 126
                                                                                                                     T
TATGGTCAACCTCGCAACAGgtgagatgactcgtgattaatactctatcttctaacaacttctggcagGTGTGCAGTGGCAGAAAGGTCCATGTGTCGGCAACACCCCTGCTATCTCGTC 840
  M  V  N  L  A  T                                                  G  V  Q  W  Q  K  G  P  C  V  G  N  T  P  A  I  S  S 150
TATACCCGGTTTTACTGGACTTTGCCTGCAAGgtatgcgttcaacgtccctacgatgtgtgcaggtcgacttattccatggtccacagACAGTCCCGTTGGTGTCCGCTACGCTGATGGA 960
  I  P  G  F  T  G  L  C  L  Q                                                          D  S  P  V  G  V  R  Y  A  D  G  171
        A
ACGTCTGTCTTCCCACCgtacgtgctgagcgatagctgtgaatctacttccctggctgaaacagttacctacagTGAAATCAACGTTGCTGCTACATGGAACCGTACGCTTATGCGCCAA 1080
 T  S  V  F  P  P                                                           E  I  N  V  A  A  T  W  N  R  T  L  M  R  Q  192
  T                       C  A                                                                 G
CGCGGAGCTGCTATGGGAGCGGAATTTAAGGGGAAAGGCGTGCACGTCGCTCTCGGTCCCATGATGAACCTCATGCGTGTCCCCGCCGCTGGCCGAAACTGGGAAGGgtgagtttctttg 1200
 R  G  A  A  M  G  A  E  F  K  G  K  G  V  H  V  A  L  G  P  M  M  N  L  M  R  V  P  A  A  G  R  N  W  E  G              228
                                                                            G              A             G
gagtgtttgacacgcttgtcgcggcagtacttaatttctccgtcttggattgtagTGGCGGTGGTGATCCTTTCCTCTCTGGCGAAGTGGCGTTCGAGACCATCACTGGCATTCAGTCCT 1320
                                                         G  G  G  D  P  F  L  S  G  E  V  A  F  E  T  I  T  G  I  Q  S  249
                                                                                                        (S)
                   G                                                                                      A
CCGGTGCCCAGGCTTGTGCCAAGCATTTCATCAACAAgtgcgtgtcttcatctccttgcagttcgcgcttactgacgtatttatgcagCGAGCAAGAACACTTTAGGGACTCCAGTTCAT 1440
S  G  A  Q  A  C  A  K  H  F  I  N  N                                                     E  Q  E  H  F  R  D  S  S  S   272
CTAACGTCGACGATAGgtgagtttaccgtgttgcgacagtcgggaaagtagctgacctgctctcatagGACGgtgagataatcgtagtttatatgtgacatgtttctcattcaagtgtt 1560
S  N  V  D  D  R                                                       T                                                  279
agGAGCATGAACTCTACGGACACCCTgtaatggttcctctatcattcgccgtcaccgactctgacgcatcaatagTTCCTTCGTAGCGTCCAGGCAAACGTAGCTTCCGTAATGTGCAGC 1680
   E  H  E  L  Y  G  H  P                                                   F  L  R  S  V  Q  A  N  V  A  S  V  M  C  S  302
TACAgtaagttgtgatctcagcctcgtgcagacgcaaactaaatcgtctgatgtttactacagATCAAAgtacgaatgcatcacgaatgattccagcagacgctaatctatatcctactc 1800
 Y                                                             N  Q                                                        305
            T
agTCAACGGAACCTTCTCTTGCGAAAATGAGAAGACCTTGTCAGGACTCCTCAAGGGAGAGTACGGCTTCCAAGGCTgtgagtgcctcttttacggggcaggcaatcaaattaactctc 1920
  I  N  G  T  F  S  C  E  N  E  K  T  L  S  G  L  L  K  G  E  Y  G  F  Q  G                                             330
agtatcgtagACGTCATGTCTGgtaggtttacagtgtgccatgtcaagctcgcggctgagctaagctgtcagACTGGTGGGCCACTCACTCTGGTGCACCCGCTGTGAACGCCGGCCTT 2040
          Y  V  M  S                                                     D  W  W  A  T  H  S  G  A  P  A  V  N  A  G  L  350
                                                                T
GACgtatgtctccgcgccttcccatggccgctactgctaactgcttttagATGACCATGCCGGGCGACGAGACACTCAGCTCTGGAACGACgtacgcagctcctgccgtaagctgtcttg 2160
 D                                                 M  T  M  P  G  D  E  T  L  S  S  G  T  T                              365
                                                                                      C
cggcattctgacgtgcgcattgccagCTACTTCGGACAGAACCTCGTCAACGCTGTGAACAGCGGCCAAGTCTCGCAGGCCAGGGTTAAGgtacgcagaatcgaccatatatctactaat 2280
                            Y  F  G  Q  N  L  V  N  A  V  N  S  G  Q  V  S  Q  A  R  V  K                               386
                                                A
aggacttgaatgttgtcttgcacagGACATGGCGACTCGCATCCTTGCGGCGTGGTACCTCCTTGGCCAAGACCAAAACTTCCCCGCGGTCAACTTCAACTCGTGGAACTCTGGTCAGGG 2400
                          D  M  A  T  R  I  L  A  A  W  Y  L  L  G  Q  D  Q  N  F  P  A  V  N  F  N  S  W  N  S  G  Q  G 418
                           C                                                                                        G
CCAGCATGTGAACGTCTCCGGCAACCATGCGAGgtactgtaaaccgggtatccactactcggcgcgcaccgctgacacgtctgtagCCTCATCCGCACCATCGGCGCGGCCTCGCAAATC 2520
  Q  H  V  N  V  S  G  N  H  A  S                                                        L  I  R  T  I  G  A  A  S  Q  I  440
             C    G                G                          C
CTGCTCAAGAACGTGAACAGTGCGCTTCCGCTCAAAAAGCCCAAGACTATTGGCATCATTGGGAACGGTGCTGGATCAAACCCTAACGGTCCCAACGCgtacggacacaatccagcttta 2640
 L  L  K  N  V  N  S  A  L  P  L  K  K  P  K  T  I  G  I  I  G  N  G  A  G  S  N  P  N  G  P  N  A                      473
            (A)   (G)
                                          T
cgttcccaaagcttgctgaagtttcgtttacgccagTTTCTCCGACCGCGCCGGAGACGTCGGTGTCCTCGCGCTTGGGTGGGGCAGTGGgtaagtcggcgcccgacgttggggttgcta 2760
                                      F  S  D  R  A  G  D  V  G  V  L  A  L  G  W  G  S  G                              491
                                         T   C                                                                T
caagctcacggtgccacaaaatatagCACTGCGAACTTCCCGTACTTGGTCGCAgtaggtctatgcttctgctgcgtttccgctgtaactgacgcgccttagCCCGTCGACGCGATTACT 2880
                            T  A  N  F  P  Y  L  V  A                                                  P  V  D  A  I  T  506
              A  C
GCTCGTGCGTCGCAGGATGGGACGACCGTGTCGTCGTCACTGAGTGACACCGACCTGACGGGCGCTGCTAACACCGCGACCGGGAAGGACGTCGCGATGGTGTTCATCACGGCCGACAGC 3000
 A  R  A  S  Q  D  G  T  T  V  S  S  S  L  S  D  T  D  L  T  G  A  A  N  T  A  T  G  K  D  V  A  M  V  F  I  T  A  D  S  546
                                                                                                   T
GGgtgcgtgtgtctggccctgttcttatggcgcacgctgatatttttacagTGAAGGATACCTTACTGTCGAGGGCAATGCAGGTGACCGTAACGACCTCCAGGCATGGCACGGAGGTgt 3120
 G                                                   E  G  Y  L  T  V  E  G  N  A  G  D  R  N  D  L  Q  A  W  H  G  G    569
                                                               C                                C
aagtgcgcctaacatcgtgactacatgcccgcttatacacagacgcctctagGATGCGCTGGTTCAGCAGGTCGCGAGCCACAACAAGAACACGATTGTCGTGATCAACAGTGTCGGACC 3240
                                                     D  A  L  V  Q  Q  V  A  S  H  N  K  N  T  I  V  V  I  N  S  V  G  P 592
TATCAATATGGAAGCCTGGGTCAACCACCCGAACGTCACCGCTATTgtgagtgaccgtgtgcattggagagccctgctgctgacgccgctcggttaactagGTTTGGAGTGGCCTCCCTG 3360
  I  N  M  E  A  W  V  N  H  P  N  V  T  A  I                                                         V  W  S  G  L  P   613
GTCAGGAAGCAGgtatgcatatacgcgcaagattcgctgtgcgaaacactcaccctgtgcccgatagGCAACGCGGTCACTGACGTTCTATTTGGCGCGGTCAACCCTGGgtacgtggtc 3480
G  Q  E  A                                                         G  N  A  V  T  D  V  L  F  G  A  V  N  P  G           632
                                                                                                 A
gaccatgtcgttaggcgtagggacggtatgctgacaccgaccacagCGGCAAGCTGCCCTTCACCATCGGCAAGTCGATCAGCGACTACTCCGCGCAGATCATCACCACCGGCTCGGGCA 3600
                                                G  K  L  P  F  T  I  G  K  S  I  S  D  Y  S  A  Q  I  I  T  T  G  S  G   656
TCGTGCCCATCCCGTACAACGAGGGCCTCTTCATCGACTACCGGCACTTTGACCAGgtgtgtaccggggtgcgtgcgcagcgcgcagcgtgctgacgggggcgtgcttagCGGGCATCG 3720
I  V  P  I  P  Y  N  E  G  L  F  I  D  Y  R  H  F  D  Q                                                        A  G  I   678
       A
CACCGCGCTTCGAGTTCGGGTTCGGCCTCTCGTACACGACGTTCGACTACTCGAACCTCGTCATCACGGGCTCGACGGCCGGCGGCACGCGTCAGCCCCCGGGGCCCGGCTCGTCGCTCG 3840
A  P  R  F  E  F  G  F  G  L  S  Y  T  T  F  D  Y  S  N  L  V  I  T  G  S  T  A  G  G  T  R  Q  P  P  G  P  G  S  S  L   718
ACCCATGgtgcgccatttccctttcgtaaagtcacctcgctgacgcaggtcgctgcgcccgcgcagGCTGCACGACTCGGTCGTCACCGTGTCCTTCACGCTGACGAACAACGGCACGGT 3960
D  P  W                                                             L  H  D  S  V  V  T  V  S  F  T  L  T  N  N  G  T  V 739
                                                                                                   G
CGACGGGACGGAGGTGCCGCAGCTGTACCTGAGCCCGCCCGCGAGCGCGAAGAGCGCGCCGCAGAACCTGAAGGGCTTCGACAGCGTATTCCTCCCCGCAGGCGCGTCGACGACGGTCTC 4080
  D  G  T  E  V  P  Q  L  Y  L  S  P  P  A  S  A  K  S  A  P  Q  N  L  K  G  F  D  S  V  F  L  P  A  G  A  S  T  T  V  S 779
   T
TTTCGAGCTCTCGCGCTACTCGTTCTCCGTCTGGGACGTCGTCTCCCAGAGCTGGCAGATCCCTGCTGGTGTCACCGGCATCTCCGTGGGCGCGAGCAGCAGAGATTTGCGCCTCAAGGG 4200
  F  E  L  S  R  Y  S  F  S  V  W  D  V  V  S  Q  S  W  Q  I  P  A  G  V  T  G  I  S  V  G  A  S  S  R  D  L  R  L  K  G 819
                   CG                  T                      A   CG
CTCAATCACGAACTGAGGG--CGAGGAACGTGCCGTGGTAAAAATGTATTCGTAGTGCACTGTGCACTATCTTTCCTGGCTTGACATATGTAAACACTTCCCATTTCGCCGACATAGCGC 4318
  S  I  T  N  *                                                                                                          823
GAGGTACACTGACTGACTTTCACATGGAGGCACGTTTGCCCATATAAGCTAGCTTCCCCTTGCGTGCGAACAGGTAACCACGACGGGTAAAAATGTTTCTGCAGCCCAACCATGTACAAA 4438
ACTACAGATTAGGGGGGATCACGTCAGTGCGAAGAGTCATGTTGCTAGAAGTGAATATGCGGGTAGTACAATGATATATCCAGACTCGGGGAGTTACGTCGCCAGTGACGCTTTTGA   4555
```

FIG. 1. Nucleotide and deduced amino acid sequences of CBGL from *P. chrysosporium*. Genomic and amino acid sequences were derived from *cbgl-2*. The amino acid sequence deduced from the cDNA sequence of *cbgl-1* was the same as that of *cbgl-2* except at positions indicated in the line below the amino acid sequence in parentheses. The exon sequence of *cbgl-1* is the same as that of *cbgl-2* except at the positions indicated in the line above the gene sequence. Nucleotides and amino acids are numbered on the right. The potential signal peptide sequence is overlined. The potential CBD is boxed. Potential N-glycosylation sites are in boldface type.

```
Saf1     1 MLMIVQLLVFALGLAVAVPIQNYTQSPSQ-RDE-SSQWVSPHYYPTPQGG--RLQDVWQEAYARAKAIVGQMTIVEKVNL
Saf2     1 MLLILELLVLIIGLGVALPVQTHNLTDNQGFDEESSQWISPHYYPTPQGG--RLQGVWQDAYTKAKALVSQMTIVEKVNL
Phc      1 -----------------------------------------------VAN---VSPEWAAAYVKAQAAVAKLSVTDMVNL
Trr      1 -------MRYRTAAALALATGPFARADS--------HSTSGASAEAVVPP---AGTPWGTAYDKAKAALAKLNLQDKVGI
Asa      1 -------MKLSWLEAAALTAASVVSADE--------LAFSPPFYPSPWAN---GQGEWAEAYQRAVAIVSQMTLDEKVNL
Sel      1 -------MVSSLFNIAALAGAVIALSHED----QSKHFTTIPTFPTPDST---GEG-WKAAFEKAADAVSRLNLTQKVAL
Pic      1 ----MKSTIIILSVLAAATAKNISKAEMENLEHWWSYGRSDPVYPSPEIS---GLGDWQFAYQRAREIVALMTNEEKTNL
Pia      1 ----MLLPLYGLASFLVLSQAALVNTSAPQASNDDPFNHSPSFYPTPQGGRI-NDGKWQAAFYRARELVDQMSIAEKVNL

Saf1    77 TTGTGWQLDPCVGNTGSVPRFG-IPNLCLQDGPLGVRFADFVTGYPSGLATGATFNKDLFLQRGQALGHEFNSKGVHIAL
Saf2    79 TTGTGWQLGPCVGNTGSVPRFG-IPNLCLQDGPLGVRLTDFSTGYPSGMATGATFNKDLFLQRGQALGHEFNSKGVHIAL
Phc     31 ATGVQWQKGPCVGNTPAISSIPGFTGLCLQDSPVGVRYADGTSVFPPEINVAATWNRTLMRQRGAAMGAEFKGKGVHVAL
Trr     63 VSGVGWNGGPCVGNTSPASKIS-YPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVIL
Asa     63 TTGTGWELEKCVGQTGGVPRLN-IGGMCLQDSPLGIRDSDYNSAFPAGVNVAATWDKNLAYLRGQAMGQEFSDKGIDVQL
Sel     66 TTGTTAGLS-CNGNIAPIPEIN-FSGLCLADGPVSVRIADLATVFPAGLTAAATWDRQLIYERARALGSEFRGKGSQVHL
Pic     74 TFGSSGDTG-CSGMISDVPDVD-FPGLCLQDAGNGVRGTDMVNAYASGLHVGASWNRQLAYDRAVYMGAEFRHKGVNVLL
Pia     76 TTGVGSASGPCSGNTGSVPRLN-ISSICVQDGPLSVRAADLTDVFPCGMAASSSFNKQLIYDRAVAIGSEFKGKGADAIL

Saf1   156 GPAVGPLGVKARGGRNFEAFGSDPYLQGTAAAATIKGLQENNVMACVKHFIGNEQEKYRQPDD-----INPATNQTTKEA
Saf2   158 GPAVGPLGVKARGGRNFEAFGSDPYLQGIAAAATIKGLQENNVMACVKHFIGNEQDIYRQPSNS---KVDPEYDPATKES
Phc    111 GPMMN-LMRVPAAGRNWEGGGGDPFLSGEVAFETISGIQSSGAQACAKHFINNEQEHFR-------------------DS
Trr    142 GPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNR-------------------ET
Asa    142 GPAAGPLGRSPDGGRNWEGFSPDPALTGVLFAETIKGIQDAGVVATAKHYILNEQEHFRQVAE------AAGYGFNISDT
Sel    144 GPASGALGRHPLGGRNWESFSPDPYLSGVAMDFSIRGIQEMGVQANRKHFIGNEQETQR--SN------TFTDDGTEIQA
Pic    152 GPVVGPIGRVATGGRNWEGFTNDPYLAGALVYETTKGIQEN-VIACTKHFIGNEQETNR----------NPSG--TYNQS
Pia    155 GPVYGPMGVKAAGGRGWEGHGPDPYLEGVIAYLQTIGIQSQGVVSTAKHLIGNEQEHFRFAKKDKHAGKIDPGMFNTSSS

Saf1   231 ISANIPDRAMHALYLWPFADSVRAGVGSVMCSYNRVNNTYACENSYMMNHLLKEELGFQGFVVSDWGAQLSGVYSAISGL
Saf2   235 ISANIPDRAMHELYLWPFADSIRAGVGSVMCSYNRVNNTYSCENSYMINHLLKEELGFQGFVVSDWAAQMSGAYSAISGL
Phc    171 SSSNVDDRTEHELYGHPFLRSVQANVASVMCSYNQINGTFSCENEKTLSGLLKGEYGFQGYVMSDWWATHSGAPAVNAGL
Trr    203 ISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWDAQHTTVQSANSGL
Asa    216 ISSNVDDKTIHEMYLWPFADAVRAGVGAIMCSYNQINNSYGCQNSYTLNKLLKAELGFQGFVMSDWGAHHSGVGSALAGL
Sel    216 ISSNIDDRTMHELYLWPFANAVRSGVASVMCSYNRLNQTYACENSKLMNGILKGELGFQGYVVSDWYATHSGVESVNAGL
Pic    219 VSANIDDKTMHELYLWPFQDSVRAGLGSIMGSYNRVNNSYACKNSKVLNGLLKSELGFQGFVVSDWGGQHTGIASANAGL
Pia    235 LSSEIDDRAMHEIYLWPFAEAVRGGVSSIMCSYNKLNGSHACQNSYLLNYLLKEELGFQGFVMTDWGALYSGIDAANAGL

Saf1   311 DMSMPGEVYGGWNTG---TSFWGQNLTKAIYNETVPIERLDDMATRILAALYATNSFPTED---HLPNFSSWTTKEYGNK
Saf2   315 DMSMPGELLGGWNTG---KSYWGQNLTKAVYNETVPIERLDDMATRILAALYATNSFPTKD---RLPNFSSFTTKEYGNE
Phc    251 DMTMPGDETLSSGTT-----YFGQNLVNAVNSGQVSQARVKDMATRILAAWYLLGQDQNF----PAVNFNSWNSGQG---
Trr    283 DMSMPG-TDFNGNNR-----LWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQ--------AGYPSFNISR----
Asa    296 DMSMPGDITFDSATS-----FWGTNLTIAVLNGTVPQWRVDDMAVRIMAAYYKVGRDRLY----QPPNFSSWTRDEYGFK
Sel    296 DMTMPGPLDSPSTALRPPPSYLGGNLTEAVLNGTIPEARVDDMARRILMPYFFLGQDTDFPTVDPSTGFVFARTYNYPDE
Pic    299 DMAMPS-------ST-----YWEEGLIEAVKNGTVDQSRLDDMATRIIAAWYKYARLDD-------PGFGMPVSLAED--
Pia    315 DMDMPCEA-----Q------YFGGNLTTAVLNGTLPQDRLDDMATRILSALIYSGVHNP-----DGPNYNAQTFLTEGHE

Saf1   385 -YADNTTEIVKVNYNVDPSNDYFTEDTALKVAEESIVLLKNENNTLPISPEKAKRLLLSGIAAG---PDPIGYQ-----C
Saf2   389 FFVDKTSPVVKVNHFVDPSND-FTEDTALKVAEESIVLLKNEKNTLPISPNKVRKLLLSGIAAG---PDPKGYE-----C
Phc    319 -------------QHVNVSGN--HASLIRTIGAASQILLKNANGALPLK--KPKTIGIIGNGAG---SNPNGPN----AF
Trr    345 -----------------NVQGN--HKTNVRAIARDGIVLLKNDANILPLK--KPASIAVVGSAAI---IGNHARNSP--SC
Asa    367 YFYPQEGPYEKVNHFVNVQRN--HSEVIRKLGADSTVLLKNNN-ALPLTG-KERKVAILGEDAG---SNSYGAN----GC
Sel    376 YLTLGGLDPYNPPPARDVRGN--HSDIVRKVAAAGTVLLKNVNNVLPLK--EPKSVGIFGNGAA---DVTEGLTFTGDDS
Pic    358 ------------HELVDARDP-AAASTIFQGAVEGHVLVKNEN-ALPLK--KPKYISLFGYDGVSTDVNTVGGGFSFFSF
Pia    379 YFKQQEGDIVVLNKHVDVRSD-INRAVALRSAVEGVVLLKNEHETLPLGREKVKRISILGQAAG---DDSKGTS-----C

Saf1   456 EDQSCTNGALFQGWGSGSVGSPKYQVTPFEEISYLARKNKMQFDYIRESYDLAQVTK--VASDAHLSIVVVSAASGEGYI
Saf2   460 SDQSCVDGALFEGWGSGSVGYPKYQVTPFEEISANARKNKMQFDYIRESFDLTQVST--VASDAHMSIVVVSAVSGEGYL
Phc    375 SDRAGDVGVLALGWGSGTANFP-YLVAPVDAITARASQDGTTVSSSLSDTDLTGAAN--TATGKDVAMVFITADSGEGYL
Trr    400 NDKGCDDGALGMGWGSGAVNYP-YFVAPYDAINTRASSQGTQVTLSNTDNTSSGAS---AARGKDVAIVFITADSGEGYI
Asa    436 SDRGCDNGTLAMAWGSGTAEFP-YLVTPEQAIQAEVLKHKGSVYAITDNWALSQVET--LAKQASVSLVFVNSDAGEGYI
Sel    449 GPWGADIGALSVGGGSGAGRHT-HLVSPLAAIRKRTESVGGRVQYLLSNSRIVNDDFTSIYPTPEVCLVFLKTWAREGTD
Pic    422 DVKAIENKTLISGGGSG-TNTPSYVDAPFNAFVAKARED-----NTFLSWDFTSAEP-VANPASDACIDFINAAASEGYD
Pia    450 SLRGCGSGAIGTGYGSGAGTFS-YFVTPADGIGARAQQEKISYEFIGDSWNQAAAMD--SALYADAAIEVANSVAGEEIG

Saf1   534 TVDGNQGDRKNLTLWNNGDKLIETVAENCANTVVVVTSTGQINFEGFADHPNVTAIVWAGPLGDRSGTAIANILFGKANP
Saf2   538 IIDGNRGDKNNVTLWHNSDNLIKAVAENCANTVVVITSTGQVDVESFADHPNVTAIVWAGPLGDRSGTAIANILFGNANP
Phc    452 TVEGNAGDRNDLQAWHGGDALVQQVASHNKNTIVVINSVGPINMEAWVNHPNVTAIVWSGLPGQEAGNAVTDVLFGAVNP
Trr    476 TVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSP
Asa    513 SVDGNEGDRNNLTLWKNGDNLIKAAANNCNNTIVVIHSVGPVLVDEWYDHPNVTAILWAGLPGQESGNSLADVLYGRVNP
Sel    528 RLS-YEND------WN-STAVVNNVARRCPNTIVVTHSGG-INTMPWADNANVTAILAAHYPGQENGNSIMDILYGDVNP
Pic    495 RPN--LAD------KY-SDKLVEAVASQCSNTIVVIHNAGIRLVDNWIEHENVTGVILAHLPGQDTGTSLIEVLYGNQSP
Pia    527 DVDGNYGDLNNLTLWHNAVPLIKNISSINNNTIVIVTSGQQIDLEPFIDNENVTAVIYSSYLGQDFGTVLAKVLFGDENP

Saf1   614 SGHLPFTIAKT---DDDYIPIETYSPSSG-EPED--NHLVENDLLVDYRYFEEKNIEPRYAFGYGLSYNEYEVSNAKVSA
Saf2   618 SGHLPFTVAKS---NDDYIPIVTYNPPNG-EPED--NTLAEHDLLVDYRYFEEKNIEPRYAFGYGLSYNEYKVSNAKVSA
Phc    532 GGKLPFTIGKS---ISDYSAQIITTGSGI-VP-----IPYNEGLFIDYRHFDQAGIAPRFEFGFGLSYTTFDYSNLVITG
Trr    556 SGKLVYTIAKS---PNDYNTRIVSGGS----------DSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLS
Asa    593 GAKSPFTWGKTREAYGDYLVRELNNGNGA-PQ-----DDFSEGVFIDYRGFDKRNETPIYEFGHGLSYTTFNYSGLHIQV
Sel    599 SGRLPYTIPKL---ATDYDFPVVNITNEA-QDPYVWQADFTEGLLIDYRHFDARNITPLYEFGYGLSYTTFEIEGVANLV
Pic    566 SGRLPYTVAKK---ASDYGGLLWPTEPEGDLDLYFPQSNFTEGVYIDYKYFIQKNITPRYEFGYGLTYTTFDYSELEVDA
Pia    607 SGKLPFTIAKD---VNDYIPVIEKVDVP--DP----VDKFTESIYVDYRYFDKYNKPVRYEFGYGLSYSNFSLSDIEIQT
```

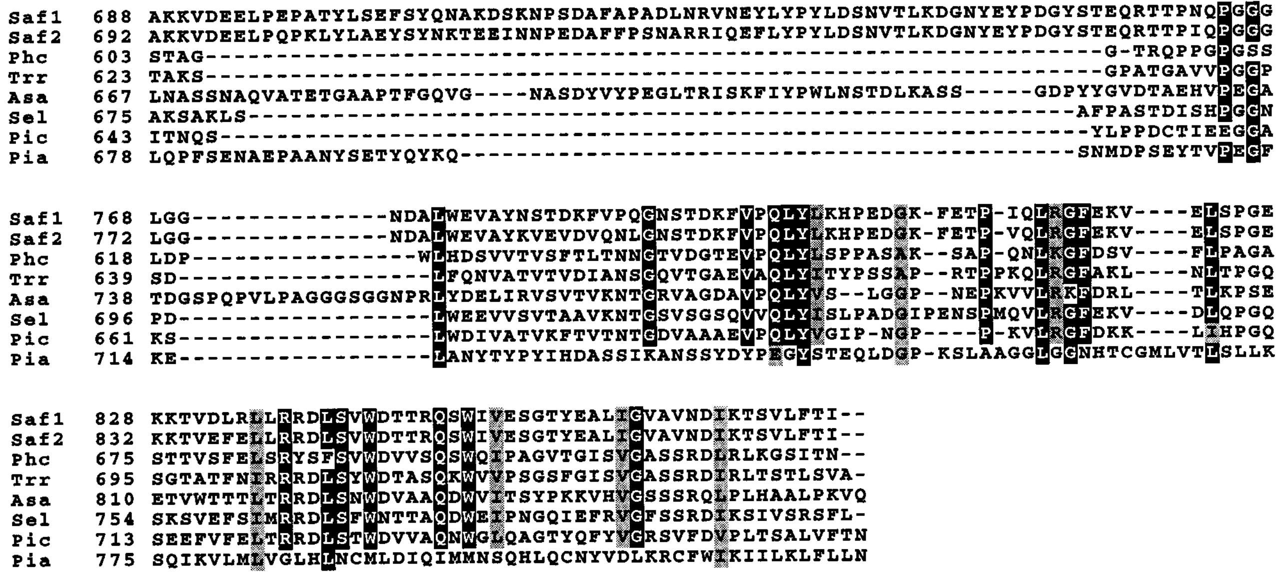

FIG. 2. Comparison of the catalytic domain sequence of CBGL with the β-glucosidases from *T. reesei* (Trr) (3), *A. aculeatus* (Asa) (15), *S fibuligera* (Saf1 and Saf2) (21), *Pichia anomala* (Pia) (17), *P. capsulata* (Pic) (14), and *Septoria lycopersici* (Sel) (24).

branes and hybridized to a ^{32}P-labeled 1.4-kb *Sal*I cDNA fragment of *cbgl* (5).

Seventy-two positive clones of *cbgl* were isolated by immunoscreening of the *P. chrysosporium* cDNA library. A full-length clone was isolated by screening the cDNA library with a *Msc*I + *Nde*I fragment from the genomic clone *cbgl-2*. Genomic clones were isolated by screening a λEMBL3 genomic library of *P. chrysosporium* with a 500-bp *Apa*I fragment from the 3′ region of the cDNA sequence. This screening yielded ~50 positive genomic clones. Restriction fragment analyses of five clones which hybridized strongly to the probe indicated that they were very similar, and one of them was subcloned and sequenced.

***cbgl* cDNA sequence.** Sequence analysis of the cDNA clone (2.4 kb) revealed an open reading frame consisting of 2,469 bp encoding 823 amino acids, including a 21-amino-acid N-terminal signal peptide sequence (Fig. 1). Prediction of the signal peptide cleavage site suggested that Gln22 was the N-terminal amino acid (22). The mature CBGL apparently consists of 802 amino acids and has an apparent molecular weight of 83,439. The CBGL molecular weight as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis was 114,000. CBGL is a glycoprotein, and the difference in molecular weight could be attributable to the carbohydrate portion. The cDNA sequence revealed six potential N-glycosylation sites conforming to the general rule Asn-X-Thr/Ser, in which X is not a proline (4). In addition, numerous O-glycosylation sites are possible.

CBD. Analysis of the *cbgl* cDNA sequence suggested that the amino acid sequence from 22 to 57 has a high sequence similarity to the conserved cellulose-binding sequence of CBHII from *P. chrysosporium* and other fungal CBD sequences (11). This confirmed our prediction that CBGL possesses a CBD similar to that of cellulases (20). This domain is connected to the C-terminal catalytic domain via a 43-amino-acid linker region.

Catalytic domain. Approximately amino acids 101 to 823 at the C terminus form the catalytic domain. Sequence analysis suggests that CBGL should be categorized as a family 3 glycosylhydrolase along with the extracellular β-glucosidases from *Trichoderma reesei*, *Aspergillus aculeatus*, *Saccharomycopsis fibuligera*, and *Pichia capsulata* (3, 13, 14, 17, 21). Asp and Glu have been found in the active sites of numerous glycosidases, including β-glucosidase, cellulase, and amylase (7, 23, 28). Analysis of the catalytic domain sequences of eight family 3 fungal and yeast glycosidases indicated conservation of 13 acidic amino acid residues. Sequences surrounding seven conserved residues are preserved (Fig. 2). Potentially, any two of the conserved residues could be involved in catalysis.

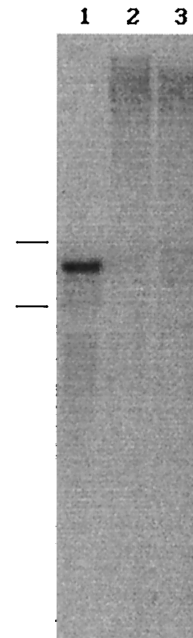

FIG. 3. Northern blot analysis of *P. chrysosporium* RNA. Total RNA was isolated from 11-day-old mycelia obtained from 1% cellulose (lane 1), glucose (lane 2), and cellobiose (lane 3) cultures. The blot was probed with ^{32}P-labeled CBGL cDNA. Bars to the left indicate the positions of 18S and 28S rRNA (from top to bottom).

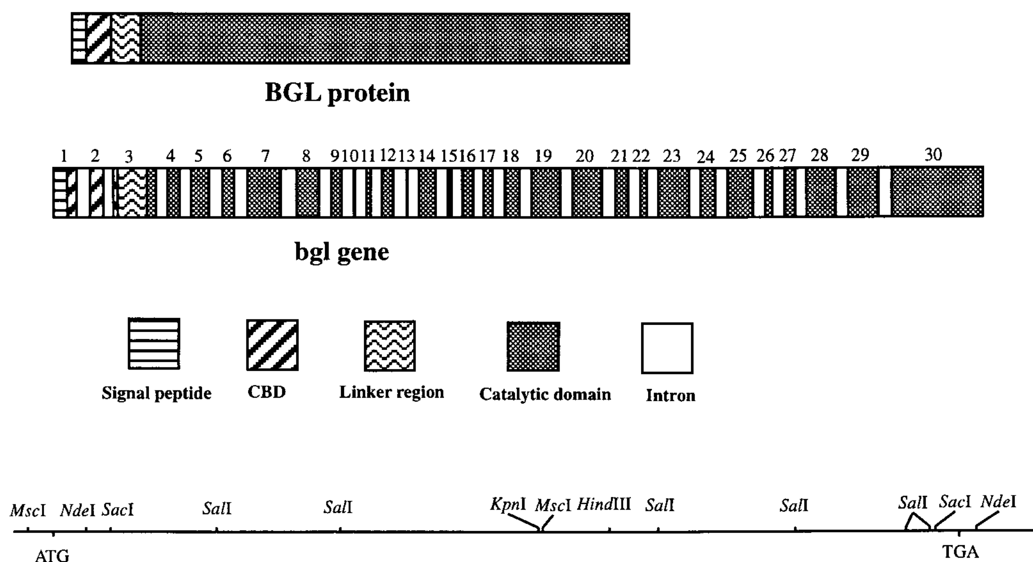

FIG. 4. Schematic representation of the protein and gene structures of CBGL and the restriction map of *cbgl-2*.

CBGL expression. *P. chrysosporium* produced CBGL abundantly only when cellulose was provided as the sole carbon source (2, 20). CBGL production was not observed in cultures supplemented with either glucose or cellobiose (2). To further understand CBGL expression in *P. chrysosporium*, total RNA was isolated from 11-day-old cellulose, cellobiose, or glucose cultures and analyzed by Northern blotting (Fig. 3). A band corresponding to 2.4 kb was observed only with the RNA isolated from cellulose-grown cells. Also, the size of this RNA was very similar to the size of the cDNA insert. These preliminary findings suggest that either cellulose or one of its degradation products might be controlling the expression of CBGL at the transcriptional level.

Gene sequence of *cbgl-2*. *cbgl-2* consisted of 4,555 bp, including 182 bp in the 5′ flanking region and 339 bp in the 3′ flanking region (Fig. 1). The 5′ upstream region contained a potential TATAA box (TATAAGT) 64 bp upstream from the translation start codon. Comparison of the genomic and cDNA sequences of CBGL indicated the presence of 29 introns varying in size from 47 to 68 bp (Fig. 1). All of the intron splice junctions conformed to the GT–AG rule. Exon 1 codes for the signal peptide and a portion of the CBD (Fig. 4). Exon 2 codes only for the CBD. Exon 3 codes for the rest of the CBD, the linker peptide, and a small sequence of the catalytic domain. Exons 4 to 30 code for the catalytic domain. Interestingly, exons 10 and 13 code for one and two amino acids, respectively (Fig. 4). In contrast to *cbgl*, *T. reesei bglu1* has only two introns (3).

Exon sequences of *cbgl-2* exhibited 98% similarity to the *cbgl* cDNA sequence. A total of 50 bp in the sequences (in the exon regions) did not match the cDNA sequence; however, the amino acid sequences differed only at four locations (Fig. 1). Restriction analysis of *P. chrysosporium* DNA indicated that, except for *Hin*dIII restriction, only one fragment from all the other restrictions hybridized to a 1.4-kb *Sal*I fragment of *cbgl-2* (Fig. 5). This result suggested that *cbgl* is probably encoded by two alleles (*cbgl-1* and *cbgl-2*) of a single gene. The cDNA sequence was presumably derived from *cbgl-1*.

***cbglu-1* and *cbglu-2* allelism.** *P. chrysosporium* is a heterokaryon with two or more genetically distinct nuclei (1). The genomic library from which the *cbgl* clones were isolated was derived from such a heterokaryon. However, the basidiospores are homokaryons and contain two identical nuclei. If *cbgl-1* and *cbgl-2* are truly allelic, then they should segregate among homokaryons. Segregation of allelic variants of lignin peroxidase, glyoxal oxidase, and cellobiose dehydrogenase from

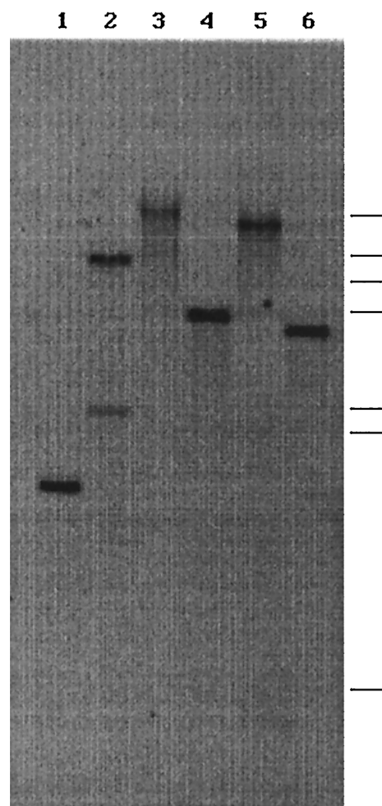

FIG. 5. Southern analysis of genomic DNA from *P. chrysosporium*. Genomic DNA, isolated by standard procedures, was digested with restriction enzymes *Sal*I (lane 1), *Hin*dIII (lane 2), *Bam*HI (lane 3), *Nde*I (lane 4), *Eco*RI (lane 5), and *Sac*I (lane 6). The blot was probed with a ^{32}P-labeled 1.4-kb *Sal*I fragment of *cbgl-2*. Bars indicate the positions of molecular size standards (from top to bottom: 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.6 kb).

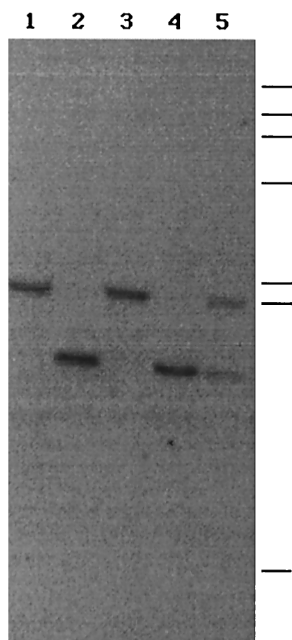

FIG. 6. Segregation of CBGL alleles into homokaryons. DNA from four separate single spore cultures (lanes 1 to 4) and one parenteral heterokaryon culture (lane 5) of *P. chrysosporium* OGC101 were restricted with *Sal*I, size fractionated on an agarose gel, and probed with a 1.4-kb *Sal*I of *cbgl-2*. Bars indicate the positions of molecular size standards (from top to bottom) of 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.6 kb.

P. chrysosporium is known (10, 16, 19). A comparison of the cDNA sequences of *cbgl-1* and the exon sequences of *cbgl-2* suggested that *cbgl-2* has at least one extra *Sal*I site at nucleotide 2866 (Fig. 4). This difference was utilized in identifying the two alleles. DNAs from homokaryotic and heterokaryotic cultures were restricted with *Sal*I and probed with a 1.4-kb *Sal*I fragment (bp 1446 to 2866). The probe was expected to hybridize to only a 2-kb fragment from *cbgl-1*, a 1.4-kb fragment from *cbgl-2*, and two fragments (1.4 and 2 kb) from the wild-type heterokaryon. Southern analysis suggested that only one fragment (1.4 or 2 kb) from homokaryon DNA and two fragments from wild-type DNA can hybridize to the probe (Fig. 6). These findings support the proposal that *cbgl-1* and *cbgl-2* are alleles.

β-Glucosidase multiplicity. Smith and Gold (26) partially purified an extracellular β-glucosidase (M_r, 90,000) from *P. chrysosporium* OGC101. CBGL is produced by the same strain in cultures optimized for low extracellular protease levels (2). Proteolytic hydrolysis of CBGL produces two non-cellulose-binding forms (M_rs, 96,000 and 98,000). Thus, the β-glucosidase isolated by Smith and Gold (26) was probably a degradation product of CBGL. At this time, there is no reason to believe that the low-molecular-weight glucosidase could arise from differential splicing of *cbgl-1* or *cbgl-2*. Deshpande et al. (9) have reported five extracellular β-glucosidases from *P. chrysosporium*, with molecular weights ranging from 165,000 to 182,000. A comparison with the current findings is not worthwhile because of strain differences and the variation in culturing conditions.

Nucleotide sequence accession numbers. The *P. chrysosporium* OGC101 CBGL cDNA and gene sequence data reported here have been deposited in GenBank under accession no. AF036873 and AF036872.

This work was supported by grant DE-FG06-92ER20065 from the U.S. Department of Energy, Office of Basic Energy Sciences.

We thank Michael Gold of the Oregon Graduate Institute for providing the λEMBL3 genomic library of *P. chrysosporium* OGC101.

REFERENCES

1. **Alic, M. A., C. Letzring, and M. H. Gold.** 1987. Mating system and basidiospore formation in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. **53:**1464–1469.
2. **Bao, W., E. Lymar, and V. Renganathan.** 1994. Production of cellobiose dehydrogenase and β-glucosidase by the cellulose-degrading cultures of *Phanerochaete chrysosporium* in shake flasks. Appl. Microbiol. Biotechnol. **42:**642–646.
3. **Barnett, C. C., R. M. Berka, and T. Fowler.** 1991. Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. Bio/Technology **9:**562–567.
4. **Bause, E.** 1983. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. **209:**331–336.
5. **Brown, T.** 1994. Analysis of DNA sequences by blotting and hybridization, p. 2.9.1–2.9.15. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
6. **Brown, T.** 1994. Analysis of RNA by northern and slot blot hybridization, p. 4.9.1–4.9.16. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
7. **Buisson, G., E. Duee, R. Haser, and F. Payan.** 1987. Three dimensional structure of porcine pancreatic α-amylase at 2.9 Å resolution. Role of calcium in structure and activity. EMBO J. **6:**3909–3916.
8. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:**5294–5299.
9. **Deshpande, V., K.-E. Eriksson, and B. Pettersson.** 1978. Production and purification of 1,4-β-glucosidase enzymes from *Sporotrichum pulverulentum*. Eur. J. Biochem. **90:**191–198.
10. **Gaskell, J., E. Dieperink, and D. Cullen.** 1991. Genomic organization of lignin peroxidase genes of *Phanerochaete chrysosporium*. Nucleic Acids Res. **19:**599–603.
11. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
12. **Gold, M. H., and M. Alic.** 1993. Molecular biology of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Microbiol. Rev. **57:**605–622.
13. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **293:**781–788.
14. **Janbon, G., R. Magnet, A. Arnaud, and P. Galzy.** 1995. Cloning and sequencing of β-glucosidase gene from *Candida molischiana* 35M5N. Gene **165:**109–113.
15. **Kawaguchi, T., T. Enoki, S. Tsurumaki, J. Sumitani, M. Ueda, and M. Arai.** 1996. Cloning and sequencing of the cDNA encoding β-glucosidase from *Aspergillus aculeatus*. Gene **173:**287–288.
16. **Kersten, P. J., C. Witek, A. Vanden Wymelenberg, and D. Cullen.** 1995. *Phanerochaete chrysosporium* glyoxal oxidase is encoded by two allelic variants: structure, genomic organization, and heterologous expression of *glx1* and *glx2*. J. Bacteriol. **177:**6106–6110.
17. **Kohchi, C., and A. Toh-e.** 1985. Nucleotide sequence of *Candida pelliculosa* β-glucosidase gene. Nucleic Acids Res. **13:**6273–6282.
18. **Li, B., S. R. Nagalla, and V. Renganathan.** 1996. Cloning of a cDNA encoding cellobiose dehydrogenase, a hemoflavoenzyme from *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. **62:**1329–1335.
19. **Li, B., S. R. Nagalla, and V. Renganathan.** 1997. Cellobiose dehydrogenase from *Phanerochaete chrysosporium* is encoded by two allelic variants. Appl. Environ. Microbiol. **63:**796–799.
20. **Lymar, E. S., B. Li, and V. Renganathan.** 1995. Purification and characterization of a cellulose-binding β-glucosidase from cellulose-degrading cultures of *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. **61:**2976–2980.
21. **Machida, M., I. Ohtsuki, S. Fukui, and I. Yamashita.** 1988. Nucleotide sequences of *Saccharomycopsis fibuligera* genes for extracellular β-glucosidases as expressed in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. **54:**3147–3155.
22. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
23. **Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A. Jones.** 1990. Three dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249:**380–386.

24. **Sandrock, R. W., D. Dellapenna, and H. D. Van Etten.** 1995. Purification and characterization of β-2-tomatinase, an enzyme involved in the degradation of α-tomatine, and isolation of the gene encoding β-2-tomatinase from *Septoria lycopersici*. Mol. Plant-Microbe Interact. **8:**960–970.
25. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
26. **Smith, M. H., and M. H. Gold.** 1979. *Phanerochaete chrysosporium* β-glucosidase: induction, cellular localization, and physical characterization. Appl. Environ. Microbiol. **37:**938–942.
27. **Strauss, E. C., J. A. Kobori, G. Siu, and L. E. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154:**353–360.
28. **Wang, Q., D. Trimbur, R. Graham, R. A. J. Warren, and S. G. Withers.** 1995. Identification of the acid/base catalyst in *Agrobacterium fecalis* β-glucosidase by kinetic analysis of mutants. Biochemistry **34:**14554–14562.